



Conservation genetics of the critically endangered Siamese rosewood (*Dalbergia cochinchinensis*)

recommendations for management and sustainable use

Hartvig, Ida; So, Thea; Changtragoon, Suchitra; Tran, Hoa Thi; Bouamanivong, Somsanith; Ogden, Rob; Senn, Helen; Vieira, Filipe Garrett; Turner, Frances; Talbot, Richard; Theilade, Ida; Nielsen, Lene Rostgaard; Kjær, Erik Dahl

Published in:
Conservation Genetics

DOI:
[10.1007/s10592-020-01279-1](https://doi.org/10.1007/s10592-020-01279-1)

Publication date:
2020

Document version
Peer reviewed version

Document license:
[CC BY](#)

Citation for published version (APA):
Hartvig, I., So, T., Changtragoon, S., Tran, H. T., Bouamanivong, S., Ogden, R., Senn, H., Vieira, F. G., Turner, F., Talbot, R., Theilade, I., Nielsen, L. R., & Kjær, E. D. (2020). Conservation genetics of the critically endangered Siamese rosewood (*Dalbergia cochinchinensis*): recommendations for management and sustainable use. *Conservation Genetics*, 21, 677-692. <https://doi.org/10.1007/s10592-020-01279-1>

Ida Hartvig^{1a*}, So Thea², Suchitra Changtragoon³, Hoa Thi Tran⁴, Somsanith Bouamanivong⁵, Rob Ogden⁶, Helen Senn⁷, Filipe Garrett Vieira^{8b}, Frances Turner⁹, Richard Talbot⁹, Ida Theilade¹⁰, Lene Rostgaard Nielsen¹, Erik Dahl Kjær¹

Conservation genetics of the critically endangered Siamese rosewood (*Dalbergia cochinchinensis*):
recommendations for management and sustainable use

¹Department of Geosciences and Natural Resource Management, University of Copenhagen,
Rolighedsvej 23, 1958 Frederiksberg C, Denmark

²Forest and Wildlife Science Research Institute, Forestry Administration, Ministry of Agriculture
Forestry and Fisheries, No. 1017, Hanoi Street, Sen Sok District, Phnom Penh, Cambodia 12101

³Forest and Plant Conservation Research Office, Department of National Parks, Wildlife and Plant
Conservation, Ministry of Natural Resources and Environment, 61 Phaholyothin, Chatuchak,
Bangkok, Thailand 10900

⁴Forest Genetics and Conservation, Center for Biodiversity and Biosafety, Institute of Agricultural
Genetics, Vietnam Academy of Agricultural Sciences, Pham Van Dong Road, Hanoi, Vietnam

⁵National Herbarium of Laos, Biotechnology and Ecology Institute, Ministry of Science and
Technology, Vientiane, Laos

19 ⁶Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Easter
20 Bush Campus, Edinburgh, EH25 9RG, UK

21 ⁷RZSS WildGenes Laboratory, Royal Zoological Society of Scotland, 134 Corstorphine Road, EH12
22 6TS, Edinburgh, UK

23 ⁸Evolutionary Genomics, Natural History Museum of Denmark, University of Copenhagen,
24 Gothersgade 128, 1123 Copenhagen K, Denmark

25 ⁹Edinburgh Genomics, Ashworth labs, University of Edinburgh, Charlotte Auerbach Road, EH9 3FL
26 Edinburgh, UK

27 ¹⁰Global Development, Department of Food and Resource Economics, University of Copenhagen,
28 Rolighedsvej 21, 1958 Frederiksberg C, Denmark

29 ^a Current affiliation: Smithsonian Environmental Research Center, 647 Contees Wharf Road,
30 Edgewater 21037 MD, USA

31 ^b Current affiliation: Center for Genomic Medicine, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen E,
32 Denmark

33 * Corresponding author: ihla@ign.ku.dk

34

- 35 ORCID IDs
- 36 Ida Hartvig: 0000-0002-1509-6864
- 37 So Thea: na
- 38 Somsanith Bouamanivong: 0000-0003-1580-4389
- 39 Hoa Thi Tran: na
- 40 Rob Ogden: 0000-0002-2831-0428
- 41 Helen Senn: 0000-0002-3711-8753
- 42 Filipe Garrett Vieira: 0000-0002-8464-7770
- 43 Francis Turner: na
- 44 Richard Talbot: 0000-0002-6173-8266
- 45 Ida Theilade: 0000-0003-3502-1277
- 46 Lene Rostgaard Nielsen: 0000-0002-7214-8691
- 47 Erik Dahl Kjær: 0000-0001-8624-1611
- 48

49 Abstract

50 The highly valuable timber species *Dalbergia cochinchinensis* is severely threatened due to habitat loss
51 and illegal logging throughout its distribution in mainland Southeast Asia and is listed on CITES
52 Appendix II. This study proposes a strategy for conservation and sustainable management of the
53 species based on assessment of genetic structure within and among natural populations. We developed
54 SNP markers from RAD sequencing and used these in combination with SSR genotypes from a
55 previous study to analyze the genetic diversity in 26 populations of *D. cochinchinensis* across its entire
56 range in Laos, Thailand, Cambodia and Vietnam. The species is able of clonal reproduction and we
57 found that trees closer than 45 meters from each other can be clones. Genetic diversity and clustering
58 analysis showed a clear division of populations into five geographical groups with differing levels of
59 diversity. Assignment tests correctly identified the region of origin for approximately 90% of the
60 samples, which demonstrates that despite a low number of successfully developed SNPs, the SSR+SNP
61 marker panel has the potential for tracking the geographic origin of *D. cochinchinensis* timber for use
62 in CITES regulation and enforcement. We propose the five identified groups to be considered as
63 Management Units and that conservation and breeding programs should be based on a network of *in*
64 *situ* and *ex situ* conservation stands representing the genetic variation among and within these units. We
65 recommend that conservation efforts are directed towards community owned and managed lands, as
66 this has proven an effective strategy locally.

67 **Keywords:** rosewood, conservation genetics, *Dalbergia*, threatened species, timber species, CITES,
68 tropical trees

70 Introduction

71 Deforestation and degradation of tropical forests are continuing at high rates, with huge impacts on
72 biodiversity as well as on the people that depend on local forest resources (Gibson *et al.* 2011; Hansen
73 *et al.* 2013; FAO 2016). For timber tree species habitat loss is often accompanied by unsustainable
74 levels of industrial logging and a large number of tropical tree species are considered threatened
75 (Newton & Oldfield 2008; FAO 2014). While the readily observable consequences are local extinctions
76 and reduced overall population size, over time, logging and fragmentation is likely to result in the more
77 subtle loss of genetic diversity as well as increased inbreeding and differentiation among populations
78 (Young *et al.* 1996; Lowe *et al.* 2005; Degen *et al.* 2006). These factors potentially reduce the species'
79 ability to adapt to a changing environment, and may negatively impact breeding success (Amaral *et al.*
80 2004a). To ensure long-term viability of natural populations of threatened trees as well as genetic
81 resources for future breeding programs, it is necessary to implement effective conservation plans.

82 Forest tree conservation plans may include *in-situ* conservation efforts aiming at protection of areas
83 with natural populations of the species, or *ex-situ* conservation, for example, establishment of seed
84 banks or living conservation stands (Graudal *et al.* 2004). *Ex-situ* conservation stands can also function
85 as seed sources, and provide material for breeding programs as well as re-introductions (So *et al.*
86 2010). Both *in-situ* and *ex-situ* conservation measures should aim to conserve genetic diversity across
87 the entire geographic range of the species as well as take into account any specific meta-population
88 structure. A key conservation challenge is to protect the population genetic processes that support long-
89 term adaptation of the species, i.e. maintaining genetic connectivity to reduce inbreeding risk due to
90 fragmentation, while enabling continued natural selection in favor of adaptation to changing local

growth conditions (Eriksson *et al.* 1993). Insight into both genetic structure and genetic processes is therefore important in order to develop long-term species conservation programs.

Breeding/plantation programs have been suggested as important components of gene conservation for timber species in high demand. The idea is that commercial plantations can supply timber in large quantities and thereby reduce the demand and price of illegal timber. If breeding programs include seed collections based on assessments of genetic diversity and population genetic structure, and encourage planting material of local origin, these could also contribute to conservation of genetic diversity (Theilade *et al.* 2004).

Conservation strategies can also include initiatives that facilitate enforcement of existing regulations. The trade of timber species is regulated by a variety of national laws and international conventions such as the Convention on International Trade in Endangered Species (CITES), which require declarations concerning the species and geographic origin of timber in trade. DNA profiling has significant potential as a timber tracking tool (Ogden *et al.* 2009; Lowe & Cross 2011), and recent studies have shown that methods can accurately verify the geographic origin of wood samples of tropical timber species (Jolivet & Degen 2012; Degen *et al.* 2013; Blanc-Jolivet *et al.* 2017).

The present study focuses on the conservation of genetic resources in the threatened timber species *Dalbergia cochinchinensis* Pierre (Fabaceae: Papilionoidea), an Indochinese species also known as Siamese rosewood.

Timber from this species is in high demand in Southeast Asia due to its red colored and durable heartwood, which is primarily used for production of luxury furniture but also for musical instruments and handicrafts (Niyomdham *et al.* 1997; CTSP 2004; GW 2015). The species is endemic to Cambodia,

112 Vietnam, Laos and Thailand, and although locally used, timber is mainly exported to the Chinese
113 market for the production of “hong mu” luxury furniture (GW 2015). Commercial plantations as well
114 as conservation stands have been established (e.g. in Cambodia and Thailand) but currently no mature
115 plantations of the species exist, which means that all logging occurs in natural forests. Logging of the
116 species from natural populations is prohibited in all four countries (CoP16 2013), but illegal logging
117 and trade of *D. cochinchinensis* is a multimillion dollar criminal activity, associated with corruption
118 and border conflicts (EIA 2015; GW 2015; Siriwat & Nijman 2018). Due to continuous
119 overexploitation coupled with the high levels of deforestation in the region the species has recently
120 experienced a severe reduction in overall population size and has been listed on CITES Appendix II
121 since 2013 (CITES 2017). The future existence of *D. cochinchinensis* as a component of natural forests
122 in Indochina and as a resource for commercial breeding programs is thus severely threatened if
123 adequate conservation plans are not implemented urgently.

124 Recent population genetic studies have shown that *D. cochinchinensis* harbors relatively high levels of
125 genetic diversity, but also shows strong geographic structure and a high level of differentiation
126 indicative of limited gene flow among populations and potentially high levels of local adaptation
127 (Moritsuka *et al.* 2017; Hartvig *et al.* 2018). Moreover, *D. cochinchinensis* can reproduce asexually,
128 which can affect effective population size (So *et al.* 2010; Hartvig *et al.* 2018).

129 Here, we use results on population genetic diversity and structure in *D. cochinchinensis* across its
130 mainland Southeast Asia range to propose recommendations for the conservation of genetic diversity in
131 the species. We also explore the potential of molecular markers for timber tracking in *D.*
132 *cochinchinensis*.

For these purposes, we developed SNP markers for *D. cochinchinensis* and genotyped samples from 26 *D. cochinchinensis* populations across the four range countries and used these data in combination with existing SSR genotypes from a previous study (Hartvig *et al.* 2018).

Material and methods

Samples and study area

We used DNA samples from leaf or cambium material collected from 677 individuals between 2010 and 2014 from 26 natural populations across Cambodia, Vietnam, Thailand and Laos (Fig 1, Table 1, Table S1). Details on sampling and permissions can be found in Hartvig *et al.* (2018). The sampling covers the majority of the currently known distribution area for *D. cochinchinensis* (Niyomdham 2002; USNH 2003; CTSP 2004; van Sam *et al.* 2004, H. Gainsberger *et al.*, unpublished), with the exemptions of populations in Southern Laos, parts of Eastern and Northeastern Thailand, Central Cambodia, and parts of Southern and Central Vietnam.

The localities of the sampled populations were visualized in relation to local deforestation trends and protected areas. We used world satellite imagery (Esri) and overlaid it with Landsat images of forest extent and change from 2000 through to 2016 (Hansen *et al.* 2013; Hansen/UMD/Google/USGS/NASA 2018) as well as information on protected areas from the World Database of Protected Areas (UNEP-WCMC & IUCN 2018) and country-specific data for Cambodia (ODC 2017).

Development of SNP markers

153 We applied the RAD sequencing approach (Baird *et al.* 2008) for the purpose of SNP discovery in *D.*
 154 *cochinchinensis*. We used a panel of 10 individual samples of *D. cochinchinensis* for the sequencing,
 155 originating from the populations RUK (four samples), BAN (two samples), TAI, KIR, HAP and KRA
 156 (one sample each) (see Fig 1 for locations of the six populations). DNA samples were normalized to 50
 157 ng/ul and one microgram of each sample was digested with the SbfI restriction enzyme. The SbfI
 158 enzyme was chosen due to an assumed genome size around $1C=1.88$ pg, equaling app. 1.8 Gbp, based
 159 on genome size estimation in closely related *Dalbergia* species (Hiremath & Nagasampige 2004;
 160 Hartvig *et al.* 2015) and in order to achieve high coverage per loci.

161 RAD libraries were prepared according to the methodology described in Etter *et al.* (2011) with
 162 modifications as described in Houston *et al.* (2012). Barcoded P1 adapters were ligated to the SbfI site
 163 of the fragmented DNA before shearing to the range 300-700bp using a Covaris E220 sonicator
 164 (Covaris Ltd. Brighton UK). The sheared DNA was size selected using a BluePippin 1.5% DF Marker
 165 R2 Gel (SAGE Science Inc. Beverly, MA) to a size range of 250 – 700 bp to clear any larger size
 166 products. The size selected DNA was blunt ended and A-tailed for the ligation of a second adapter
 167 (P2). Following the second ligation the DNA was purified using Ampure beads (Beckman Coulter,
 168 High Wycombe, UK) before extending and enriching the library by PCR. Ampure beads were used to
 169 purify the PCR products and the final library was assessed using D1000 Screentape (Agilent,
 170 Edinburgh, UK) and quantified using the Kapa Library Quant kit (Roche, Welwyn Garden City). The
 171 library was sequenced on a single lane on an Illumina HiSeq 2000 instrument using 150 base paired-
 172 ends reads and later re-sequenced on a further two lanes. The library preparation and sequencing were
 173 conducted at Edinburgh Genomics, University of Edinburgh. Resulting sequence data were examined
 174 and filtered for sequence quality, trimmed to remove primers and collated by barcode.

175 Bioinformatic analyses

176 The raw data was demultiplexed using the process_radtags program from STACKS 1.20 (Catchen *et al.*
 177 2013) and each sample was then processed using the PALEOMIX pipeline v1.2.13.4 (Schubert *et al.*
 178 2014). Briefly, we used AdapterRemoval v2.1.7 (Schubert *et al.* 2016) to remove adapters, trim low
 179 quality (BaseQ < 5) and missing bases, as well as exclude reads with more than 50bp of missing data or
 180 length shorter than 30bp. Since there is no reference genome for *Dalbergia*, we constructed a pseudo-
 181 reference sequence from the clean reads of all samples. To reduce the computational effort, only the
 182 first read from each pair was used. We iteratively clustered all reads with VSEARCH v2.8.0 (Rognes *et*
 183 *al.* 2016) (--fasta_width 0 --minseqlength 50 --strand both --fastq_maxns 30 --id 0.98 --sizein --
 184 sizeout), until no new clusters were formed. We used the consensus sequence of each cluster (flanked
 185 by 150 bps of Ns) as a pseudo-reference sequence, excluding those with less than 50 bps or made from
 186 less than 5 sequences. Each sample was then mapped against this pseudo-reference using bwa-mem
 187 v0.7.15 (Li *et al.* 2009), and GATK v3.3.0 (DePristo *et al.* 2011) used to perform a local realignment
 188 step around indels.

189 With the mapped reads, we then used ANGSD v0.929 (Korneliussen *et al.* 2014) to call SNPs and
 190 genotypes (-GL 2 -doGlf 2 -doMajorMinor 1 -doMaf 1 -SNP_pval 1e-4 -doPost 2 -doGeno 3 -
 191 geno_minDepth 5 -postCutoff 0.8), applying some standard quality filters (-baq 1 -C 50 -minMapQ 15
 192 -minQ 10 -rmTriallelic 1e-4), empirically defined depth cutoffs (-setMinDepth N_IND*2 -
 193 setMaxDepth N_IND*50), and keeping only positions with at most 30% missing data. Lastly, we
 194 selected all loci with a single bi-allelic SNP, with at least one heterozygous genotype (since lack of
 195 heterozygous genotypes can be indicative of genome duplication), and where the SNP was located at a

196 minimum of 50 bps from either end of the read (to make sure it could be tested with the SinglePlex
197 assay). This resulted in a panel of 43 SNP loci.

198 SNP genotyping

199 Kompetitive Allele Specific PCR (KASP [™]) assays were designed based on information on flanking
200 sequences for the 43 selected loci and used to genotype 659 individual samples, as 18 of the 677
201 original samples did not contain sufficient amount of DNA for genotyping. LGC Genomics
202 (Hertfordshire, UK) performed both assay design and genotyping. Eleven of the 43 loci failed LGC
203 quality control. We manually checked all genotype calls for the remaining 32 loci in SNPviewer (LGC
204 Ltd, 2014) and discarded loci that were either effectively monomorphic or failed to produce clear
205 clusters. This resulted in a final dataset of 19 SNP loci.

206 Data analysis

207 We combined the obtained 19 SNP loci with the dataset of 9 SSRs generated in Hartvig *et al.* (2018),
208 retaining only individuals with data from both kind of markers, and with a maximum of six loci
209 missing data (21%). This resulted in a total dataset of 643 individuals. Because of the missing locus
210 criterion, the final set of individuals used in the present study is not identical to the one in Hartvig *et al.*
211 (2018).

212 Clonality

213 Due to known presence of redundant genotypes in the SSR dataset (Hartvig *et al.* 2018) we calculated
214 the Probability of Identity (PI) for both the combined dataset as well as for separate SSR and SNP
215 datasets, using GenAlEx 6.501 (Peakall & Smouse 2006, 2012). The population level PI values for the

216 SNP data ranged between 10^{-2} to 10^{-4} , for the SSR dataset between 10^{-3} to 10^{-9} , and for the combined
 217 dataset between 10^{-6} to 10^{-13} . We concluded that identical genotypes observed in the combined dataset
 218 were representative of clonal individuals and used this dataset to identify unique genotypes with the
 219 “Matches” function in GenAlEx. We manually checked whether near-matches in fact represented true
 220 matches masked by missing data in one of the samples, or genotyping errors, and corrected data if that
 221 was the case. Summary statistics for the SNP dataset for 643 individuals were then computed in
 222 GenAlEx.

223 In order to explore the spatial distribution of genotype matches, we compared genotype matches
 224 between individuals with distance classes (0-15 m, 15-25 m, 25-35 m, 45-55 m, 55-65 m, 75-85 m and
 225 85-100 m) and then for each distance class calculated the frequency of clonality as the number of pairs
 226 of trees with identical genotypes as a percentage of all pairs of trees in that distance class. For all
 227 further analysis of the genetic data, duplicate genotypes were removed, reducing the dataset to 500
 228 genotypes (Table 1).

229 Genetic diversity

230 For estimation of genetic diversity levels, we conducted analysis on SSR and SNP datasets separately,
 231 including all 500 individuals. We used GenAlEx to calculate population means of the following
 232 measures: observed heterozygosity (H_O), unbiased expected heterozygosity ($U H_e$) (Peakall & Smouse
 233 2012), the Fixation index (F_{IS}), and for the SNP data also percentage of polymorphic loci (P). These
 234 measures were also summed over populations for the 19 SNP loci, as well as number of alleles (N_A)
 235 and number of effective alleles (N_E). The pegas package in R3.4.3 (Paradis 2010) was used to calculate

236 exact probabilities of HW proportions of the SNP markers across populations as well as within each
 237 population, and results evaluated against a table-wide alpha level of 0.05.

238 Genetic differentiation and clustering

239 A Discriminant Analysis of Principal Coordinates (DAPC) (Jombart *et al.* 2010) as implemented in the
 240 adegenet package in R3.4.3 (Jombart & Ahmed 2011) was used to describe population structure. We
 241 used the combined dataset and populations as units for the discriminant analysis, retaining 50 principal
 242 components in the first step and all 25 discriminant functions. A scatterplot of the first two principal
 243 component axes was then produced to examine population structure.

244 We also performed a genetic clustering analysis with STRUCTURE (Pritchard *et al.* 2000) on the
 245 SSR+SNP dataset. We used the admixture model and the correlated allele frequencies model, without
 246 prior population information. The most likely number of genetic clusters K was found by estimating the
 247 likelihood for each K from 1-27, with 10 iterations per K . We used a burnin of 100,000 MCMC and a
 248 simulation run of 1,000,000 MCMC. The result data were analyzed using the StructureHarvester web
 249 programme (Earl & Vonholdt 2012) to find the average likelihood value of each K , as well as
 250 estimating ΔK for identifying K (Evanno *et al.* 2005). The software CLUMPP (Jakobsson & Rosenberg
 251 2007) was used to average the outcome over the 10 iterations with the identified optimal number of
 252 clusters.

253 We used GenAlEx to perform an AMOVA (Weir & Cockerham 1984; Excoffier *et al.* 1992) with 1000
 254 permutations to analyze the partitioning of genetic variation at different hierarchical levels; within
 255 populations, among populations and among regions. The clusters identified by discriminant analysis
 256 and STRUCTURE defined the regions (Northern Laos, Central Laos & Northeast Thailand, Dangrek

257 Mountain area, Cardamom Mountains area, East Cambodia & Vietnam) and the analysis was
 258 conducted separately for SSR and SNP datasets. We also calculated U_{H_E} and number of private alleles
 259 per region for SSR loci and U_{H_E} and percentage of polymorphic loci per region for SNP loci.

260 To test for the effect of geographical distance on the levels of genetic differentiation among
 261 populations, we conducted paired Mantel tests (Mantel 1967), as implemented in GenAlEx, on SNP
 262 and SSR sets separately. We used matrices of pairwise geographical distances and pairwise F_{ST} values,
 263 and tested the significance with 999 permutations.

264 Assignment tests

265 As a preliminary assessment of the timber tracking potential of the genetic markers we performed self-
 266 assignment tests (which includes leave-one-out) in the program GeneClass (Piry *et al.* 2004). The aim
 267 was to test whether the sampled trees could be allocated to their true geographic origin (among the
 268 sampled geographic origins) based only on their marker genotype. The self-assignment ability was
 269 computed at both population and region level (same regions as used in the AMOVA) and for combined
 270 SSR+SNP data as well as for separate datasets. We used the Rannala & Moutain (1997) Bayesian
 271 method and enabled Monte-Carlo resampling with 1000 simulated individuals.

272

273 Results

274 SNP loci

275 The RAD sequencing generated 191,519,491 reads, ranging from 10,744,182 to 34,099,267 reads per
 276 sample. Initial filtering with the PALEOMIX pipeline retained 191,051,623 reads (10711254 to
 277 34043825 reads per samples). Of these, 144,257,495 reads (6,056,181 to 27,902,377 reads per sample)
 278 mapped to the generated pseudo-reference and resulted in the identification of 23,647 putative SNP
 279 loci. Selecting tags with only one SNP reduced the number to 428 and additional filtering resulted in a
 280 panel of 43 SNPs. Of these, 19 provided interpretable genotypes after genotyping with the KASP assay.
 281 Missing data across the 659 samples genotyped with KASP was 3% and were mainly due to a few very
 282 low-quality samples. After combination with the SSR data, 643 samples were kept for further analyses.
 283 Summary statistics for the 19 SNP loci across the 643 samples are found in Table S2. There was no
 284 indication of deviations from HW proportions within populations, however all but two SNPs were
 285 fixed for one allele in one or more populations. The level of polymorphism varied from 15-100%
 286 among loci, and from 32-79% among populations.

287 Clonal distribution

288 An analysis of the spatial distribution of genotype matches for all 643 genotypes revealed that close
 289 trees often were results of clonal reproduction (i.e. same genotypes for both SNP and SSR markers). A
 290 large fraction (41%) of the pairs of trees sampled closer than 15 meters from each other thus had
 291 identical genotypes, and matching genotypes occurred among 24% of all trees separated up to 45
 292 meters. Above 45 meters, all pairs of trees had different genotypes (Fig 2).

293 The level of clonality varied greatly among populations as several populations had none or only a few
 294 duplicate genotypes, while in other populations a high proportion of duplicate genotypes were detected
 295 (e.g. RUK, ANG, RO, KRA, DP; Table 1). The proportion of closely sampled tree pairs was however

296 not equal among sites (as the density of accessible individuals varied greatly), which may affect
 297 comparisons among populations.

298 Genetic diversity

299 The level of genetic diversity (excluding clonal individuals) as measured by mean unbiased expected
 300 heterozygosity U_{HE} , was 0.57 for SSR markers. However, the U_{HE} level varied from around 0.4 in the
 301 most northern populations in Laos and Thailand (NIA to PY) to above 0.6 in Cardamom Mountains
 302 area in southwest Cambodia (RO, KRA, SPEU) and above 0.7 in the central part of the sampling area,
 303 the Dangrek Mountains area (POS, ANG, RUK, PYA) (Table 1, Fig 1). The SNP markers reported an
 304 average U_{HE} of 0.18, with variation from 0.09 (TAI and UDO), to values above 0.2 for a few
 305 populations from Dangrek Mountains and Cardamom Mountains. The variation in heterozygosity level
 306 largely followed the variation in proportion of polymorphic SNPs within each population, and on
 307 average only around 50% of the SNP markers were polymorphic within any individual population
 308 (Table 1).

309 There was only a weak correlation between observed heterozygosity at SSR and SNP markers ($r^2=0.17$,
 310 $p<0.05$), but a slightly stronger relationship when unbiased heterozygosity was compared ($r^2=0.35$,
 311 $p<0.01$) (Fig S1). The comparison revealed a pattern where SSR and SNP diversity values correlated
 312 best when comparing populations with high levels of diversity, while for the populations with lower
 313 diversity levels, the deviation between estimated diversity levels based on SNP and SSR markers was
 314 larger (Table 1, Figs 3 + S1).

315 Genetic differentiation

316 The DAPC and STRUCTURE analyses revealed clear geographic structure in the data, and both
 317 marker types expressed evidence for isolation by distance (Figs 4+5, S2). The strongest level of
 318 structuring was observed between a northern cluster (Laos+ Northeast Thailand) and a southern cluster
 319 (Vietnam, Cambodia and the PJ Thai populations) (Figs 4+5, S3). A clear peak for $K=5$ in the
 320 STRUCTURE results supported further substructure in the data into five geographical regions, which
 321 could also be recognized in the DAPC analyses: North Laos (NIA to BAN), Central Laos & Northeast
 322 Thailand (TAI to PY), Dangrek Mountains area (PJ to PYA), Cardamom Mountains area (RO to DNT)
 323 and East Cambodia and Vietnam (HAP to YOD) (Figs 1,4+5, S3).

324 We used these five regions to further explore the levels of genetic diversity and structure in *D.*
 325 *cochinchinensis*. The five regions showed different levels of genetic diversity, with the highest level
 326 found in the Dangrek Mountains area, followed by the Cardamom Mountains area, East Cambodia &
 327 Vietnam, and the lowest diversity levels found in North Laos and Central Laos & Northeast Thailand
 328 (Table 2, Fig 3).

329 This tendency was also true at a population level for the SSR data, while the SNP U_{H_E} values showed a
 330 higher degree of variation among populations, especially within the Laos & Northeast Thailand region.
 331 In this region, the SNP diversity level varied greatly, nearly covering the total span in U_{H_E} levels (Fig
 332 3).

333 An AMOVA showed a very high level of genetic differentiation with overall $\Phi_{PT}=0.41$ for SSR and
 334 $\Phi_{PT}=0.52$ for SNP. Variation among the five regions were three times higher than the level of
 335 differentiation among populations for the SNP markers ($\Phi_{RT}=0.44$ vs $\Phi_{PR}=0.15$), while the SSR

336 markers showed a more equal level of differentiation among regions and populations ($\Phi_{RT} = 0.27$ and
 337 $\Phi_{PR} = 0.19$) (Table 3).

338 Assignment tests

339 The molecular markers assigned in most cases individuals correctly to geographic origin when
 340 applying a leave-one-out (LOO) assignment test. When using the combined dataset, 94 % of all
 341 individuals were assigned to the correct region (North Laos, Central Laos & Northeast Thailand,
 342 Dangrek Mountains, Cardamom Mountains or East Cambodia & Vietnam). This rate was slightly lower
 343 when using the SSR data alone (89%) and considerably lower when only applying SNP data (80%).

344 Assignment at the population level resulted in a correct identification rate of 80 % for the combined
 345 dataset and 74% for the SSR data, while the SNP markers only assigned 41% of individuals back to the
 346 correct population.

347 Discussion

348 The conservation status of *D. cochinchinensis* has decreased over the last decades, and the continued
 349 protection of the species throughout its natural distribution area is of crucial importance and requires
 350 urgent action. While acknowledging the limitations of the data used in this study, we consider the data
 351 on genetic diversity and structure to be by far the best available data on which to make informed
 352 decisions on best strategies for the conservation of genetic diversity in *D. cochinchinensis*.

353 Limitations of molecular markers used in this study

354 The number of usable SNP loci identified through RAD sequencing was very low and did not meet our
 355 expectations. Even though the initial number of SNPs was comparable to the number of loci predicted

for other tree species given the enzyme of choice and expected genome size (Hiremath & Nagasampige 2004; Parchman *et al.* 2018), the use of a rare cutting enzyme together with the small sample size used for SNP discovery (10 samples) might have hindered our ability to find variable positions. Either way, the loss of loci during subsequent filtering was surprisingly high, with 98% of the SNPs lost due to filtering for multiple SNPs on the same RAD locus. This could be due to extensive presence of duplicated regions leading to mis-assembly of paralogous loci and a high rate of false positive SNP discovery (Hodel *et al.* 2016; Ravindran *et al.* 2018). The experienced high dropout after filtering for sequences with only 1 SNP could be indicative of this scenario. More than 1 SNP per 150 bp would not be expected to occur frequently under normal conditions, and two or more SNPs per tag could be an artifact produced by assembly of paralogous loci. In line with this, a RADseq study of the legume *Robinia pseudoacacia* found high levels of paralogy, which had a strong impact on number of SNPs detected (Verdu *et al.* 2016). The authors related it to the ancestral genome duplications in the Fabaceae (Cannon *et al.* 2015), which could be of impact in this study as well. With the experienced high SNP dropout rate during quality filtering, it would have been beneficial to use a more frequently cutting restriction enzyme to achieve a better starting point in terms of number of loci and potential SNPs identified in the first step.

In the final steps of filtering, the use of a KASP genotyping assay also reduced number of SNPs roughly by a third, as it requires a minimum of 50 bp either side of the SNPs for proper primer development. An assay less stringent in requirements for size of flanking regions would have increased the number of usable SNPs. An exhaustive exploration on the performance of the RAD sequencing is though considered beyond the scope of this paper, and it will therefore not be dealt with further.

377 SSR markers are generally assumed selectively neutral and to detect mainly the demographic history of
 378 the species and current processes such as gene flow, genetic drift and inbreeding, while the SNP
 379 markers might be located in genomic regions under selection and affected by different selection
 380 pressures due to differing biotic and abiotic conditions in the localities (Ouborg *et al.* 2010). While the
 381 number of SNP markers in this study is far too low for detection of patterns of local adaptation or
 382 selection, the discrepancy in population level diversity patterns between SSR data and SNP data (Table
 383 1, Fig 3) could indicate that they are affected differently by external factors. The variation in SNP
 384 diversity at the population level is associated with variation in the percentage of polymorphic loci and
 385 is due to fixation of alleles in the more low-diversity populations. Whether this is a result of random
 386 processes or different selection pressures among populations is however not possible to tell.

387 The combined data produced by SNP and SSR markers in this study is thus best seen as an assessment
 388 of the neutral genetic variation in *D. cochinchinensis* and further research to support conservation of
 389 this species should focus on detecting adaptive genetic variation.

390 Priority areas for conservation of genetic resources in *D. cochinchinensis*

391 *Dalbergia cochinchinensis* exhibit a very strong spatial structure in the distribution of genetic diversity,
 392 high levels of isolation by distance and varying levels of genetic diversity among five well-separated
 393 genetic clusters (Figs 3+4, S2, Table 2). This is possibly induced by limited seed and pollen dispersal,
 394 and could be related to ancient as well as current landscape barriers in the Indochinese landscape, such
 395 as the Mekong River and Tonle Sap Lake (Hartvig *et al.* 2018). For further exploration of potential
 396 causes underlying the observed patterns, we refer to the discussion in Hartvig *et al.* (2018), and here
 397 focus on the implications of the results for conservation and management.

398 Given the high level of genetic differentiation and correlation between genetic clusters and geography,
 399 we suggest that the five identified clusters (North Laos, Central Laos & Northeast Thailand, Dangrek
 400 Mountains, Cardamom Mountains and East Cambodia & Vietnam, Figs 1,4,5) could potentially be
 401 recognized as Management Units (MUs) (Palsboll *et al.* 2007) when designing a regional plan for
 402 conservation of the genetic diversity in the species. The parts of the occurrence area for *D.*
 403 *cochinchinensis* not covered in this study, most significantly Southern Laos and Eastern Thailand,
 404 should be target to supplementary analyses, as to evaluate whether populations in these areas might
 405 represent additional genetically distinct clusters, or could be included in the units suggested above.

406 The MUs can form the geographical areas from which more detailed *in situ* and *ex situ* conservation
 407 strategies can be implemented to conserve genetic diversity both within and among clusters and ensure
 408 sufficient genetic resources for breeding programs.

409 An ideal *in situ* conservation strategy for *D. cochinchinensis* should target a network of selected
 410 populations from each of the identified MUs for strict protection and monitoring. Populations in the
 411 areas around Dangrek Mountains and Cardamom Mountains should receive high priority as
 412 conservation stands, since the highest levels of diversity appear to be in these areas. These areas are
 413 also among the most threatened in terms of deforestation and illegal logging (Fig 1, (GW 2015)), and it
 414 is probable that several of the sampled populations have been extirpated since time of collection in
 415 2010-2012. It is thus a matter of urgency to increase protection levels for selected stands in these areas.

416 The peripheral clusters in Vietnam and Northeast Thailand & Laos should also be targeted for specific
 417 conservation management, since these evidently harbor distinct genetic variation. Fady *et al.* (2016)
 418 argue that as geographically peripheral populations may possess unique adaptations to the extreme

ecological conditions within the distribution range of the species, they are likely to have a disproportionately high influence on a forest tree species' ability to adapt to future climatic conditions and should be of special interest for conservation. Populations not covered in this study, such as in Southern Laos, East/Northeast Thailand, Central Cambodia and Vietnam should also be included in the network to maximize geographic spread and hence genetic diversity, given the strong pattern of differentiation.

The network of *in situ* conservation stands could simultaneously be combined with collection of seeds for the establishment of *ex situ* conservation stands, which could enhance the likelihood of conservation success. The natural populations of *D. cochinchinensis* are under a great threat especially from illegal logging targeted directly at mature trees, and adequate protection have proven difficult even within protected areas. *Ex situ* stands thus seem a crucial component in effective conservation of *D. cochinchinensis* and such stands could eventually act as local seed sources for both restoration and breeding purposes. There is a growing demand for seed supplies in the range countries, for example in Cambodia, *D. cochinchinensis* was the second-most frequently planted tree species in 2015 (Institute of Forest and Wildlife Research and Development, FA, Cambodia, unpublished report). The combination of highly valuable timber and high growth rate in young stages make *D. cochinchinensis* a good candidate for a successful plantation species (So, unpublished data and So *et al.* 2010). Given the high level of genetic differentiation and the large ecological variation present within the natural distribution area, it is recommendable to base planting programs on tree planting zones each served by a set of local seed sources (Kjaer *et al.* 2001; Amaral *et al.* 2004b).

The Breeding Seedling Seed Orchard (BSO) concept suggested first by Barnes (1995) can be implemented by collection of seeds from a number (>50) of different healthy trees ('plus trees') in a

441 number of stands within a planting zone, planting of the offspring in a test trial, and - after evaluation
 442 and thinning within families – the use of the remaining trees as a seed source for plantation programs.
 443 The BSO activities involve directional selection that over time favors healthy and straight trees with
 444 vigorous juvenile growth under planting conditions. *D. cochinchinensis* is a pioneer species in its native
 445 habitats (So 2000; So *et al.* 2010) where rapid juvenile growth under non-shaded conditions is likely to
 446 contribute to fitness then colonizing new canopy openings. The artificial and natural selection in the
 447 multiple BSOs are therefore not a severe concern in relation to the objectives of dynamic gene
 448 conservation, although the *ex situ* conservation in BSOs should by no means replace *in situ*
 449 conservation efforts. The most import aspect from a conservation point of view is that high levels of
 450 effective population sizes (genetic diversity) are maintained at all stages. This can be obtained by
 451 basing the program on initial selection of a large number of mother trees, or – when number of founder
 452 trees are limited – direct selection towards within families (Dhakal *et al.* 2005). In this context the BSO
 453 concept seems especially suitable for *D. cochinchinensis*, because it can combine high levels of
 454 diversity in the seed crop with the capturing of diversity among regions and continued selection in
 455 favor of local growth conditions. The integration of genetic diversity as an essential part of planting
 456 activities will also limit future concerns for gene flow from plantations into remaining *in situ*
 457 conservation areas. The size and delineation of the planting zones can be based on the revealed
 458 population structure, but the planting zones should not cover populations from very different ecological
 459 conditions since the DNA markers often fail to reveal fine scale local adaption in fitness traits
 460 (Gonzalez-Martinez *et al.* 2006).
 461 Clonality affects recommended size of conservation stands and establishment of *ex situ* stands

462 *Dalbergia cochinchinensis* can reproduce clonally by root suckers (So *et al.* 2010; Hartvig *et al.* 2018)
 463 and the analyses in this study revealed that trees as far apart as 45 meters could be ramets of the same
 464 clone. *Dalbergia cochinchinensis* is known as an intermediate pioneer species (So 2000; So *et al.* 2010)
 465 able to establish dense stands in forest clearings (the authors, personal observations). This is most
 466 likely correlated to its clonal reproduction ability, which may be considered a strategy to quickly
 467 colonize new habitats, such as after fire or other disturbances (Kammesheidt 1999; Mostacedo *et al.*
 468 2009). The high frequency of clonality in some populations, such as in the DNT population, where the
 469 area had been cleared ten years prior to sampling, may be the result of rapid re-colonization by root
 470 suckers. This clonality is an important factor to consider when implementing conservation and breeding
 471 programs.

472 The size of the conservation stand is of crucial importance to avoid loss of genetic diversity.
 473 Accumulated evidence suggest that an effective population size, N_e , of at least 100 is needed to avoid
 474 inbreeding while $N_e=1000$ is recommended to maintain evolutionary potential (Frankham *et al.* 2014).
 475 The effective population size N_e will often be substantially lower than the actual number of individuals,
 476 and conversion ratio from census population size to N_e may in general be around 0.1-0.2 (Frankham *et*
 477 *al.* 2014), i.e. the actual number of mature trees ideally should be 5 to 10 times N_e to meet the
 478 conservation objectives. This may indeed be the case for *Dalbergia cochinchinensis* due to the
 479 observed levels of clonality, and selfing values of 20-30% in disturbed populations (Hartvig *et al.*
 480 2018). This points to designing large continuous conservation stands of *D. cochinchinensis*, with
 481 ideally at least 500-1000 mature trees, such as the stands in Kampong Speu (500 mature trees) and Sre
 482 Noy (more than 1000 trees) in Southwest and Central Cambodia, respectively. However, such large
 483 populations are rare, and although large stands should be prioritized for conservation, it is still

484 important to take actions to conserve populations across different geographical areas and in threatened
485 areas, despite small population sizes.

486 To ensure sufficient genetic diversity in material used for breeding, the seed sources for each planting
487 zone should maintain a suitable effective population number and the BSO concept recommends
488 collection from at least 50 trees to establish the seed source. To avoid sampling clones, it is
489 recommendable to only use seeds from trees more than 50 meters from each other (Fig 2).

490 Potential of efficient timber tracking with SNP and SSR markers in *D. cochinchinensis*

491 The developed SSR and SNP markers show a significant potential for use in verification of origin of
492 traded timber samples when implementing CITES regulations for *D. cochinchinensis*. The assignment
493 rates found here were higher than found for other tropical trees (Jolivet & Degen 2012; Degen *et al.*
494 2013; Blanc-Jolivet *et al.* 2017), but could be overestimated due to the assignment method and lack of
495 blind samples. For final verification of the geographic tracking ability of the markers, the method
496 should be applied to timber samples with other origin than the present collection of samples, which
497 would likely decrease assignment rates if the population of origin is not covered in the reference data.
498 In a real custom seizure situation though, assignment rates may be affected positively by two factors.
499 Likely a batch of timber samples rather than a single sample are available for testing, and the question
500 is rather to confirm or reject a stated geographic origin instead of matching genotypes of uncertain
501 origin to all possible reference populations. Both factors will likely increase the assignment
502 performance of the molecular markers (Jolivet & Degen 2012; Degen *et al.* 2013), but with the
503 completeness of the reference database being a crucial factor for the success rate.

504 The slightly lower power of the SNP data to detect spatial resolution of the sampled populations is
 505 most probably due to the low number of SNP loci ($n=19$). The number of bi-allelic SNPs required to
 506 match the power of multiallelic SSR markers is significantly higher, up to an order of magnitude, with
 507 several thousand SNPs required to accurately detect population structure in some plant species,
 508 (Fischer *et al.* 2017). The remarkably high resolution at a regional level for the 19 SNP markers in this
 509 study can be attributed to the high level of spatial genetic structure in the species.

510 For implementation in timber trade control, it would be more practical to choose only one type of
 511 marker. SNPs are considered very feasible for timber tracking since marker development can be very
 512 cost-effective and genotyping can be performed with high-plex, automated assays allowing for easier
 513 standardization among laboratories (Blanc-Jolivet & Liesebach 2015; Jardine *et al.* 2016; Pakull *et al.*
 514 2016; Blanc-Jolivet *et al.* 2017). This contrasts to the fragment length calling of SSR markers, where
 515 use of calibrated standards is needed in order to guarantee similar scoring among laboratories. The
 516 benefit of SSR markers is their high levels of polymorphism resulting in high precision rate with a low
 517 number of markers. While the current markers are immediately applicable for single cases of timber
 518 tracking and have been used in a case of illegal logging from a protected area in Thailand, development
 519 of a large number of SNP markers would be preferable for a large-scale efficient timber tracking
 520 system for *D. cochinchinensis*.

521 Governance of protected areas

522 State-controlled protected areas (PAs) have dominated conservation strategies globally, yet their
 523 performance relative to other governance regimes is rarely assessed comprehensively (Schleicher *et al.*
 524 2017). The populations of *D. cochinchinensis* in Thailand and Vietnam are exclusively located within

525 well-protected national parks and protected areas (Fig 1), although targeted illegal logging is difficult
 526 to avoid even within these areas. A more official designation of these populations as conservation
 527 stands and sources for breeding and re-introduction programs could help ensure their continued
 528 protection. In Cambodia, Royal Decrees have been issued to give conservation stands additional
 529 protection on top of the existing legislation and such initiatives may heighten the awareness of local
 530 authorities and park rangers.

531 Several studies have shown that forests are equally well protected on indigenous territories and lands
 532 managed by local communities as they are in national parks (Nepstad *et al.* 2006; Bray *et al.* 2008;
 533 Ellis & Porter-Bolland 2008; Porter-Bolland *et al.* 2012). In Cambodia, local communities have
 534 protected many of the larger, remaining stands of *D. cochinchinensis* in the form of community forests.
 535 This is the case for stands in Dang Peng (Koh Kong), Kampong Speu and Rukkavan (Otdor
 536 Meanchey). This observation adds to the growing body of research showing that local forest
 537 governance can be equally or more effective than centralized state regimes and that local conservation
 538 initiatives deserve more political, financial and legal support than they currently receive (Schleicher *et*
 539 *al.* 2017). On this background, we recommend that future *in situ* conservation efforts on *D.*
 540 *cochinchinensis* are directed towards community owned and managed lands and local communities be
 541 consulted and involved to a much larger extent in the practical management and conservation of the
 542 species.

543 Conclusion

544 The natural populations of *D. cochinchinensis* are disappearing fast, and a combination of strategically
 545 chosen *in situ* and *ex situ* conservation stands and well-designed breeding programs, is urgently needed

in order to conserve the genetic resources within the species. Recognizing that illegal logging is very difficult to combat, high priority should be given to establishment of living germplasm collections from populations across the distribution area. Given the high level of spatial structure detected with the developed markers, *D. cochinchinensis* however seems to be the ideal case for developing a molecular timber tracking system to limit the trade of illegally logged timber.

Further research should focus on assessing variation in adaptive traits in the species, which could help target conservation efforts and guide breeding programs, but also on the ability of clonal reproduction, as it might prove an important trait for natural populations to recover from illegal logging and for vegetative propagation of germplasm material needed for planting or restoration programs.

Acknowledgements

Funding: This work was supported by DANIDA grant no 11073-LIFE and IH was funded by a PhD scholarship from University of Copenhagen.

Author contributions

Study concept and design: IH ST SC HTT SB RO HS IT LRN EDK. RAD sequencing design: IH RO FT RT. Bioinformatics: HS FGV. SNP genotyping: IH. Analysis and interpretation of data: IH EDK LRN. Drafting of manuscript: IH IT LRN EDK Critical revision: IH ST SC HTT SB RO HS FGV FT RT IT LRN EDK.

Data availability

564 A FASTA file with the 43 SNPs and flanking sequences and the dataset with 500 SNPs+SSR
 565 genotypes are available from the University of Copenhagen Electronic Research Data Archive at
 566 <https://sid.erda.dk/public/archives/e701584e549fe59472596f5f4e793959/published-archive.html> and
 567 <https://sid.erda.dk/public/archives/6c53d7edfa9195829eec09903a8fb110/published-archive.html>,
 568 respectively.

569 Supporting Information

570 Table S1 Sample and locality information for the 26 *D. cochinchinensis* populations included in this
 571 study.

572 Table S2 Summary statistics for the 19 SNP loci, summed across 26 populations of *D. cochinchinensis*.

573 Figure S1 Correlation between SSR and SNP diversity levels in 26 populations of *D. cochinchinensis*.

574 Figure S2 Isolation by distance among 26 *D. cochinchinensis* populations.

575 Figure S3 DeltaK values (Evanno *et al.* 2005) for STRUCTURE runs from K=1 to K=27 for SSR+SNP
 576 genotypes of 26 *D. cochinchinensis* populations.

577 References

- 578 Amaral W, Thomson L, Yanchuk A (2004a) Conservation of genetic resources in their natural
 579 environment. In: *Forest genetic resources conservation and management. Overview, concepts*
 580 *and some systematic approaches* (ed. FAO/FLD/IPGRI), pp. 1-4. International Plant Genetic
 581 resources Institute, Rome.
- 582 Amaral W, Yanshuk A, Kjaer ED (2004b) Methodologies for ex situ conservation. In: *Forest Genetic*
 583 *Resources: Conservation and management. Vol 3: In plantations and gene banks (ex situ)* (ed.
 584 FAO/FLD/IPGRI), pp. p. 3-8. International Plant Genetic Resources Institute, Rome, Italy.
- 585 Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP Discovery and Genetic Mapping Using
 586 Sequenced RAD Markers. *Plos One* **3**. doi:10.1371/journal.pone.0003376
- 587 Barnes R (1995) The breeding seedling orchard in the multiple population breeding strategy. *Silvae*
 588 *Genetica* **44**, 81-88

- Blanc-Jolivet C, Kersten B, Dainou K, *et al.* (2017) Development of nuclear SNP markers for genetic tracking of Iroko, *Milicia excelsa* and *Milicia regia*. *Conservation Genetics Resources* **9**, 531-533. doi:10.1007/s12686-017-0716-2
- Blanc-Jolivet C, Liesebach M (2015) Tracing the origin and species identity of *Quercus robur* and *Quercus petraea* in Europe: a review. *Silvae Genetica* **64**, 182-193. doi:10.1515/sg-2015-0017
- Bray DB, Duran E, Ramos VH, *et al.* (2008) Tropical Deforestation, Community Forests, and Protected Areas in the Maya Forest. *Ecology and Society* **13**, 56
- Cannon SB, McKain MR, Harkess A, *et al.* (2015) Multiple Polyploidy Events in the Early Radiation of Nodulating and Nonnodulating Legumes. *Molecular Biology and Evolution* **32**, 193-210. doi:10.1093/molbev/msu296
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140. doi:10.1111/mec.12354
- CITES (2017) *Appendices I, II and III*. <https://cites.org/eng/app/appendices.php>
- CoP16 C (2013) *CoP16 Prop. 60. Proposal for amendment of appendices I and II. Dalbergia cochinchinensis*. <http://www.cites.org/sites/default/files/eng/cop/16/prop/E-CoP16-Prop-60.pdf>
- CTSP (2004) *Cambodian Tree Species. Monographs*. Cambodia Tree Seed Project, FA, Cambodia and DANIDA, Denmark.
- Degen B, Blanc L, Caron H, *et al.* (2006) Impact of selective logging on genetic composition and demographic structure of four tropical tree species. *Biological Conservation* **131**, 386-401. doi:10.1016/j.biocon.2006.02.014
- Degen B, Ward SE, Lemes MR, *et al.* (2013) Verifying the geographic origin of mahogany (*Swietenia macrophylla* King) with DNA-fingerprints. *Forensic Science International-Genetics* **7**, 55-62. doi:10.1016/j.fsigen.2012.06.003
- DePristo MA, Banks E, Poplin R, *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491-498. doi:10.1038/ng.806
- Dhakal LP, Lillesø JPB, Kjær ED, Jha PK, Aryal HL (2005) *Seed sources of agroforestry trees in a farmland context - a guide to tree seed source establishment in Nepal*, Hørsholm, Denmark.
- Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361. doi:10.1007/s12686-011-9548-7
- EIA (2015) *Addressing ASEAN's regional rosewood crisis: An urgent call to action* Environmental Investigation Agency, London, UK.
- Ellis EA, Porter-Bolland L (2008) Is community-based forest management more effective than protected areas? A comparison of land use/land cover change in two neighboring study areas of the Central Yucatan Peninsula, Mexico. *Forest Ecology and Management* **256**, 1971-1983. doi:10.1016/j.foreco.2008.07.036
- Eriksson G, Namkoong G, Roberds JH (1993) Dynamic Gene Conservation for Uncertain Futures. *Forest Ecology and Management* **62**, 15-37. doi:10.1016/0378-1127(93)90039-P
- Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011) Local De Novo Assembly of RAD Paired-End Contigs Using Short Sequencing Reads. *Plos One* **6**. doi:10.1371/journal.pone.0018561
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x

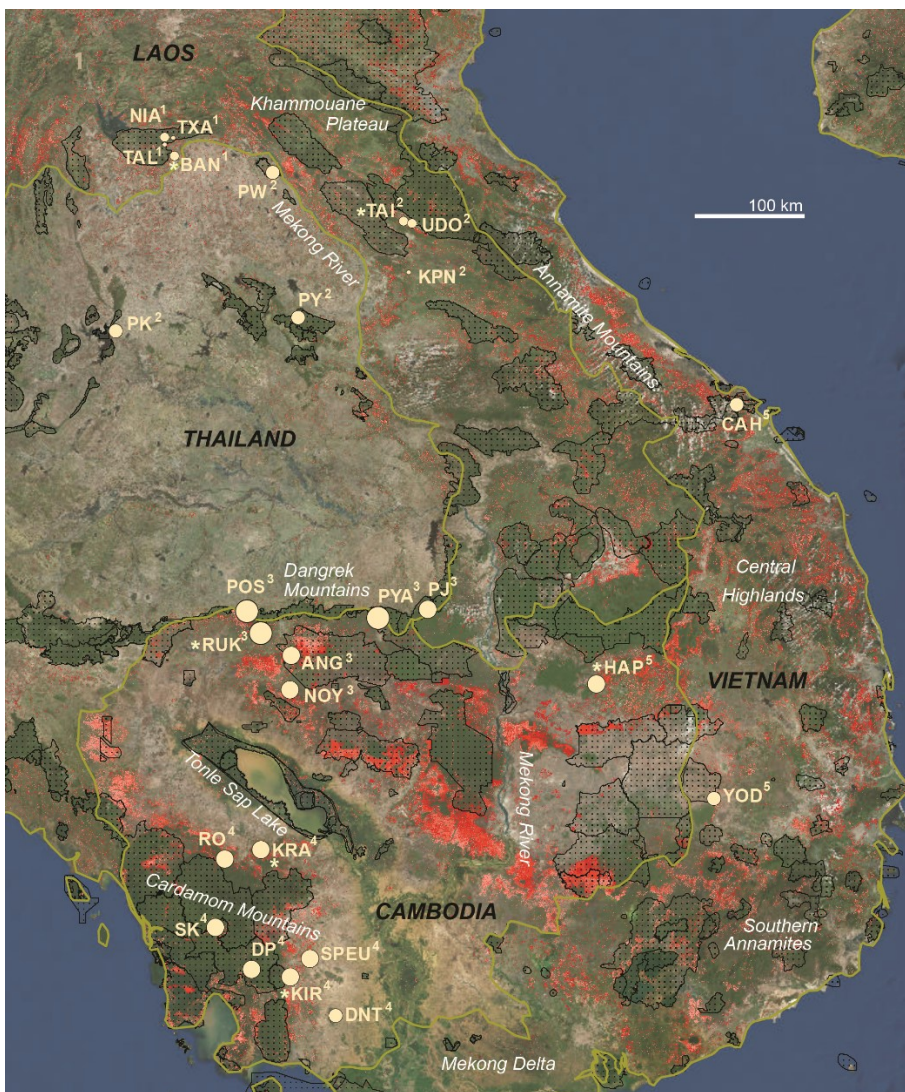
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes - Application to Human Mitochondrial-DNA Restriction Data. *Genetics* **131**, 479-491
- Fady B, Aravanopoulos FA, Alizoti P, *et al.* (2016) Evolution-based approach needed for the conservation and silviculture of peripheral forest tree populations. *Forest Ecology and Management* **375**, 66-75. doi:10.1016/j.foreco.2016.05.015
- FAO (2014) *State of the World's Forest Genetic Resources*. Food and Agriculture Organization of the United Nations, Rome, Italy.
- FAO (2016) *State of the World's Forests 2016. Forest and agriculture: land-use challenges and opportunities*. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Fischer MC, Rellstab C, Leuzinger M, *et al.* (2017) Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *Bmc Genomics* **18**. doi:10.1186/s12864-016-3459-7
- Frankham R, Bradshaw CJA, Brook BW (2014) Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation* **170**, 56-63. doi:10.1016/j.biocon.2013.12.036
- Gibson L, Lee TM, Koh LP, *et al.* (2011) Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature* **478**, 378-381. doi:10.1038/Nature10425
- Gonzalez-Martinez SC, Krutovsky KV, Neale DB (2006) Forest-tree population genomics and adaptive evolution. *New Phytologist* **170**, 227-238. doi:10.1111/j.1469-8137.2006.01686.x
- Graudal L, Yanshuk A, Kjaer ED (2004) National Planning. In: *Forest Genetic Resources: Conservation and management. Vol 1: Overview, concepts and some systematic approaches* (ed. FAO/FLD/IPGRI), pp. 25-36. International Plant Genetic Resources Institute, Rome, Italy.
- GW (2015) *The cost of luxury: Cambodia's illegal trade in precious wood with China* Global Witness, London, UK.
- Hansen MC, Potapov PV, Moore R, *et al.* (2013) High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850-853. doi:10.1126/science.1244693
- Hansen/UMD/Google/USGS/NASA (2018) Global Forest Change 2000–2016.
- Hartvig I, Czako M, Kjaer ED, Nielsen LR, Theilade I (2015) The Use of DNA Barcoding in Identification and Conservation of Rosewood (*Dalbergia* spp.). *Plos One* **10**. doi:10.1371/journal.pone.0138231
- Hartvig I, So T, Changtragoon S, *et al.* (2018) Population genetic structure of the endemic rosewoods *Dalbergia cochinchinensis* and *D. oliveri* at a regional scale reflects the Indochinese landscape and life-history traits. *Ecology and Evolution* **8**, 530-545. doi:10.1002/ece3.3626
- Hiremath SC, Nagasampige MH (2004) Genome size variation and evolution in some species of *Dalbergia* Linn.f. (Fabaceae). *Caryologia* **57**, 367-372. doi:10.1080/00087114.2004.10589418
- Hodel RDGJ, Segovia-Salcedo MC, Landis JB, *et al.* (2016) The Report of My Death Was an Exaggeration: A Review for Researchers Using Microsatellites in the 21st Century. *Applications in Plant Sciences* **4**. doi:10.3732/apps.1600025
- Houston RD, Davey JW, Bishop SC, *et al.* (2012) Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *Bmc Genomics* **13**. doi:10.1186/1471-2164-13-244
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806. doi:10.1093/bioinformatics/btm233

- Jardine DI, Blanc-Jolivet C, Dixon RRM, *et al.* (2016) Development of SNP markers for Ayous (*Triplochiton scleroxylon* K. Schum) an economically important tree species from tropical West and Central Africa. *Conservation Genetics Resources* **8**, 129-139. doi:10.1007/s12686-016-0529-8
- Jolivet C, Degen B (2012) Use of DNA fingerprints to control the origin of sapelli timber (*Entandrophragma cylindricum*) at the forest concession level in Cameroon. *Forensic Science International-Genetics* **6**, 487-493. doi:10.1016/j.fsigen.2011.11.002
- Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070-3071. doi:10.1093/bioinformatics/btr521
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *Bmc Genetics* **11**. doi:10.1186/1471-2156-11-94
- Kammesheidt L (1999) Forest recovery by root suckers and above-ground sprouts after slash-and-burn agriculture, fire and logging in Paraguay and Venezuela. *Journal of Tropical Ecology* **15**, 143-157. doi:10.1017/S0266467499000723
- Kjaer ED, Graudal L, Nathan I (2001) Ex situ conservation of genetic resources of commercial tropical tree species. In: *In situ and ex situ conservation of commercial tropical trees* (eds. Theilges BA, Sastrapradja SD, Rimbawanto A), pp. 127-146. Gadjah Mada University, Yogyakarta, Indonesia.
- Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *Bmc Bioinformatics* **15**. doi:10.1186/s12859-014-0356-4
- Li H, Handsaker B, Wysoker A, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079. doi:10.1093/bioinformatics/btp352
- Lowe AJ, Boshier D, Ward M, Bacles CFE, Navarro C (2005) Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity* **95**, 255-273. doi:10.1038/sj.hdy.6800725
- Lowe AJ, Cross HB (2011) The Application of DNA Methods to Timber Tracking and Origin Verification. *Iawa Journal* **32**, 251-262
- Mantel N (1967) Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research* **27**, 209-220
- Moritsuka E, Chhang P, Tagane S, *et al.* (2017) Genetic variation and population structure of a threatened timber tree *Dalbergia cochinchinensis* in Cambodia. *Tree Genetics & Genomes* **13**. doi:10.1007/s11295-017-1199-8
- Mostacedo B, Putz FE, Fredericksen TS, Villca A, Palacios T (2009) Contributions of root and stump sprouts to natural regeneration of a logged tropical dry forest in Bolivia. *Forest Ecology and Management* **258**, 978-985. doi:10.1016/j.foreco.2008.09.059
- Nepstad D, Schwartzman S, Bamberger B, *et al.* (2006) Inhibition of Amazon deforestation and fire by parks and indigenous lands. *Conservation Biology* **20**, 65-73. doi:10.1111/j.1523-1739.2006.00351.x
- Newton AC, Oldfield S (2008) Red Listing the world's tree species: a review of recent progress. *Endangered Species Research* **6**, 137-147
- Niyomdham C, Hô PH, Dy Phon P, Vidal JE (1997) Leguminosae-Papilionoideae Dalbergieae. In: *Flore du Cambodge du Laos et du Viêt Nam* (ed. Morat P). Muséum National d'Histoire Naturelle, Paris, France.

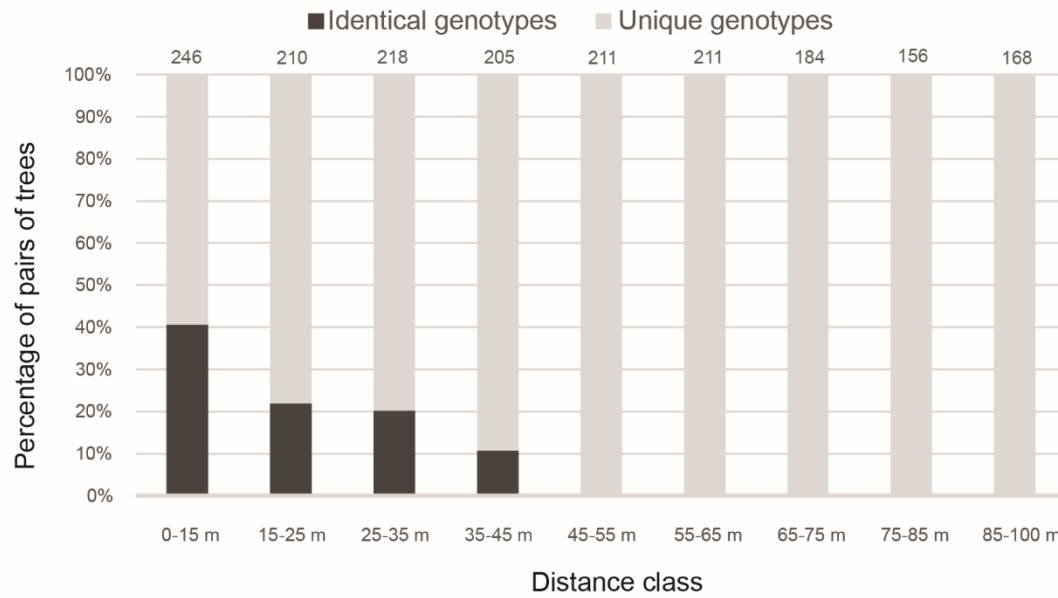
- 722 ODC (2017) Natural Protected Areas in Cambodia (1993-2017). Open Development Cambodia, Phnom
 723 Penh, Cambodia <https://opendevelopmentcambodia.net/>.
- 724 Ogden R, Dawney N, McEwing R (2009) Wildlife DNA forensics - bridging the gap between
 725 conservation genetics and law enforcements. *Endangered Species Research* **9**, 179-195
- 726 Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick PW (2010) Conservation genetics in
 727 transition to conservation genomics. *Trends in Genetics* **26**, 177-187.
 728 doi:10.1016/j.tig.2010.01.001
- 729 Pakull B, Mader M, Kersten B, *et al.* (2016) Development of nuclear, chloroplast and mitochondrial
 730 SNP markers for *Khaya* sp. *Conservation Genetics Resources* **8**, 283-297. doi:10.1007/s12686-
 731 016-0557-4
- 732 Palsboll PJ, Berube M, Allendorf FW (2007) Identification of management units using population
 733 genetic data. *Trends in Ecology & Evolution* **22**, 11-16. doi:10.1016/j.tree.2006.09.003
- 734 Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach.
 735 *Bioinformatics* **26**, 419-420. doi:10.1093/bioinformatics/btp696
- 736 Parchman TL, Jahner JP, Uckele KA, Galland LM, Eckert AJ (2018) RADseq approaches and
 737 applications for forest tree genetics. *Tree Genetics & Genomes* **14**. doi:10.1007/s11295-018-
 738 1251-3
- 739 Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for
 740 teaching and research. *Molecular Ecology Notes* **6**, 288-295. doi:10.1111/j.1471-
 741 8286.2005.01155.x
- 742 Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for
 743 teaching and research-an update. *Bioinformatics* **28**, 2537-2539.
 744 doi:10.1093/bioinformatics/bts460
- 745 Piry S, Alapetite A, Cornuet JM, *et al.* (2004) GENECLASS2: A software for genetic assignment and
 746 first-generation migrant detection. *Journal of Heredity* **95**, 536-539. doi:10.1093/jhered/esh074
- 747 Porter-Bolland L, Ellis EA, Guariguata MR, *et al.* (2012) Community managed forests and forest
 748 protected areas: An assessment of their conservation effectiveness across the tropics. *Forest*
 749 *Ecology and Management* **268**, 6-17. doi:10.1016/j.foreco.2011.05.034
- 750 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus
 751 genotype data. *Genetics* **155**, 945-959
- 752 Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of*
 753 *the National Academy of Sciences of the United States of America* **94**, 9197-9201.
 754 doi:10.1073/pnas.94.17.9197
- 755 Ravindran PN, Bentzen P, Bradbury IR, Beiko RG (2018) PMERGE: Computational filtering of
 756 paralogous sequences from RAD-seq data. *Ecology and Evolution* **8**, 7002-7013.
 757 doi:10.1002/ece3.4219
- 758 Rognes T, Flouri T, Nichols B, Quince C, Mahe F (2016) VSEARCH: a versatile open source tool for
 759 metagenomics. *PeerJ* **4**. doi:10.7717/peerj.2584
- 760 Schleicher J, Peres CA, Amano T, Lactayo W, Leader-Williams N (2017) Conservation performance
 761 of different conservation governance regimes in the Peruvian Amazon. *Scientific Reports* **7**.
 762 doi:10.1038/s41598-017-10736-w
- 763 Schubert M, Ermini L, Sarkissian CD, *et al.* (2014) Characterization of ancient and modern genomes
 764 by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature*
 765 *Protocols* **9**, 1056-1082. doi:10.1038/nprot.2014.063

- Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes* **9**. doi:10.1186/s13104-016-1900-2
- Siriwat P, Nijman V (2018) Using online media-sourced seizure data to assess the illegal wildlife trade in Siamese rosewood. *Environmental Conservation* **45**, 352-360. doi:10.1017/S037689291800005x
- So NV (2000) The potential of local tree species to accelerate natural forest succession on marginal grasslands in Southern Vietnam. In: *Proceedings of the Workshop on Forest 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, Chiang Mai, Thailand.
- So T, Theilade I, Dell B (2010) Conservation and utilization of threatened hardwood species through reforestation - an example of *Azelia xylocarpa* (Kruz.) Craib and *Dalbergia cochinchinensis* Pierre in Cambodia. *Pacific Conservation Biology* **16**, 110-116
- Theilade I, Yanshuk A, Hald S (2004) Establishment and management of *ex situ* conservation stands In: *Forest genetic resources conservation and management. Vol 3: In plantations and genebanks (ex situ)*. (ed. FAO/FLD/IPGRI). Rome, Italy, International Plant Genetic Resources Institute.
- UNEP-WCMC, IUCN (2018) Protected Planet: The World Database on Protected Areas (WDPA). UNEP-WCMC and IUCN., Cambridge, UK.
- Verdu CF, Guichoux E, Quevauvillers S, *et al.* (2016) Dealing with paralogy in RADseq data: in silico detection and single nucleotide polymorphism validation in *Robinia pseudoacacia* L. *Ecology and Evolution* **6**, 7323-7333. doi:10.1002/ece3.2466
- Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population-Structure. *Evolution* **38**, 1358-1370. doi:10.2307/2408641
- Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution* **11**, 413-418. doi:10.1016/0169-5347(96)10045-8

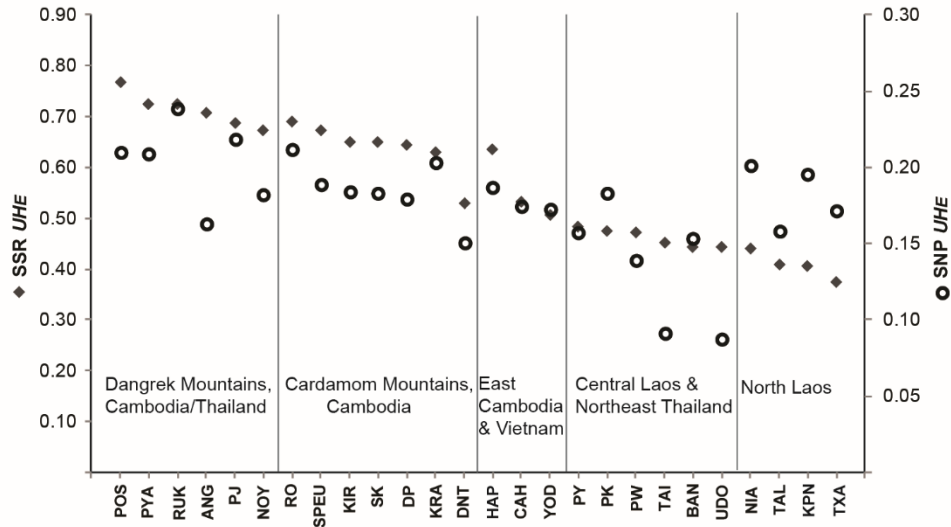
792 **Fig 1 Indochina map with positions of sampled *D. cochinchinensis* populations and overlaid with data for forest**
 793 **cover loss 2001-2016 (red), and protected areas (dotted areas).** Size of population symbol corresponds to estimates of
 794 unbiased expected heterozygosity (UH_e) based on SSR data, and populations used for RAD library are marked with *.
 795 Population names refer to Table 1 and Table S1, and populations are marked corresponding to the proposed Management
 796 Units: 1. North Laos (populations NIA, TXA, TAL, BAN), 2. Central Laos and Northeast Thailand (TAI, UDO, KPN, PW,
 797 PK, PY), 3. Dangrek Mountains (PJ, POS, RUK, ANG, NOY, PYA), 4. Cardamom Mountains (RO, KRA, SK, DP, KIR,
 798 SPEU, DNT) and 5. East Cambodia & Vietnam (HAP, CAH, YOD). (Sources: UNEP-WCMC & IUCN 2018,
 799 Hansen/UMD/Google/USGS/NASA 2018, EsriDigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA,
 800 USGS, AeroGRID, IGN, and the GIS User Community).



803 **Fig 2 Spatial distribution of clones in populations of *D. cochinchinensis*.** Percentage of pairs of trees with identical or
 804 unique genotypes, within each distance class. Numbers above bars indicate total number of pairwise comparisons within
 805 each distance class.

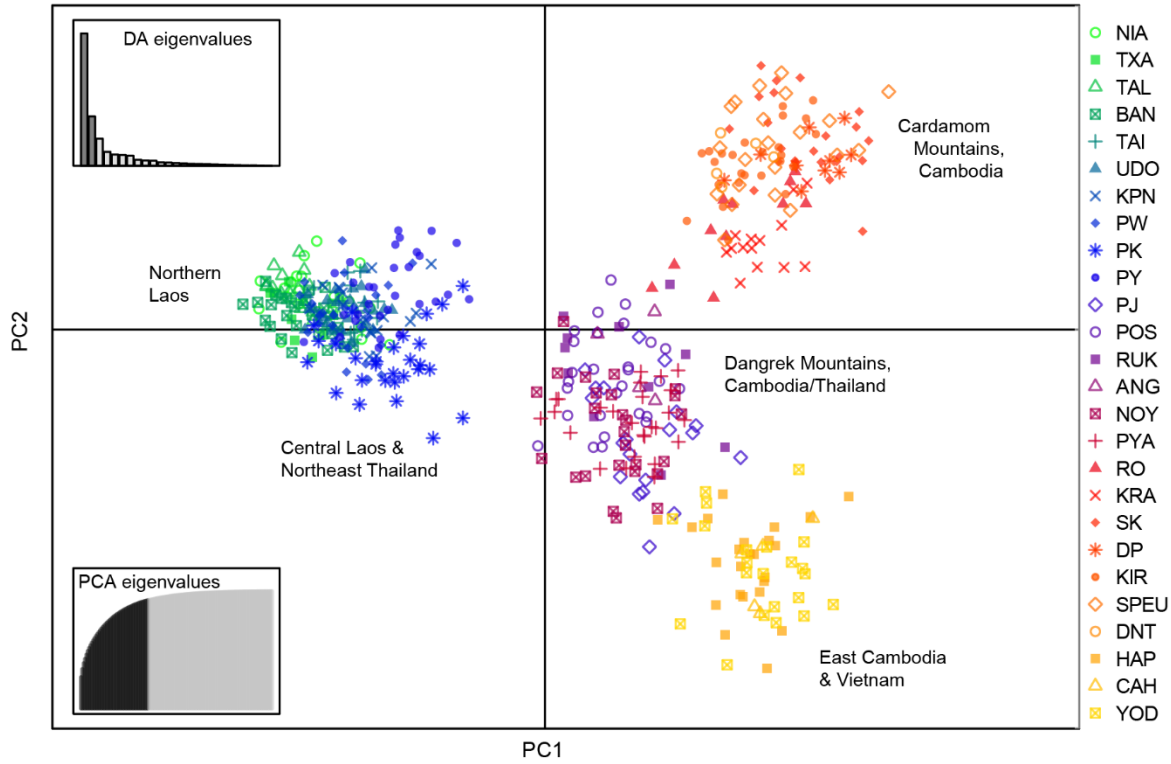


808 **Fig 3 SSR and SNP diversity levels in *D. cochinchinensis* populations in the five regions detected by DAPC and**
 809 **STRUCTURE analyses, sorted by descending SSR U_{H_E} levels. Populations within regions also sorted by SSR U_{H_E} .**
 810 **Regions defined as the clusters identified in the DAPC and STRUCTURE analyses.**

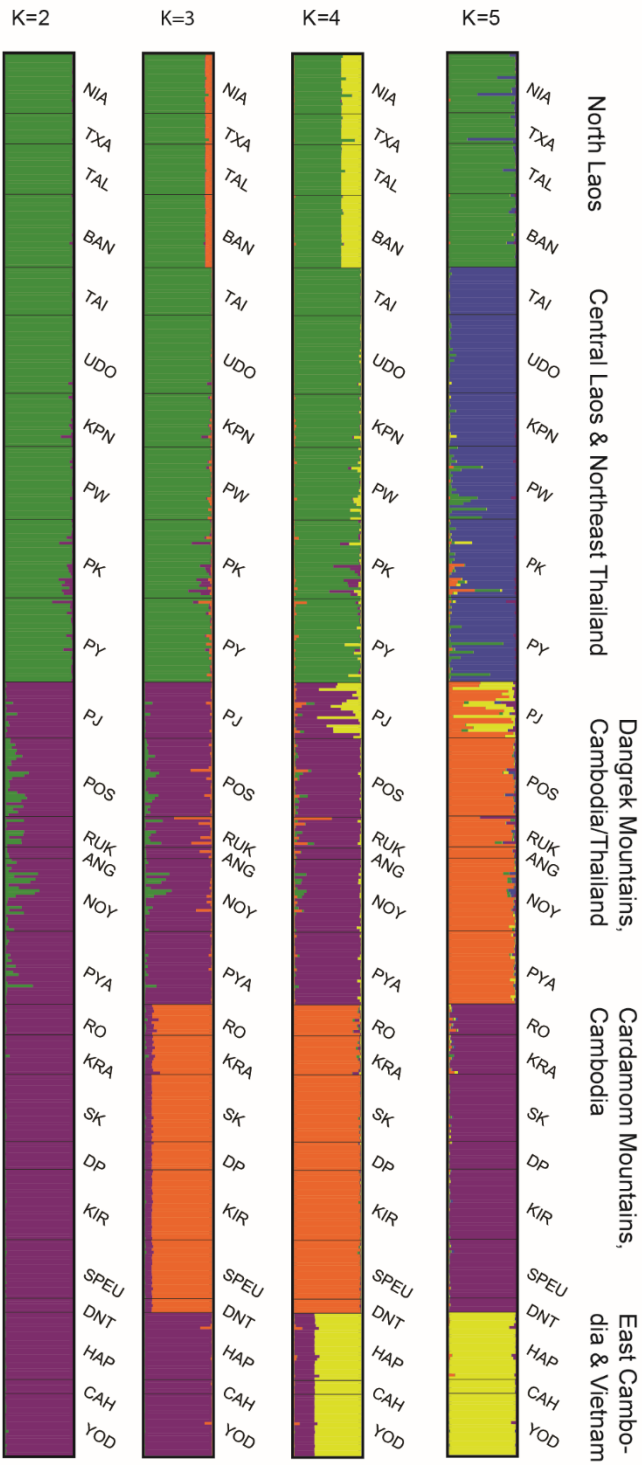


811

812 **Fig 4 Discriminant Analysis of Principal Components for combined SSR and SNP genotypes for 26 populations of *D.***
 813 ***cochinchinensis*.** Bottom-left corner: Proportion of principal component eigenvalues retained as predictors for linear
 814 discriminant analysis shown in black. Upper left corner: The first two DA eigenvalues correspond to the horizontal and
 815 vertical axes, respectively (shown in dark grey).



818 Fig 5 Structure results from K=2 to K=5 for combined SSR+SNP data for 26 populations of *D. cochinchinensis*.



820 **Table 1. Genetic characteristics of 26 *D. cochinchinensis* populations estimated with SSR and SNP markers.** NP=
 821 National Park, N_S = number of individual samples, N_G = number of unique genotypes used for further analyses, H_O =observed
 822 heterozygosity, $U H_E$ = unbiased expected heterozygosity, F_{IS} = Fixation index, % Pol= Percentage polymorphic SNP loci.

					SSR			SNP			
Country	Locality and region	Pop	N_S	N_G	H_O	$U H_E$	F_{IS}	%Pol	H_O	$U H_E$	F_{IS}
Laos	Phu Khao Khuay NP, Da Niai, Bolikhamsay	NIA	25	21	0.39	0.44	0.11	47%	0.20	0.20	-0.01
	Phu Khao Khuay NP, Tad Xai, Bolikhamsay	TXA	14	11	0.37	0.38	-0.07	47%	0.20	0.17	-0.20
	Phu Khao Khuay NP, Tad Leuk, Bolikhamsay	TAL	22	18	0.36	0.41	0.08	47%	0.16	0.16	-0.05
	Phu Khao Khuay NP, Ba Na, Bolikhamsay	BAN	26	26	0.46	0.44	-0.05	58%	0.15	0.15	-0.03
	Na Kai Tai, Khammouane	TAI	24	17	0.46	0.45	-0.04	32%	0.10	0.09	-0.11
	Udomsuk, Khammouane	UDO	28	28	0.40	0.44	0.05	37%	0.10	0.09	-0.11
	Khaun Pheung Neua, Khammouane	KPN	26	19	0.45	0.41	-0.09	47%	0.25	0.20	-0.27
Thailand	Phu Wua Wildlife Sanctuary, Bueng Kan	PW	26	26	0.41	0.47	0.12	58%	0.15	0.14	-0.10
	Phukao-Phuphankham NP, Nongbualamphu	PK	28	28	0.40	0.47	0.12	63%	0.18	0.18	-0.03
	Phu Pha Yon NP, Sakon Nakhon	PY	30	30	0.46	0.48	0.01	53%	0.16	0.16	0.04
	Phu chong-nayoi NP, Ubon Ratchathani	PJ	20	20	0.68	0.69	0.01	68%	0.21	0.22	0.07
Cambodia	Phnom O'Smach, Otdor Meanchey	POS	30	28	0.79	0.77	-0.05	79%	0.21	0.21	0.01
	Rukkavan, Otdor Meanchey	RUK	26	11	0.71	0.72	-0.02	68%	0.22	0.24	0.00
	Anglong Veng, Otdor Meanchey	ANG	20	4	0.72	0.71	-0.16	37%	0.14	0.16	-0.04
	Sre Noy, Siem Reap	NOY	27	26	0.65	0.67	0.01	74%	0.18	0.18	0.00
	Prey Eth, Prey Vihear	PYA	31	26	0.72	0.73	-0.01	68%	0.21	0.21	-0.01
	Rovieng, Pursat	RO	25	11	0.72	0.69	-0.08	53%	0.21	0.21	0.03
	Kravanh, Pursat	KRA	25	14	0.71	0.63	-0.17	58%	0.20	0.20	0.12
	Spin Kdah, Koh Kong	SK	30	24	0.68	0.65	-0.08	53%	0.18	0.18	-0.03
	Daeng Peng, Koh Kong	DP	28	10	0.68	0.64	-0.08	42%	0.22	0.18	-0.29
	Kirirom NP, Kampong Speu	KIR	28	25	0.66	0.65	0.00	53%	0.19	0.18	0.06
	Kampong Speu, Kampon Speu	SPEU	26	21	0.71	0.67	-0.08	53%	0.18	0.19	0.01
	Damnak Neakta Thmapoun, Kampot	DNT	9	5	0.44	0.53	0.12	42%	0.16	0.15	-0.13
Vietnam	Hark Port, Rattanakiri	HAP	24	24	0.62	0.64	0.03	58%	0.22	0.19	-0.16
	Cam Ha, Da Nang	CAH	6	5	0.53	0.53	-0.08	42%	0.16	0.17	0.01
	Yok Don NP, Dak Lak	YOD	39	22	0.47	0.51	0.04	47%	0.17	0.17	-0.01
Overall mean			24.7	19.1	0.56	0.57	-0.02	53%	0.18	0.18	-0.04

824 **Table 2 Genetic diversity levels calculated at region level for *D. cochinchinensis*.** Regions defined as clusters identified
 825 in the DAPC and STRUCTURE analyses. U_{H_E} = expected heterozygosity, % Pol= Percentage polymorphic loci.

	SSR U_{H_E}	SSR private alleles	SNP U_{H_E}	SNP % Pol
Dangrek Mountains	0.79	11	0.22	100
Cardamom Mountains	0.72	12	0.20	68
East Cambodia and Vietnam	0.61	4	0.19	58
Central Laos & Northeast Thailand	0.53	5	0.16	68
Northern Laos	0.44	0	0.18	58

826

827 **Table 3 Results of hierarchical AMOVA analysis with 26 populations and 5 regions of *D. cochinchinensis*.** * $p=0.001$.

828 Regions defined as the clusters identified in the DAPC and STRUCTURE analyses.

Source of variation	df	SSR			SNP		
		Variance components	Percentage of variation		Variance components	Percentage of variation	
Among regions	4	2.453	27	$\Phi_{RT}=0.27^*$	3.767	44	$\Phi_{RT}=0.44^*$
Among populations within regions	21	1.279	14	$\Phi_{PR}=0.19^*$	0.690	8	$\Phi_{PR}=0.15^*$
Within populations	974	5.337	59		4.060	48	
Total	999	9.069	100	$\Phi_{PT}=0.41^*$	8.517	100	$\Phi_{PT}=0.52^*$

829