

Evaluating the genetic diversity in two tropical leguminous trees, *Dalbergia cochinchinensis* and *D. nigrescens*, in lowland forests in Cambodia and Thailand using MIG-seq

Moeko Hirata¹, Chika Mitsuyuki², Etsuko Moritsuka^{3,8}, Phourin Chhang⁴, Shuichiro Tagane^{3,5}, Hironori Toyama^{3,6}, Heng Sokh⁴, Sukid Rueangruea⁷, Somran Suddee⁷, Yoshihisa Suyama², Tetsukazu Yahara³, Kousuke M. Teshima³, Hidenori Tachida³ and Junko Kusumi^{8*}

¹Graduate School of Systems Life Sciences, Kyushu University, Fukuoka, Fukuoka 819-0395, Japan

²Graduate School of Agricultural Science, Tohoku University, Osaki, Miyagi 989-6711, Japan

³Department of Biology, Faculty of Science, Kyushu University, Fukuoka, Fukuoka 819-0395, Japan

⁴Institute of Forest and Wildlife Research Development, Forestry Administration, Phnom Penh 12205, Cambodia

⁵The Kagoshima University Museum, Kagoshima University, Kagoshima, Kagoshima 890-0065, Japan

⁶Center for Environmental Biology and Ecosystem Studies, National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan

⁷The Forest Herbarium, Department of National Parks, Wildlife and Plant Conservation, Bangkok 10900, Thailand

⁸Department of Environmental Changes, Faculty of Social and Cultural Studies, Kyushu University, Fukuoka, Fukuoka 819-0395, Japan

(Received 14 May 2020, accepted 30 November 2020; J-STAGE Advance published date: 18 March 2021)

It is vital to measure the levels of genetic diversity and differentiation between populations in a species to understand the current genetic structure and evolution of the species. Here, MIG-seq (multiplexed inter-simple sequence repeat genotyping by sequencing) was employed to assess the genetic variation in two tropical leguminous tree species, *Dalbergia cochinchinensis* and *D. nigrescens*, in Cambodia and Thailand. Sequence data for 255–618 loci, each with an approximate length of 100 bp, were obtained, and the nucleotide diversity, Tajima's *D* and *F_{ST}* were computed. The estimates calculated from the data obtained by MIG-seq were compared to those obtained by Sanger sequencing of nine nuclear coding genes in *D. cochinchinensis* in our previous study. The nucleotide diversity at the MIG-seq loci was slightly higher than that at silent sites in the coding loci, whereas the *F_{ST}* values at the MIG-seq loci were generally lower than those at the coding loci, although the differences were not significant. Moreover, nucleotide diversities within populations of the two species were similar to each other, at approximately 0.005. Three and four population clusters were genetically recognized in *D. cochinchinensis* and *D. nigrescens*, respectively. Although the populations were differentiated from each other, the levels of differentiation among them, as measured by *F_{ST}*, were higher in *D. cochinchinensis* than in *D. nigrescens*. This indicates higher levels of gene flow between the populations in the latter species. We recommend using MIG-seq for quick surveys of genetic variation because it is cost-effective and results in smaller variance in the estimates of population genetic parameters.

Key words: MIG-seq, genetic variation, *Dalbergia*, population structure

INTRODUCTION

Monitoring genetic variation within a species is important for understanding the current genetic structure of the populations in the species and also for inferring

Edited by Yoshihiko Tsumura

* Corresponding author. jkusumi@scs.kyushu-u.ac.jp

DOI: <https://doi.org/10.1266/ggs.20-00026>

the evolutionary history of the species (Nielsen and Beaumont, 2009; Ellegren and Galtier, 2016). Moreover, it also enables the prediction of the future genetic structure of the species using standard population genetic theory (Yoder et al., 2018; see also general textbooks of population genetics, e.g., Charlesworth and Charlesworth, 2010). Although a desirable way to undertake the monitoring is to resequence the whole genome of many individuals with sufficient read depth, as has been done, for example, in humans (The 1000 Genomes Project Consortium, 2015), it is not yet feasible for most non-model organisms due to the lack of reference genome sequences and to the high costs and labor that are required to resequence an entire genome using the genome resequencing approach. However, many non-model organisms are of interest from the perspective of evolution or for their conservation. For those organisms, simpler and more cost-effective approaches for monitoring genetic variation have been conducted (e.g., sequencing of genes in organelle genomes and genotyping microsatellite marker loci; Ohtani et al., 2013). Such lower-cost approaches still enable us to know the current population structure, the levels of diversity within the populations and the differentiation between the populations, all of which can be used for monitoring genetic variation (Mimura et al., 2017). In addition, information about whether a population is expanding or shrinking is desirable because the non-equilibrium status of populations affects both neutral and non-neutral future genetic variation (Brandvain and Wright, 2016; Simons and Sella, 2016). This kind of information is also relevant from the conservation perspective (Shafer et al., 2015; Yoder et al., 2018). A simple way to infer the non-equilibrium status of a population is to calculate some of the neutrality statistics, such as Tajima's D (Tajima, 1989a, 1989b; Ramos-Onsins and Rozas, 2002). To obtain good estimates of these statistics, it is desirable to survey as many loci as possible. The genes in the genome of an organelle are often not suitable for monitoring genetic variation because they are completely linked, and are considered as one locus. Additionally, developing many microsatellite loci is time-consuming. Moreover, as the levels of diversity at microsatellite loci depend on how the loci were chosen based on their polymorphisms in a small sample, they are, for example, not comparable between species (see Mimura et al., 2017).

To assess the genetic variation at many loci more cost-effectively, several reduced-representation sequencing (RRS) methods, which include restriction site-associated DNA sequencing (Baird et al., 2008; reviewed by Andrews et al., 2016) and genotyping by sequencing (Davey et al., 2012), have been developed. In these methods, after amplifying specific parts of the genome from each sample and adding an identifying tag, the DNA from all samples is pooled and subjected to next-generation sequencing

(NGS), thus generating sequence data at many loci in multiple samples at a reasonable cost. Since RRS methods generate sequence data, an estimation of nucleotide diversity (Nei and Tajima, 1981) can be made; the estimation in turn allows approximately unbiased estimates of F_{ST} (Excoffier et al., 1992; Hudson et al., 1992) to be obtained. Among the different RRS methods, multiplexed inter-simple sequence repeat (ISSR) genotyping by sequencing (MIG-seq; Suyama and Matsuki, 2015), which amplifies regions flanked by microsatellite loci using ISSR primers, is advantageous for the measurement of genetic variation because it requires only a small amount (as little as 2 ng) of DNA of a wide range of qualities (Suyama and Matsuki, 2015). This is because it targets regions of the genome without using restriction enzymes, which require a larger amount of clean DNA. Therefore, MIG-seq can be applied to more species than the other RRS methods. In addition, because the MIG-seq method selects regions flanked by microsatellites, most of the loci are expected to be located in noncoding regions, which are considered less affected by selection. Thus, if the MIG-seq method is used to monitor genetic variation, we can postulate that we are likely measuring neutral genetic variation (although not all noncoding regions are selectively neutral), which gives us information about the present genetic diversity, population structure and past demographic features of the populations of the species.

In this study, we used MIG-seq to assess the genetic variation in two species of *Dalbergia*, *D. cochinchinensis* Pierre and *D. nigrescens* Kurz, which belongs to Fabaceae. *Dalbergia cochinchinensis* and *D. nigrescens* are trees that are 15–30 m and 15–20 m high, respectively (Niyomdham et al., 1997; Niyomdham, 2002). Both trees grow in lowland mixed deciduous or dry evergreen forests in both the monsoon and the savanna climates in Southeast Asia (Niyomdham, 2002; Toyama et al., 2013). Although the ecological features of *D. nigrescens* have not been thoroughly investigated, *D. cochinchinensis* is known to be insect-pollinated, to have wind-dispersed seedpods, and to occasionally reproduce clonally (CITES 2013, Appendices I, II and III; retrieved from <https://cites.org/eng/cop/index.php>). Fabaceae is a key target of the Global Legume Diversity Assessment because of its ecological and economic importance (Yahara et al., 2013), and *Dalbergia* is one of the prioritized genera in this effort because of its species richness and pantropical distribution (e.g., Vatanparast et al., 2013). A common name for *D. cochinchinensis* is Siamese rosewood, which is often used for expensive furniture and musical instruments. Because of its commercial value, it has been logged extensively, and its numbers have recently been reduced so drastically that the IUCN categorized it as Vulnerable in the IUCN Red List of Threatened Species 1998 (<https://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T32625A9719096.en>). Therefore, genetic studies

using allozyme (Soonhuay, 1994), chloroplast (Yooyuen et al., 2011), RAPD and ISSR (Hien and Phong, 2012) and microsatellite (Hartvig et al., 2017) markers have been conducted to elucidate its population structure. While *D. nigrescens* does not have a high commercial value, its abundance has also decreased because it has been logged for firewood and other noncommercial uses. By exploring and comparing the genetic variation and the patterns of genetic variation of these two *Dalbergia* species, which grow in similar environments and have an overlapping distribution range, we may be able to observe differences in the population structure of the two species, and to relate such differences to the species' ecological features. In addition, our results should be applicable to designing conservation plans for the two species.

Previously, we analyzed the genetic variation of *D. cochinchinensis* by resequencing nuclear coding genes (Moritsuka et al., 2017). That study revealed that the level of silent nucleotide diversity within populations of *D. cochinchinensis* ranged from 0.0027 to 0.0050, and the populations were strongly differentiated. Although nuclear gene markers are useful for investigating the genetic variation in non-model species, it cannot be ignored that natural selection may influence estimates of the levels of genetic diversity and population differentiation (e.g., *Adh* in *Drosophila*; see Kreitman and Hudson, 1991). Indeed, in our previous study we found a few candidate loci under natural selection in this species (Moritsuka et al., 2017); therefore, it was worthwhile to now examine the genetic variation in the noncoding regions using the MIG-seq method, which is considered to experience less pressure from natural selection, and to compare the results with those obtained for the coding genes. Additionally, the accuracy of the estimates obtained using a small number of loci (~10 loci) may not be adequate to evaluate the population structure or to measure the genetic variation and levels of differentiation because the coefficient of variation for the estimate of nucleotide diversity is of the order of one (Tajima, 1983). Therefore, data for more than 100 loci are desirable to obtain reasonable estimates of nucleotide diversity.

In this paper, we analyzed the genetic variation in *D. cochinchinensis* and *D. nigrescens* using MIG-seq. The aims of this study were threefold: 1) to assess the basic features of the genetic variation at the loci surveyed by MIG-seq (MIG loci) in the two species to infer the population structure and the levels of nucleotide diversity and differentiation in these two species; 2) to compare the results for *D. cochinchinensis* with those of the nuclear coding genes obtained by the conventional sequencing method for this species (Moritsuka et al., 2017), focusing on the accuracy of the estimates and differences in genetic variation between the coding regions and those surveyed by MIG-seq; and 3) to compare the features of

the genetic variation between the two species, relating them to their ecological features.

MATERIALS AND METHODS

Sampling Leaf samples of *D. cochinchinensis* and *D. nigrescens* were collected from trees in ten populations in Cambodia and one population in Thailand (Fig. 1, Supplementary Table S1). As there were clones among the sampled trees, we chose one representative from each set of clones (see Moritsuka et al., 2017 for the method of identifying clones). As a consequence, samples from 15 trees in Kampong Thom (KT), seven trees in Kampot1 (KM1), 13 trees in Kampot2 (KM2) and 13 trees in Siem Reap (SR), totaling 48 samples, of *D. cochinchinensis* were used for the analysis. These samples were the same as those used in Moritsuka et al. (2017). For *D. nigrescens*, samples from four trees in KT, four trees in KM2, 11 trees in SR, four trees in Monduliri Province (MN), six trees in Kratie Province (KR), five trees in Tboung Khmum Province (TB) and 15 trees in Phu Kradung (TP) were used. In total, samples from 49 *D. nigrescens* trees were used. All locations, except for Phu Kradung, Thailand, were in Cambodia. The individual trees used for the analyses were at least 3 m apart. DNA was extracted from the dried leaves using the CTAB method (Murray and Thompson, 1980).

MIG-seq MIG-seq (Suyama and Matsuki, 2015) was employed to obtain the sequence data. We followed the procedure described in Suyama and Matsuki (2015) to obtain quality-filtered reads. Although 2×75 bp paired-end sequencing is usually used in NGS with MIG-seq, we used 2×300 bp paired-end sequencing with the MiSeq Reagent Kit v3 (600 cycles) to obtain longer reads. The raw reads were grouped into those from individual samples using the index read option. Additionally, using the dark cycle option of the MiSeq system, the sequences of the first 17 bp (2^{nd} PCR anchor + SSR + 1^{st} PCR anchor) of read1 (forward) and the first 3 bp (2^{nd} PCR anchor) of read2 (reverse) were skipped. Therefore, we obtained sequences of up to 251 bp from read1 and 265 bp from read2.

Quality control of the raw reads First, the first 14 bp of read2 were trimmed using the program fastx_trimmer in the FASTX-Toolkit 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/) so that the length of read2 also became 251 bp. Reads containing the sequence of the other primer (GTCAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC for read1 and CAGAGATCGGAAGAGCGTCGTGTAGGGAAAGA for read2) were removed using the program cutadapt 1.12 (Martin, 2011) with the discard-trimmed option, because the available sequences from such reads were too short for the analysis. Since

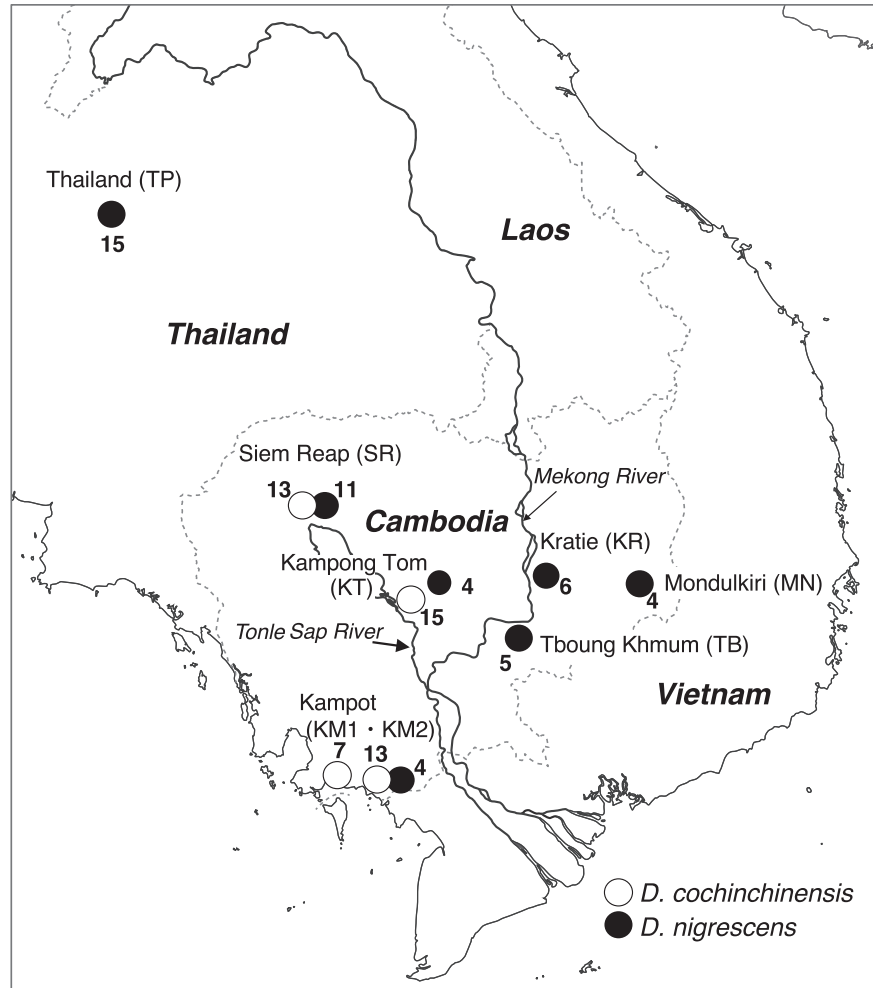


Fig. 1. Sampling locations. The number of samples collected at each location is shown next to the sampling location.

a preliminary inspection of the base calls indicated that the quality at the tail ends of the reads was deficient, we decided to trim bases from the tail end in each read. We tried three trimming lengths: the tail ends of 1, 101 or 151 bp of read1 and read2 were trimmed using the `fastx_trimmer` in the FASTX-Toolkit so that the final length (L) of each read became 250, 150 and 100 bp, respectively. Finally, the reads were filtered using the `quality_filter` of the FASTX-Toolkit with $q = 25$ and $p = 80$. Only those from read1 and read2 were chosen using `cmpfastq` (<http://compbio.brc.iop.kcl.ac.uk/software/cmpfastq.php>). The data from read1 and read2 of each sample were merged and used for subsequent analyses.

Compiling reference sequences We first built reference sequences separately for *D. cochinchinensis* and *D. nigrescens*. The trimmed and quality-filtered reads with $L = 100, 150$ and 250 bp were processed separately using *Stacks* ver. 1.4 (Catchen et al., 2011). First, stacks were built by *de novo* assembly using *ustacks*. The *ustacks* m parameter was set to 20, which is the mini-

mum number of raw reads required to form a stack (putative allele). The M parameter of *ustacks* was changed depending on the read length, which controls the number of mismatches allowed between the stacks to merge them at a putative locus ($M = 2, 3$ and 6 for $L = 100, 150$ and 250 bp, respectively). Then, using *cstacks*, a catalog of stacks was created with the option of $n = 5, 7$ and 12 for $L = 100, 150$ and 250 bp, respectively, allowing the maximum mismatches of n among the stacks. The algorithm of *cstacks* generated a catalog of consensus sequences for each stack (putative loci). Hence, these sequences were used as reference sequences in the subsequent mapping step. Because the presence of repeated sequences in the reference sequences would disturb the mapping of the reads, we applied RepeatMasker (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>) to the consensus sequences to mask known repeated sequences by 'N'. If a consensus sequence had 30 or more Ns, it was discarded. As a reference database for the repeated sequences, *Arabidopsis thaliana* was used. Finally, the program *cd-hit-est* from CD-HIT 4.6 (Fu et al., 2012) with

the sequence identity set to 0.95 was used to bundle consensus sequences with 95% identity to each other.

Two catalogs of consensus sequences, one for each of *D. cochinchinensis* and *D. nigrescens*, were initially created and used as references for mapping the reads from the samples of each species. We separately ran the same process of building the references for each genetic cluster because two separate genetic clusters, which shared only a small number of stacks (putative loci), were identified in the samples of *D. nigrescens*. This is explained further in the Results section.

Mapping reads to the reference sequences The trimmed and quality-filtered reads were mapped to the reference sequences using bwa mem (BWA ver. 0.7.15-r1140; Li et al., 2009) as the single reads, and the other options were set to their default values. These files were then sorted, and a pileup file was generated using SAMtools (Li et al., 2009). The variant sites, as well as the monomorphic sites, were called with BCFtools (Li and Durbin, 2009; Li et al., 2009). Then, by applying the filter subcommand of BCFtools, sites with minimum depth 10 and maximum depth 200 were obtained. Next, loci with lengths of less than 80, or shared by less than 80% of the samples, were removed. Finally, for each locus, the deviation from Hardy–Weinberg equilibrium (HWE) in each species was assessed at FDR = 0.01 with the χ^2 test via a custom script; any loci that did not conform to HWE were removed.

Population genetics analyses For the filtered loci, we randomly picked one polymorphic site from each filtered polymorphic locus (single SNPs) to conduct principal component analysis (PCA) and STRUCTURE analysis (STRUCTURE 2.3, Pritchard et al., 2000; Hubisz et al., 2009). PCA was performed using Tassel 5.2.33 (Bradbury et al., 2007) with the eigen-decomposition approach. We applied the analysis to the genotype data at 393, 234 and 503 single SNPs for DC, DN and DN-C groups, respectively. The vcf files that included only single SNPs were converted to the input file format for the STRUCTURE program using PGDspider ver. 2.1.1.5 (Lischer and Excoffier, 2012). For STRUCTURE analyses, we set the number of clusters, K , from 1–7. For each K value, 10 independent runs were executed with burn-in and Markov chain Monte Carlo iterations of 100,000 each. We used pophelper ver.2.3 (Francis, 2017) to summarize the data for each K and to calculate Evanno’s ΔK (Evanno et al., 2005).

FASTA files were created from the vcf files that included all monomorphic and polymorphic sites using a custom script. The indel sites were represented by ‘N’ in the FASTA files. From these FASTA files, the nucleotide diversities within the populations (π) were first calculated for each population using a custom script. A

small number of loci showed very high values of π , suggesting that duplicated loci were still included. As the inclusion of such loci skews the statistical estimation, we arbitrarily set the threshold at 0.05 for π and excluded loci with $\pi \geq 0.05$. We then estimated π and F_{ST} (Hudson et al., 1992) from the reduced sets of loci using the custom script. The standard error of the estimate of F_{ST} was calculated using a custom script that implemented the jackknife procedure described by Weir and Cockerham (1984). The relationship of F_{ST} with the geographical distance between the populations was examined using the function mantel.test in the R package ape v5.3 with 9,999 permutations. Tajima’s D at each locus was calculated using PopGenome (Pfeifer et al., 2014), and the means and standard errors were calculated.

RESULTS

Quality control of the raw reads Supplementary Table S2 shows the number of reads, total number of sites, mean number of reads per sample and mean number of sites per sample before and after quality filtering for the trimmed read lengths of 100, 150 and 250 bp. In total, 40,976,618 raw reads were obtained from MiSeq. As the mean number of available sites per sample was the highest, we decided to use the reads of the trimmed lengths of 100 bp. Consequently, trimmed read1 and read2 did not overlap because the untrimmed reads of length 251 bp that included the opposite primers had already been removed. After the quality control of the raw reads, the total number of reads in the whole sample was 10,556,844, and the average number of reads per sample was 112,307. The average length per sample was 28,076,713 bp.

Compiling reference sequences for each species and mapping the reads After the quality control of

Table 1. The number of loci and polymorphic sites after filtering

Group ^a	Loci	Total sites	Polymorphic sites
DC	485	48,461	1,919
	(451.71)	(45,021.27)	
DN	255	25,500	1,605
	(235.47)	(23,461.34)	
DN_C	618	61,797	2,166
	(586.82)	(58,543.56)	
TP	512	51,161	1,342
	(455.93)	(45,406.80)	

Average number of loci and sites per individual are in parentheses. ^a DC, *D. cochinchinensis*; DN, *D. nigrescens*; DN_C, Cambodian populations of *D. nigrescens*; TP, Thai population of *D. nigrescens* (see the text for more details).

the raw reads, separate runs of *Stacks* were conducted for *D. cochinchinensis* (DC) and *D. nigrescens* (DN). The total numbers of reference sequences after clustering by similarity using the CD-HIT program and removal of the short sequences are shown in Supplementary Table S3. Although the numbers of reference sequences were different between the population groups, more than 1,500 sequences for each species were used for the following mapping process.

The quality-controlled reads of each species were mapped to those reference sequences. The numbers of loci and available sites resulting from the mapping and filtering by the criteria described above are shown in Table 1. The number of loci after the filtering decreased to ~30% and ~10% of the number of reference sequences, for DC and DN, respectively. A major reason for the substantial reduction in the number of loci was the removal of loci shared by less than 80% of the samples. The reduc-

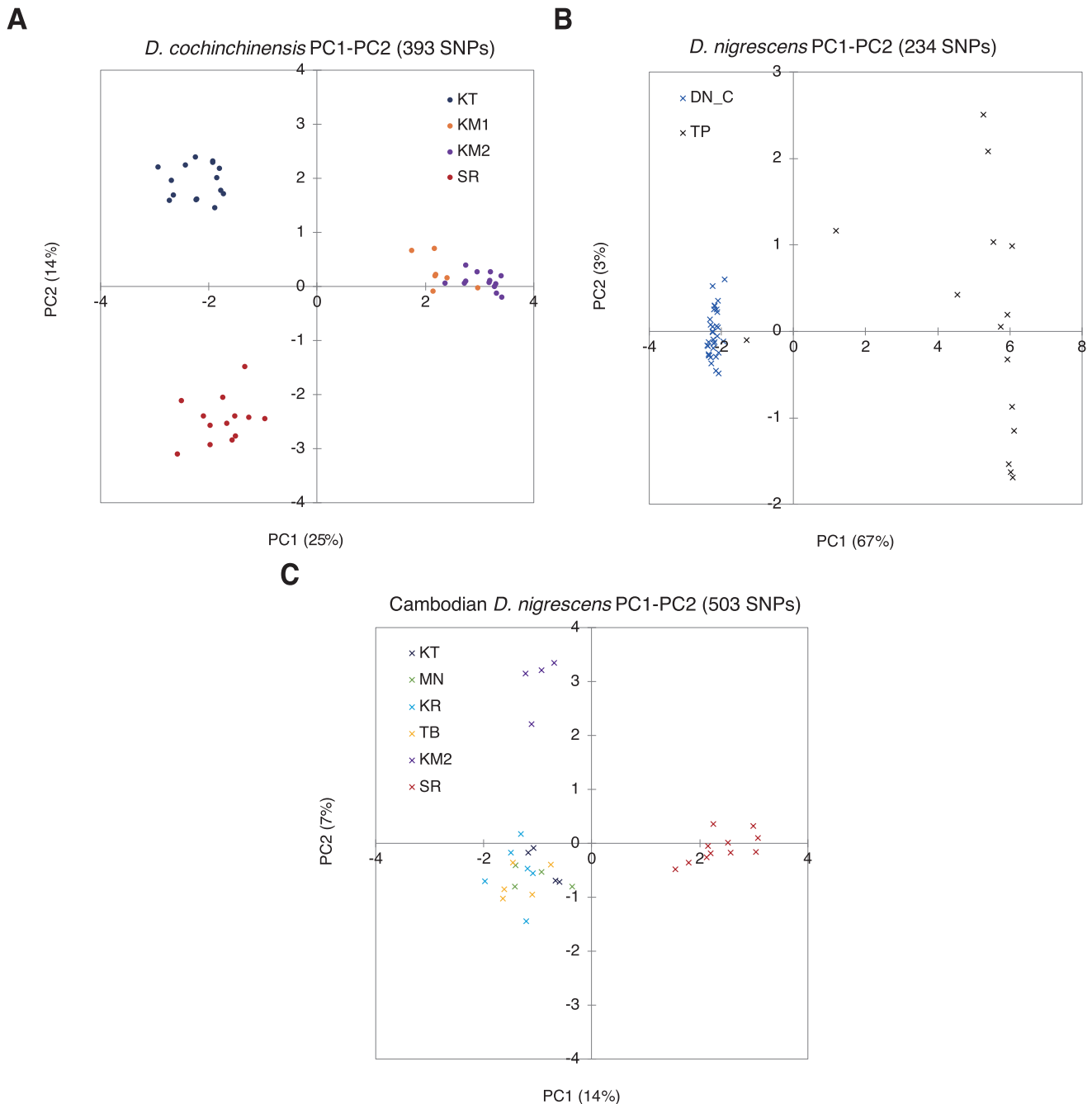


Fig. 2. Results of PCA for the whole *D. cochinchinensis* population (A), the whole *D. nigrescens* population (B) and the Cambodian populations of *D. nigrescens* (C).

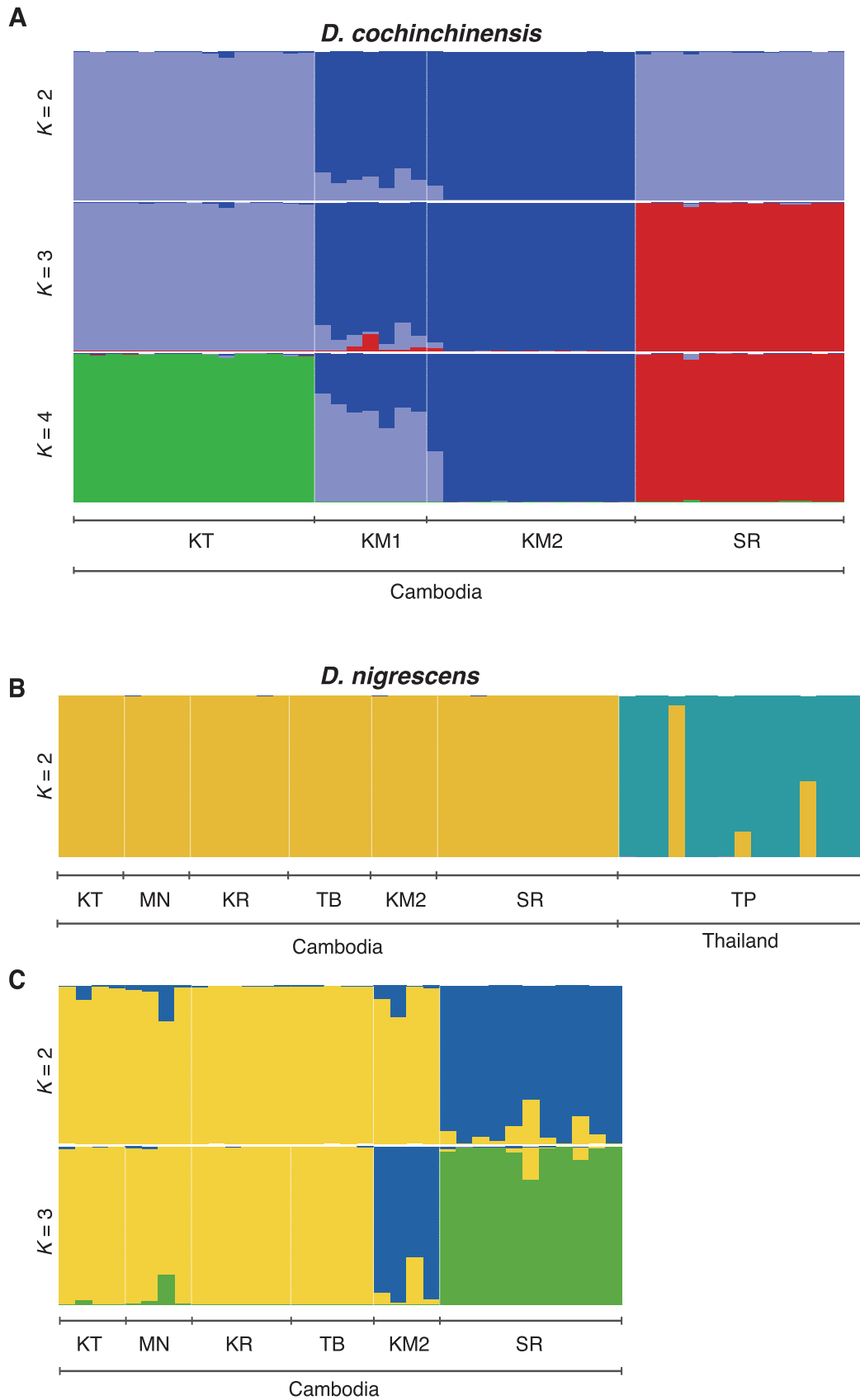


Fig. 3. Results of STRUCTURE analysis for the individual populations of *D. cochinchinensis* (A), all *D. nigrescens* populations (B) and the Cambodian populations of *D. nigrescens* (C).

tion was considerable for DN as there was a strong differentiation between the Thai population and the Cambodian populations (as described later) and these two population groups shared only a small number of loci. Hence, we divided DN into two groups, the TP and the Cambodian populations (DN_C). *Stacks* was then run separately for TP and DN_C, to build the reference sequences for each group (Supplementary Table S3). After mapping the reads to the reference sequence and filtering loci that did not meet the criteria, we could obtain a greater number of available loci and sites (Table 1). Finally, we used 485, 255, 618 and 512 loci for the following analyses of DC, DN, DN-C and TP groups, respectively (Table 1).

Population structure To identify the population structure, we conducted PCA for each species, and the results are shown in Fig. 2. The samples of *D. cochinchinensis* were clustered into three groups, corresponding to the sampling locations, KT, SR and KM1+KM2, with weak clustering recognizable for both KM1 and KM2 (Fig. 2A). The results using STRUCTURE also agreed with this population structure (Fig. 3A), although the likelihood was the highest when $K = 4$, and Evanno's ΔK was the highest when $K = 2$ (Supplementary Fig. S1).

PCA for *D. nigrescens* was conducted using the loci shared by the two divergent genetic groups. The individuals were clustered into two groups, which mostly corresponded to the TP and Cambodian populations (DN_C); two exceptional individuals from Thailand were similar to those from DN_C in the PCA plot (Fig. 2B). The results of STRUCTURE analysis are shown in Fig. 3B. The results also indicated that there were two separate groups; however, they further indicated that there were two or

three exceptional individuals that were considered to be recent putative hybrids between the TP and Cambodian populations. The genetic differentiation between TP and DN_C was extreme; therefore, we built the reference sequences and mapped the reads separately for TP and DN_C. Doing this increased the number of loci usable for the analyses, as described earlier. The result of PCA analysis using DN_C is shown in Fig. 2C. The individuals of DN_C were further divided into three groups: SR, KM2 and the remaining populations located in eastern Cambodia (KT, MN, KR and TB). On the other hand, STRUCTURE analysis identified only two genetic groups, the SR population and the others, and the likelihood and Evanno's ΔK were highest when $K = 2$ (Fig. 3C, Supplementary Fig. S1).

Population statistics The means of π and Tajima's D across loci are shown in Table 2. The estimates of π were similar among the populations of *D. cochinchinensis*; they ranged from 0.00399 (KM2) to 0.00531 (SR). For *D. nigrescens*, the highest value of π was found in TP, 0.00779; it was still high when the two putative hybrids were excluded (0.00584). Both estimates were higher than that for DN_C (0.00534 ± 0.00047), even though DN_C consisted of six populations. The Cambodian populations of *D. nigrescens* showed similar levels of nucleotide diversity. The estimates of π ranged from 0.00422 (KR) to 0.00479 (SR) and were similar to each other, yet significantly lower than TP. The means of Tajima's D across all loci were positive for *D. cochinchinensis*; however, they were predominantly negative for *D. nigrescens*. Only the KM2 population (*D. cochinchinensis*) showed a significant deviation from zero, and the estimate was significantly

Table 2. Mean nucleotide diversity (π) and Tajima's D across loci in each population

Species	Population	n^a	π		Tajima's D		
			Mean	SE	l^b	Mean	SE
<i>D. cochinchinensis</i>	KT	15	0.00456	0.00033	68	0.1492	0.1356
	KM1	7	0.00515	0.00034	74	0.0679	0.1136
	KM2	13	0.00399	0.00032	54	0.8503	0.1349
	SR	13	0.00531	0.00032	77	0.2351	0.1128
<i>D. nigrescens</i>	TP	15	0.00779	0.00058	—	—	—
	TPex ^c	13	0.00584	0.00057	190	-0.1448	0.0774
	KT	4	0.00438	0.00028	110	-0.1375	0.0933
	MN	4	0.00459	0.00030	110	0.0198	0.1044
	KR	6	0.00422	0.00027	123	-0.1364	0.0964
	TB	5	0.00443	0.00029	120	-0.1061	0.0916
	KM2	4	0.00449	0.00029	112	0.0214	0.0930
	SR	11	0.00479	0.00028	144	-0.0384	0.0797

^a Number of samples (diploid).

^b Number of polymorphic loci used to compute Tajima's D .

^c Putative hybrids were excluded.

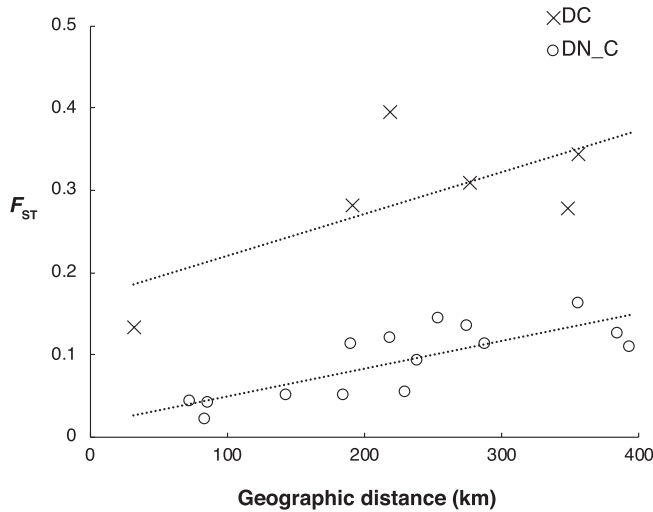


Fig. 4. Relationship between the geographical distance between the populations and the F_{ST} estimates for *D. cochinchinensis* and *D. nigrescens* in Cambodia. The regression lines are shown as dotted lines. Isolation by distance, as examined by the Mantel test, was significant for DN_C ($P = 0.0343$) but not for DC ($P = 0.3347$).

positive.

The means of pairwise F_{ST} estimates across the loci ranged from 0.133 (KM1 and KM2) to 0.395 (KT and KM1) in *D. cochinchinensis* (Supplementary Table S4). In *D. nigrescens*, a high level of genetic differentiation was found between TP and DN_C (Supplementary Table S5). The F_{ST} between TP and DN_C was 0.626 when the putative hybrids were included and 0.7 when they were excluded. In the Cambodian populations of *D. nigrescens*, the F_{ST} values ranged from 0.0209 (KT and TB) to 0.1628 (KM2 and SR) (Supplementary Table S6). The values of F_{ST} among the eastern Cambodian populations (KT, MN, KR and TB) ranged from 0.0209 to 0.0555 and were much lower than those between the pairs involving KM2 or SR (0.0933–0.1628). These findings were in accordance with the PCA and STRUCTURE results.

The pairwise F_{ST} estimates were plotted against the geographical distance between the populations to compare the levels of differentiation between the Cambodian populations in the two species (Fig. 4). The F_{ST} estimates were higher in *D. cochinchinensis* than in *D. nigrescens* when they were compared at the same distance. Isolation by distance (IBD) was significant for the populations of *D. nigrescens* ($P = 0.0343$), although it was not significant for those of *D. cochinchinensis* ($P = 0.3347$).

DISCUSSION

Comparison with the results obtained by Sanger sequencing of several coding genes The estimates by Moritsuka et al. (2017) of nucleotide diversity at silent

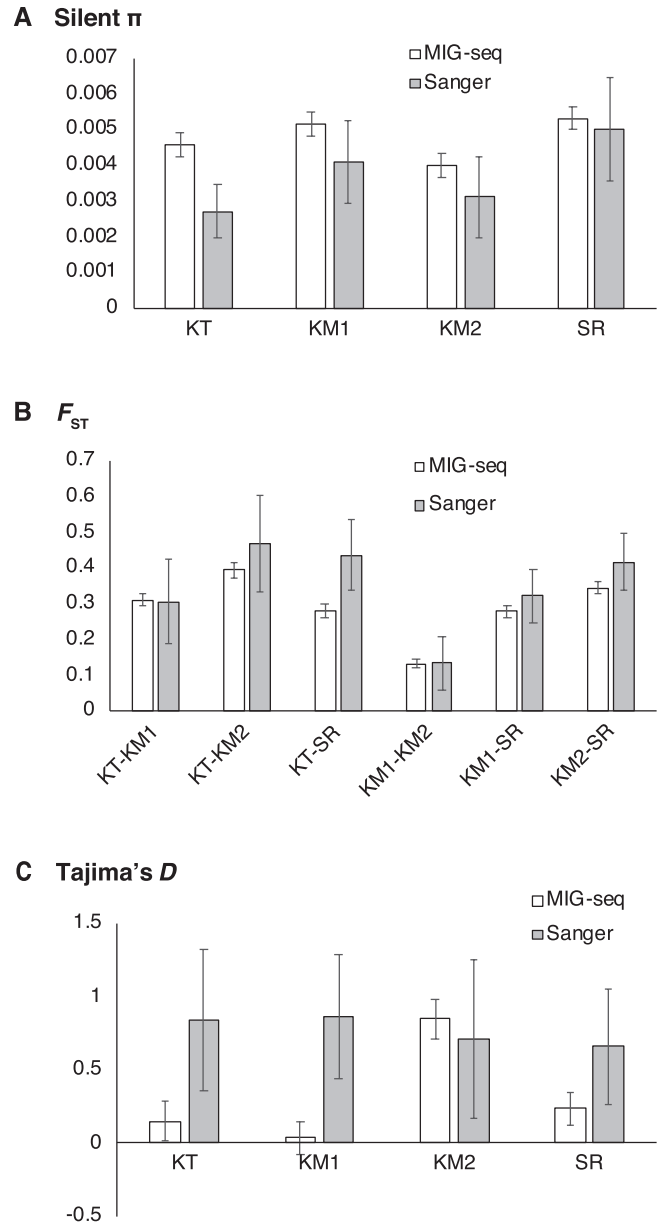


Fig. 5. Comparison of the estimates of diversity measured by MIG-seq with those by the Sanger method (Moritsuka et al., 2017). The error bars for the estimates indicate one standard error.

sites, which may be considered selectively neutral and comparable to that in the noncoding regions targeted by MIG-seq in this study, were smaller than those found here (Fig. 5A). The standard errors of the estimates in Moritsuka et al. (2017) were larger than those in the present study, and a definite conclusion will require data for more coding loci. Nonetheless, if the differences are true, they may be explained by at least three hypotheses. First, selection at the silent sites, such as that on codon usage at synonymous sites (see Akashi, 1994), might reduce the silent nucleotide diversity in cod-

ing genes. Second, a selective sweep (genetic draft; see Gillespie, 2001) or background selection (Charlesworth et al., 1993) may have reduced the silent nucleotide diversity in the coding genes, as these are more likely to be subjected to either positive or negative selection. Finally, errors in genotyping may have increased the estimates of nucleotide diversity in our MIG-seq study. Further studies, such as those using the Sanger method for some of the amplified fragments from MIG-seq, may elucidate whether the inflation of diversity was due to sequencing and/or genotyping errors in NGS.

On the other hand, the estimates of F_{ST} in Moritsuka et al. (2017) were mostly higher than those in this study (Fig. 5B). Again, the standard errors in the estimates of Moritsuka et al. (2017) were larger. However, if the differences were not due to sampling errors, divergent selection on the coding regions may explain the observation. Indeed, Moritsuka et al. (2017) suggested that divergent selection had increased the differentiation between the populations at a few loci, and that this might have inflated the mean of the F_{ST} values across the coding genes, although they also found one locus with a very low F_{ST} value that may have been subject to some type of balancing selection. Other factors, such as selective sweep and background selection, also increase F_{ST} estimates by reducing the variation within populations.

The means of Tajima's D were all positive and they were mostly larger in the coding regions than in the regions surveyed by MIG-seq (Fig. 5C); however, the differences were not significant, due to the larger standard errors of the estimates in the coding regions. A few loci showed significantly negative or positive values of Tajima's D in Moritsuka et al. (2017), which indicated the action of selection. Therefore, the variance may be inflated if Tajima's D is estimated in the coding regions.

If we wanted to evaluate the geographical differentiation or infer the history of the population structure, estimates based on neutral markers with smaller variances would have been preferable. Therefore, the estimates obtained by MIG-seq should be suitable for the general evaluation of the levels of neutral genetic diversity, differentiation and population history of the focal species.

Comparison between the two species The levels of nucleotide diversity in the Cambodian populations of the two species were fairly similar, although they were lower than those for *D. nigrescens* from TP (Thailand). Populations of *D. cochinchinensis* are thought to have dramatically shrunk recently as a result of commercial logging and deforestation (UNEP-WCMC, 2018). However, this does not seem to have had a considerable impact on its genetic diversity, because the nucleotide diversity estimates of these two species (0.00422–0.00584) were not particularly low compared to those of other long-term perennials (Chen et al., 2017). This is understandable

because neutral nucleotide diversity is reduced every generation by $1/(2N_e)$, where N_e is the effective population size, and a few generations of reduced N_e does not cause a significant reduction of the variation unless N_e is extremely small. The nucleotide diversity in KM2 of *D. cochinchinensis* was reduced by about 20% relative to those of the other Cambodian populations of the species, and the positive Tajima's D indicated a recent reduction of population size if we assume that most of the MIG loci were neutral. However, for such a reduction in nucleotide diversity to occur in a few generations, N_e must be less than ten, which is even smaller than the sample size in KM2; therefore, we think the size reduction happened over a longer period, possibly due to some natural cause such as changes of vegetation, rather than to human exploitation. The high nucleotide diversity in *D. nigrescens* from TP may be explained by gene flow from other populations, such as those in another part of Thailand, Cambodia or elsewhere, as we found a few putative hybrids in Thailand. Although we removed those putative hybrids from our estimation of nucleotide diversity, not all recent migrants from other populations may have been recognized by our PCA or STRUCTURE analysis, and they might have had increased nucleotide diversity. Alternatively, TP may have a larger N_e than populations in Cambodia. In any case, more samples from Thailand are needed to elucidate and determine which explanation is more plausible.

The levels of differentiation between the populations seemed generally higher in *D. cochinchinensis* than in *D. nigrescens*. In a more comprehensive survey, Hartvig et al. (2017) reported that populations of *D. cochinchinensis* are more differentiated than those of *D. oliveri*. The higher estimates of F_{ST} suggest either lower levels of gene flow between populations or smaller local population size. Hartvig et al. (2017) suggested that a lower seed dispersal rate by water, pioneer characteristics, and/or a narrower ecological niche of *D. cochinchinensis*, compared to *D. oliveri*, caused the higher differentiation in the former species. In our case, *D. cochinchinensis* grows in mixed deciduous forests and dry evergreen forests, while *D. nigrescens* grows only in mixed deciduous forests (Niyomdham, 2002). This suggests that *D. cochinchinensis* has a broader habitat than *D. nigrescens*, which may imply larger population size or higher levels of gene flow between local populations, contradicting the inference based on the higher level of differentiation in *D. cochinchinensis*. However, a vegetation survey in Thailand by Lamotte et al. (1998) indicated that *D. nigrescens* could grow in more open forests than *D. cochinchinensis*. Therefore, the populations of *D. nigrescens* may be more interconnected than those of *D. cochinchinensis*. The drier climate of glacial periods (Louys and Meijaard, 2010) may allow more gene flow between populations, or permit a larger population size

of *D. nigrescens*. Another explanation for the lower differentiation in *D. nigrescens* may be its higher ability of water dispersal of seeds, as suggested to explain lower differentiation in *D. oliveri* than in *D. cochinchinensis* by Hartvig et al. (2017), although we do not have any evidence for this. However, this may explain the weak differentiation between KT and the three populations on the other side of the Mekong River, KR, MN and TB, because, except for the river, these four populations are connected by deciduous forests according to a vegetation map by Stibig et al. (2004).

The means of Tajima's D were positive in *D. cochinchinensis*, although only that for KM2 was significant. Thus, the populations of this species appear to be decreasing if we assume that most of the MIG loci were neutral. In addition, Moritsuka et al. (2017) suggested that these populations have decreased based on the inference obtained by applying the likelihood-based method implemented in IMA2 (Hey, 2010) to the data of the nine nuclear genes. Hartvig et al. (2017) also suggested that a reduction of the population size in *D. cochinchinensis* and *D. oliveri* occurred during the Pleistocene, based on M_m , which is a statistic to detect bottlenecks, as proposed by Garza and Williamson (2001). However, it is not clear whether the timings of the size reductions detected in the two studies using different marker loci and methods were the same or not. On the other hand, the means of Tajima's D in *D. nigrescens* were close to zero, and none were significantly different from zero. This might suggest that the sizes of the populations in *D. nigrescens* had experienced no notable recent change. However, the levels of gene flow also affect Tajima's D (see Tajima, 1989b). Furthermore, the sample sizes were generally smaller for *D. nigrescens*, which may have reduced the power to detect changes in population size. In any case, to understand why the mean of Tajima's D in *D. cochinchinensis* was larger than that in *D. nigrescens*, we need to infer the history of the populations quantitatively in a similar manner to Moritsuka et al. (2017), which we did not try due to the missing data at respective loci.

Limitations of the samples According to Niyomdham (2002), *D. cochinchinensis* is distributed in Thailand, Laos, Cambodia and Vietnam, and *D. nigrescens* is distributed in Thailand, Cambodia, Myanmar and Vietnam. Thus, our samples covered only parts of the ranges of the two species, and the conclusions drawn from our study may not be applicable to all populations of either species. However, for *D. cochinchinensis*, we could infer the proportion of the genetic variation of the species that were studied by referring to Hartvig et al. (2017), who surveyed species-wide ranges of *D. cochinchinensis* and *D. oliveri* using microsatellite markers. The SR, KM1 and KM2 sites in our study are located in northern and southern Cambodia, which corresponded to only two of

the five regions where a cluster of *D. cochinchinensis* populations was previously identified by Hartvig et al. (2017). Whether the KT samples belonged to the eastern Cambodia/Vietnam cluster identified by Hartvig et al. (2017) could not be determined because those authors did not include samples from the region near KT, which is located between the Mekong River and Tonle Sap River. Considering their suggestion of the importance of large rivers in restricting gene flow, KT may represent an unidentified sixth cluster. The F_{ST} between the northern and southern populations of Cambodia was estimated to be 0.25 or less by Hartvig et al. (2017) (see their Fig. S2), which is slightly lower than our estimates of F_{ST} between the northern (SR) and southern (KM1 or KM2) populations. However, the difference might have been due to differences in the markers and statistics used to evaluate genetic differentiation. Finally, although no statistical significance of IBD in *D. cochinchinensis* was found, this may be due to the small number (four) of populations investigated, because IBD exists in this species as was previously shown by Hartvig et al. (2017) using samples from 26 populations.

Conclusion Assessing the genetic variation in a species is important for understanding its evolution and for predicting the future of the species. Here, MIG-seq, which uses NGS and is applicable to a wide variety of species, was shown to be a useful and efficient method for quickly evaluating the genetic variation of evolutionarily interesting non-model species. The two species studied here seem to have different population histories and/or migratory characteristics as indicated by the lower levels of differentiation between populations, and near-zero values of Tajima's D in *D. nigrescens*. We could also identify three and four genetic clusters in *D. cochinchinensis* and *D. nigrescens*, respectively. Although *D. nigrescens* has not been cut for its high commercial value, as *D. cochinchinensis* has been, it is also vulnerable to the effects of deforestation (Theilade and de Kok, 2015). Future conservation efforts in the area studied here must take this genetic clustering into account.

DATA ARCHIVING STATEMENT

The sequence data from this study have been submitted to the DDBJ Sequence Read Archive (<https://www.ddbj.nig.ac.jp/dra/index.html>) under accession no. DRA009421.

All samples were legally gathered under agreements between Kyushu University and the Forest Administration of Cambodia or the Forest Herbarium, Thailand. We cordially thank both institutions for permits and support for our fieldwork in each country. The Environment Research and Technology Development Fund, Grant Number S-9 (T. Y.), of the Ministry of the Environment, Japan, and JSPS KAKENHI, Grant Numbers JP16H02553 (Y. S.) and JP16K07466 (J. K.), supported this research.

REFERENCES

- Akashi, H. (1994) Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* **136**, 927–935.
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* **17**, 81–92.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., and Johnson, E. A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**, e3376.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635.
- Brandvain, Y., and Wright, S. I. (2016) The limits of natural selection in a nonequilibrium world. *Trends Genet.* **32**, 201–210.
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011) *Stacks*: building and genotyping loci *de novo* from short-read sequences. *G3 (Bethesda)* **1**, 171–182.
- Charlesworth, B., Morgan, M. T., and Charlesworth, D. (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.
- Charlesworth, B., and Charlesworth, D. (2010) *Elements of Evolutionary Genetics*. Roberts and Company, Greenwood Village, Colorado, USA.
- Chen, J., Glémin, S., and Lascoux, M. (2017) Genetic diversity and the efficacy of purifying selection across plant and animal species. *Mol. Biol. Evol.* **34**, 1417–1428.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2012) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **12**, 499–510.
- Ellegren, H., and Galtier, N. (2016) Determinants of genetic diversity. *Nat. Rev. Genet.* **17**, 422–433.
- Evanno, G., Regnaut, S., and Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620.
- Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.
- Francis, R. M. (2017) pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* **17**, 27–32.
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics* **28**, 3150–3152.
- Garza, J. C., and Williamson, E. G. (2001) Detection of reduction in population size using data from microsatellite loci. *Mol. Ecol.* **10**, 305–318.
- Gillespie, J. H. (2001) Is the population size of a species relevant to its evolution? *Evolution* **55**, 2161–2169.
- Hartvig, I., So, T., Changtragoon, S., Tran, H. T., Bouamanivong, S., Theilade, I., Kjær, E. D., and Nielsen, L. R. (2017) 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.
- Hey, J. (2010) Isolation with migration models for more than two populations. *Mol. Biol. Evol.* **27**, 905–920.
- Hien, V. T. T., and Phong, D. T. (2012) Genetic diversity among endangered rare *Dalbergia cochinchinensis* (Fabaceae) genotypes in Vietnam revealed by random amplified polymorphic DNA (RAPD) and inter simple sequence repeats (ISSR) markers. *Afr. J. Biotechnol.* **11**, 8632–8644.
- Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. (2009) Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9**, 1322–1332.
- Hudson, R. R., Slatkin, M., and Maddison, W. P. (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589.
- Kreitman, M., and Hudson, R. R. (1991) Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**, 565–582.
- Lamotte, S., Gajasen, J., and Malaisse, F. (1998) Structure diversity in three forest types of north-eastern Thailand (Sakaerat Reserve, Pak Tong Chai). *Biotechnol. Agron. Soc. Environ.* **2**, 192–202.
- Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Lischer, H. E. L., and Excoffier, L. (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299.
- Louys, J., and Meijaard, E. (2010) Palaeoecology of southeast Asian megafauna-bearing sites from the Pleistocene and a review of environmental changes in the region. *J. Biogeogr.* **37**, 1432–1449.
- Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12.
- Mimura, M., Yahara, T., Faith, D. P., Vázquez-Domínguez, E., Colautti, R. I., Araki, H., Javadi, F., Núñez-Farfán, J., Mori, A. S., Zhou, S., et al. (2017) Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evol. Appl.* **10**, 121–139.
- Moritsuka, E., Chhang, P., Tagane, S., Toyama, H., Sokh, H., Yahara, T., and Tachida, H. (2017) Genetic variation and population structure of a threatened timber tree *Dalbergia cochinchinensis* in Cambodia. *Tree Genet. Genomes* **13**, 115.
- Murray, M. G., and Thompson, W. F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8**, 4321–4325.
- Nei, M., and Tajima, F. (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**, 145–163.
- Nielsen, R., and Beaumont, M. A. (2009) Statistical inferences in phylogeography. *Mol. Ecol.* **18**, 1034–1047.
- Niyomdham, C., Pham, H. H., Dy Phon, P., and Vidal, J. E. (1997) Leguminosae-Papilionoideae Dalbergieae, Flore du Cambodge, du Laos et du Vietnam, Vol. 29. Muséum National d'Histoire Naturelle, Paris, France.
- Niyomdham, C. (2002) An account of *Dalbergia* (Leguminosae-Papilionoideae) in Thailand. *Thai. For. Bull.* **30**, 124–166.
- Ohtani, M., Kondo, T., Tani, N., Ueno, S., Lee, L. S., Ng, K. K. S., Muhammad, N., Finkeldey, R., Na'iem, M., Indrioko, S., et al. (2013) Nuclear and chloroplast DNA phylogeography reveals Pleistocene divergence and subsequent secondary

- contact of two genetic lineages of the tropical rainforest tree species *Shorea leprosula* (Dipterocarpaceae) in South-East Asia. *Mol. Ecol.* **22**, 2264–2279.
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., and Lercher, M. J. (2014) PopGenome: An efficient swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
- Ramos-Onsins, S. E., and Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**, 2092–2100.
- Shafer, A. B., Wolf, J. B., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., et al. (2015) Genomics and the challenging translation into conservation practice. *Trends Ecol. Evol.* **30**, 78–87.
- Simons, Y. B., and Sella, G. (2016) The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Curr. Opin. Genet. Dev.* **41**, 150–158.
- Soonhuay, P. (1994) Estimation of genetic variation in Thailand rosewood (*Dalbergia cochinchinensis* Pierre). University of British Columbia, Vancouver, Canada.
- Stibig, H.-J., Achard, F., and Fritz, S. (2004) A new forest cover map of continental southeast Asia derived from SPOT-VEGETATION satellite imagery. *Appl. Veg. Sci.* **7**, 153–162.
- Suyama, Y., and Matsuki, Y. (2015) MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. *Sci. Rep.* **5**, 16963.
- Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460.
- Tajima, F. (1989a) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
- Tajima, F. (1989b) DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model. *Genetics* **123**, 229–240.
- The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* **526**, 68–74.
- Theilade, I., and de Kok, R. (2015) The status of botanical exploration and plant conservation in Cambodia. *Cambodian J. Nat. Hist.* **2015**, 117–120.
- Toyama, H., Tagane, S., Chhang, P., Kajisa, T., Ichihashi, R., Samreth, V., Vuthy, M., Sokh, H., Katayama, A., Itadani, H., et al. (2013) Inventory of the woody flora in permanent plots of Kampong Thom and Kampong Chhnang Provinces, Cambodia. *Acta Phytotax. Geobot.* **64**, 45–105.
- UNEP-WCMC (2018) Report on species/country combinations selected for review by the Plants Committee following CoP17. UNEP-WCMC, Cambridge, UK.
- Vatanparast, M., Klitgård, B. B., Adema, F. A. C. B., Pennington, R. T., Yahara, T., and Kajita, T. (2013) First molecular phylogeny of the pantropical genus *Dalbergia*: implications for infrageneric circumscription and biogeography. *S. Afr. J. Bot.* **89**, 143–149.
- Weir, B. S., and Cockerham, C. C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.
- Yahara, T., Javadi, F., Onoda, Y., de Queiroz, L. P., Faith, D. P., Prado, D. E., Akasaka, M., Kadoyam, T., Ishihama, F., Davies, S., et al. (2013) Global legume diversity assessment: concepts, key indicators, and strategies. *Taxon* **62**, 249–266.
- Yoder, A. D., Poelstra, J. W., Tiley, G. P., and Williams, R. C. (2018) Neutral theory is the foundation of conservation genetics. *Mol. Biol. Evol.* **35**, 1322–1326.
- Yooyuen, R., Duangjai, S., and Changtragoon, S. (2011) Chloroplast DNA phylogeography of *Dalbergia cochinchinensis* Pierre in Thailand and Laos. *In* World Series vol. 30 Asia and the Pacific Workshops: Multinational and Transboundary Conservation of Valuable and Endangered Forest Tree Species. (eds.: Heok-Choh, S., Hamid, S. A., and Mei, L.), pp. 84–87. IUFRO, Kuala Lumpur, Malaysia.