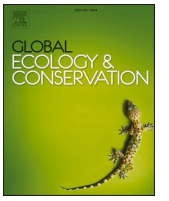
Global Ecology and Conservation 40 (2022) e02321

Contents lists available at ScienceDirect

Global Ecology and Conservation

journal homepage: www.elsevier.com/locate/gecco



Population genetic structure of wild Malayan tapirs (*Tapirus indicus*) in Peninsular Malaysia revealed by nine cross-species microsatellite markers

Qi Luan Lim [a,b,1], Christina Seok Yien Yong [a,2], Wei Lun Ng [c,3], Ahmad Ismail [a], Jeffrine J. Rovie-Ryan [d,e,4], Norsyamimi Rosli [d], Miho Inoue-Murayama [b,5], Geetha Annavi [*,a]

[a] *Department of Biology, Faculty of Science, Universiti Putra Malaysia, Serdang, Selangor, Malaysia*
[b] *Wildlife Research Center, Kyoto University, Kyoto, Japan*
[c] *China-ASEAN College of Marine Sciences, Xiamen University Malaysia, Sepang, Selangor, Malaysia*
[d] *National Wildlife Forensic Laboratory, Ex-Situ Conservation Division, Department of Wildlife and National Parks, Kuala Lumpur, Malaysia*
[e] *Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia*

ARTICLE INFO

Keywords:
Asian tapir
Conservation genetics
Cross-species microsatellite
Sex-biased dispersal
Effective population size
Central forest spine

ABSTRACT

The Malayan tapir (*Tapirus indicus*) is an endangered species endemic to Southeast Asia. Previous research have used maternally inherited, mitochondrial DNA genetic markers to investigate the population genetics of the species but not the biparentally inherited, nuclear genetic markers. However, the increasing pressure of habitat fragmentation and roadkill on the Malayan tapirs has called for an urgent need to assess the species' genetic status. In this study, the genetic diversity, population genetic structure, and sex-biased dispersal patterns of the tapir population in Peninsular Malaysia were investigated using nine cross-species microsatellite markers, using sixty-seven tapir samples (39 wild, 11 captive-born, and 17 of unknown origin) provided by the wildlife department, zoo and conservation centre. Low genetic diversity of the wild population was found based on heterozygosity and allelic richness, and a cryptic pattern of population genetic structure was inferred. One (K = 3) to two (K = 4) genetic clusters in proximity to the southwestern part of the major forest complexes seemed to be experiencing restricted gene flows compared to the other more widespread clusters or clines. Effective to census population size was estimated at 0.39–0.46. Sex-biased dispersal was not found in the current dataset. Even though these results may need cautious interpretation due to possible sampling bias, this study is the first study that investigated the genetic diversity distribution for the species using nuclear markers, and therefore will have conservation implications for the implementation of the Central Forest Spine Master Plan in Peninsular Malaysia to connect major forest complexes.

* Corresponding author.
E-mail address: geetha@upm.edu.my (G. Annavi).
[1] ORCID: 0000-0002-1499-3627
[2] ORCID: 0000-0002-1399-971X
[3] ORCID: 0000-0002-6056-7549
[4] ORCID: 0000-0003-2910-5431
[5] ORCID: 0000-0003-4249-876X

https://doi.org/10.1016/j.gecco.2022.e02321
Received 18 February 2022; Received in revised form 31 October 2022; Accepted 31 October 2022
Available online 8 November 2022

1. Introduction

The Malayan tapir (*Tapirus indicus* Desmarest, 1819; also, *Acrocodia indica* used by some authorities), locally known as *tenuk* or *cipan* in the Malay language, is a large herbivorous mammal in the family Tapiridae and the only extant tapir species in the Old World. This species is listed in the International Union for the Conservation of Nature (IUCN) Red List as 'Endangered' (Traeholt et al., 2016) and in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Its current natural habitats in Southeast Asia are found along the Tenasserim Hills, spanning from southern Myanmar and the west border of Thailand to southern Thailand and Peninsular Malaysia, and on the Sumatran Island of Indonesia (Traeholt et al., 2016). While the worldwide population size estimated by IUCN is less than 2500 mature individuals, the number of individuals within the boundary of Peninsular Malaysia has been estimated at 1300–1500, and in other habitats in Thailand-Myanmar and Indonesia at only a few hundred (Traeholt et al., 2016). Unfortunately, the number of wild tapirs in Peninsular Malaysia may be smaller than previously estimated, according to unpublished surveys carried out by the Department of Wildlife and National Parks (PERHILITAN). Their current effective range in Peninsular Malaysia has also been reduced to the Belum forest complex, greater Taman Negara forest complex, Pahang peat swamp forest, and Endau-Rompin forest complex (PERHILITAN, 2012), in contrast to a report back in 1980, which states that tapirs could be found in all states of Peninsular Malaysia (Williams and Petrides, 1980).

Two decades ago, the Malayan tapir was claimed to be safe from extinction because it was believed that they were as abundant as wild boars and barking deer in some places (Kawanishi et al., 2002). Additional factors added to this false sense of security, including that the consumption of its meat was unfavoured by the locals, only few natural predators existed, and reports of killings by natural predators, i.e., the Malayan tiger, was scarce (Kawanishi et al., 2002). Today, the conservation status of the Malayan tapir in Malaysia, considering its population size, is still comparatively more optimistic than other large mammals in the country, e.g., the Malayan tiger (fewer than 200 individuals) and the Malayan gaur (fewer than 500 individuals). Nonetheless, this endangered species requires conservation intervention to ensure its long-term survival in the plight of increasing habitat loss and forest fragmentation. *Ex-situ* conservation includes zoos and conservation centres such as the Sungai Dusun Wildlife Conservation Centre (SDWCC; originally for the captive-breeding of the Sumatran rhinoceros) and the National Wildlife Rescue Centre in Sungkai, Perak, that function to care for wild-caught, displaced, or injured tapirs. Zoos and SDWCC also contribute to the captive breeding programme of the tapirs. The Malayan tapir conservation centre under construction in the Kenaboi forest reserve in the state of Negeri Sembilan is expected to enhance the conservation of this species once established. As an *in-situ* countermeasure to habitat fragmentation, the Central Forest Spine (CFS) Master Plan for Ecological Linkages was adopted by the Federal Government of Malaysia in 2010 (Brodie et al., 2016). The master plan emphasises on the connectivity between the four main forest complexes via 37 ecological linkages to promote wildlife movements, in which Malayan tapir is one of the focal mammalian species other than the Asian elephant and Malayan tiger. However, its implementation has been challenging and the adequacy of its linkages has not been thoroughly evaluated before their designation (Jain et al., 2014). Only 85% of 28 linkages were found to provide high to acceptable connectivity based on the movements of the Asian elephants (Torre et al., 2019).

Habitat loss and fragmentation, a consequence of landscape conversions into agricultural lands and human settlements, has resulted in tapir displacement into residential areas and engagement in risky road crossings which cut through their home ranges, often with disastrous outcomes (Magintan and Traeholt, 2012; Magintan et al., 2021). For instance, an average of 8.21 Malayan tapirs were killed each year from 2006 to 2019 due to collision with vehicles and the trend is increasing (Magintan et al., 2021). Not only does habitat loss and fragmentation increase human-wildlife conflicts, it also leads to population sub-structuring of an otherwise panmictic population into smaller and isolated populations (Allendorf et al., 2013). Mammals with large body mass, being terrestrial, herbivorous, and forest dependent are found to be more negatively and consistently affected by the effects of habitat fragmentation, showing an overall decrease in genetic diversity and greater risk of extinction risk (Lino et al., 2019; Crooks et al., 2017). This situation is particularly concerning when the negative genetic consequences due to habitat modification are coupled with the direct consequences of roadkill accidents—immediate removal of animals from already small and isolated populations. This could pose great threats to reproductively non-viable populations in affected areas, such as in the increasingly fragmented forests of the southern Peninsular Malaysia. As such, the implementation of CFS and maintenance of the linkages i.e., wildlife corridors between fragmented forests is therefore of great importance to promote gene flow and maintain reproduction viability of the tapir populations. In fact, even though tapirs can survive in fragmented forest landscapes when provided suitable habitats, their occurrence, found to be negatively correlated with areas of human disturbance (Samantha et al., 2020; Linkie et al., 2013), is higher in larger forest landscapes that have more connectivity to other non-reserve forests (Samantha et al., 2020).

Changes in genetic diversity due to natural or anthropogenic barriers shape the pattern of population genetic structure. Habitat fragmentation can reduce animal movement (Allendorf et al., 2013; Lino et al., 2019; Schlaepfer et al., 2018), resulting in gene flow reduction, increasing random genetic drift, and inbreeding, which has an overall negative effect on the genetic variation of the remnant populations, such as a reduction in allelic diversity, allelic richness, observed and expected heterozygosity (Allendorf et al., 2013; Lino et al., 2019; Schlaepfer et al., 2018). Low genetic variation, loss of genetic diversity, and genetic differentiation have been found to be associated with anthropogenic habitat fragmentation (Ernest et al., 2012; De et al., 2021; Goossens et al., 2006). Nonetheless, when migration and gene flows are still permitted at sufficient levels, habitat fragmentation may not cause an evident population genetic structure (Griciuvienė et al., 2021). Levels of genetic diversity will also reflect on the effective population size (Ne), a key component in conservation biology and defined as the size of an ideal population experiencing the same rate of random genetic change (i.e., genetic drift) over time as the census population (Nc). Ne to Nc ratio is typically less than 1.0 for a wide range of taxa and reproductive scenarios (Palstra and Fraser, 2012). Ne estimates provide historical baselines to be maintained in a population, whereby

provides the prospects for the sustainability of the population (Wang et al., 2016).

Sex-biased dispersal can influence the population genetic structure of a population but is not well-understood in the Malayan tapir. Sex-biased dispersal happens when one sex disperses more readily and over greater distances than the other. Greater dispersal distance is generally related to larger body mass, and both attributes might be linked to higher reproductive success, procurement of resources, conspecific competition (both intra- and intersexual), and inbreeding avoidance (Kilanowski and Koprowski, 2016; Li and Kokko, 2019; Gros et al., 2008; Jenkins et al., 2007). Relatively uncommon among mammals, tapirs show female-biased sexual size dimorphism, for females are often larger and heavier than males by 25–100 kg (Barongi, 1986), which may suggest differing resource requirements between the sexes (Kilanowski and Koprowski, 2016). Geographic range area and home range area are also important predictors for dispersal distance (Whitmee and Orme, 2013). Tapirs have a wide home range estimated in excess of 12 km^2 that can overlap with the home ranges of other individuals including female tapirs (Williams, 1979, 1978; Campos-Arceiz et al., 2012). There is however not enough information to ascertain if female tapirs have a larger home range than male tapirs as expected from the body size differences.

Genetic assessment and conservation of the Malayan tapir populations is an important aspect highlighted by the first Malayan tapir workshop held in 2002 (Medici et al., 2003) that should be considered in formulating tapir conservation and management plans e.g., relocation, reintroduction, captive breeding etc. Since then, several population genetic studies have been carried out. For example, genetic investigations using the mitochondrial DNA cytochrome *b* (mtDNA cytB) gene (Rovie-Ryan et al., 2008; Muangkram et al., 2013) and more recently the mtDNA control region (Muangkram et al., 2017; Lim et al., 2021) have discovered two distinct clades of the Malayan tapir in the Malay Peninsula, estimated to have diverged since 1.46 million years ago. Nevertheless, the co-existence of two clades in Peninsular Malaysia but only one in the Thai captive line suggests secondary contact in the southern Malay Peninsula after isolation. Furthermore, only one haplotype is shared between the populations, indicating restricted gene flow (between the Indo-Chinese and Sundaic regions) and population genetic structuring in the Thai-Malay Peninsula but not in Peninsular Malaysia alone (Lim et al., 2021). The genetic diversity of the tapir population in Peninsular Malaysia is also found higher than the captive Thai population despite a comparable number of haplotypes (Lim et al., 2021). However, mtDNA mutations only represent genetic diversity in the maternal lineage, and therefore nuclear genetic markers such as microsatellites will be needed for investigating the nuclear genetic diversity in the Malayan tapir for a clearer picture of the species' genetic diversity and population genetic structure.

Following an unsuccessful attempt to develop microsatellite markers with random amplified microsatellite markers (Lim et al., 2019), this study attempts cross-species amplification of microsatellite markers that were developed for other tapir species. The study attempts the first genetic assessment using nuclear DNA markers on the Malayan tapirs in Peninsular Malaysia, reporting levels of genetic diversity, population genetic structure, sex-biased dispersal, and effective population size of the species. The genetic information will be useful to better inform the conservation management plans of the Malayan tapir in Peninsular Malaysia as well as other tapir populations in Southeast Asia.

2. Methodology

2.1. Sample collection and DNA extraction

A total of 83 Malayan tapir samples were collected from the Wildlife Genetic Resources Bank (WGRB) of the Department of Wildlife and National Parks (PERHILITAN), Sungai Dusun Wildlife Conservation Centre (SDWCC), and the National Zoo of Malaysia (Supplementary Table S1). Seventy-four samples were from WGRB, consisting of dried blood spots (DBS; n = 38) on FTA® Blood Card (Whatman, UK), tissue samples (n = 28; mainly muscle tissue, a few were liver and others unknown), and hair samples (n = 8). These samples were caught/collected and stored in WGRB from the period 2002–2017 by PERHILITAN during various operations including for the purpose of rescue, translocation, research, human-wildlife intervention, etc. Whole blood samples were drawn from the tapirs in SDWCC (n = 6), and at the National Zoo of Malaysia, whole blood (n = 1) and hair samples (n = 2) were obtained; these samples were collected in 2013–2018. Genomic DNA extraction was carried out with the QIAamp® DNA Mini Kit (Qiagen, Germany) following the manufacturer's spin protocol, and quantified using Quantus™ Fluorometer with ONE dsDNA dye (Promega, USA) following the recommended protocol. For hair samples, dithiothreitol (DTT) was added to a final concentration of 40 mM during the tissue lysis step. The samples were grouped by their origins: the wild-born or wild-caught animals (*WILD*), the captive-born (*CAPTIVE*) and the unknown (*Unk*); regardless of their status as living, deceased or unknown. Sex information (in the database or sexed by markers) was extracted from a previous study (Lim et al., 2020).

Table 1
Thirty-four microsatellite markers and their references or accession numbers.

Markers	References/Accession no.
Tte1, Tte5, Tte9, Tte12, Tba15, Tba20, Tba21, Tba23, Tba25	Norton and Ashley (2004)
Tter3, Tter4, Tter5, Tter7, Tter9, Tter11, Tter13, Tter14, Tter17, Tter18	Sanches et al. (2009)
TtGT021, TtGT038, TtGT048, TtGT053, TtGT070, TtGT137, TtGT138, TtGT215	Gonçalves da Silva et al. (2009)
TpGT001, TpGT002, TpGT005, TpGT009, TpGT047, TpGT062, TpGT068	FJ904319.1-FJ904323.1, FJ904325.1-FJ904326.1 (Also see Supplementary Table S2)

2.2. Screening of microsatellite markers

Thirty-four microsatellite markers (Table 1) were screened in all tapir samples, including 22 initially designed for the lowland tapir (*Tapirus terrestris*) (Gonçalves da Silva et al., 2009; Norton and Ashley, 2004; Sanches et al., 2009), five for the Baird's tapir (*Tapirus bairdii*) (Norton and Ashley, 2004), and seven designed in this study (Supplementary Table S2, with Primer3Plus (Untergasser et al., 2007); primer size set to 18–24 bp, optimum 20 bp; melting temperature, Tm set to 57–63 °C, optimum 60 °C; and primer GC% set to 40–60%, optimum 50%) in this study from DNA sequences containing microsatellites of the mountain tapir (*Tapirus pinchaque*) that were deposited in the National Center for Biotechnology Information (NCBI) GenBank database (FJ904319.1-FJ904323.1, FJ904325.1, FJ904326.1) but however with no primer design described elsewhere.

Polymerase chain reaction (PCR) amplification was run in 10 μL-singleplex reactions containing 5 μL 2 ×MyTaq™ Red Mix (Bioline, Germany), 0.3 μM of both forward and reverse primers, and 2–20 ng gDNA, with a touchdown PCR profile on a T100™ Thermal Cycler (Bio-Rad, USA). The PCR profile consisted of a pre-denaturation step at 95 °C (3 min), 40 cycles of three-step touchdown PCR: 1) denaturation step at 95 °C, 2) annealing step at 65→55 °C (−1 °C per step for the first 10 cycles), and 3) extension step at 72 °C (each step 20 s), followed by a final extension step at 72 °C (3 min). The PCR amplicons were first examined on a 2% agarose gel prepared with 1 ×TAE buffer (40 mM tris, 20 mM acetate, and 1 mM EDTA) and pre-stained with RedSafe™ Nucleic Acid Staining Solution (iNTRON, South Korea) at 1 μL:20 mL agarose solution. Microsatellite loci that were positively and consistently amplified, yielding single band or at most two bands < 50 bp apart in the range of 100–400 bp, were retained as the candidate markers.

The NCBI sequences of the loci (see Table 2) were searched against the Malayan tapir's chromosome-length Hi-C genome assembly (Dudchenko et al., 2017) using BLAST+ (Camacho et al., 2009). Sex-linked chromosome-length scaffolds (HiC_scaffold_26 and HiC_scaffold_27) in the genome assembly were identified via the Blast search with the sequences of X-linked (DQ519376.1, MN786409.1) or Y-linked (KM347952.1) zinc finger gene sequences in the Malayan tapir. Only microsatellite loci that were not sex-linked and contained a motif (regardless of the motif observed in the original taxon) with at least five repeats for dinucleotide motifs and four repeats for tri- and tetranucleotide motifs, perfect or imperfect, were retained for analysis.

2.3. Genotyping of microsatellite marker loci

For each microsatellite marker, either the forward or reverse primer amplifying the marker was modified by extending the 5′-end to contain any of three tail sequences: Neomycin rev (*6-FAM*), Hill (*HEX*), and T7term (*ROX*), in order to label PCR amplicons in a cost-effective way (Vartia et al., 2014; Culley et al., 2013). Primer-dimer and hairpins of all possible combinations of modified primers, unmodified primers, and tail primers were checked on AutoDimer 1.0 (Vallone and Butler, 2004) to aid the selection of tail sequences.

Singleplex or multiplex PCR reactions (for details of multiplex PCR see Supplementary Table S3) were prepared for each sample in a final volume of 10 μL containing 5 μL 2 ×MyTaq™ Red Mix (Bioline, Germany), 0.3–0.4 μM each of unmodified primer and tail primer, 0.075–0.1 μM modified primer, and 2–6 ng gDNA. PCR was performed with a touchdown profile like the profile described above, however, the denaturation, annealing and extension steps were extended to 30 s, 60 s, and 30 s respectively, and the final extension step was increased to 15 min. The PCR amplicons were then pooled post-PCR, and the fragment sizes were analysed on an ABI3730XL (Applied Biosystems, USA) using LIZ500 as the size standard. Genotype scoring was manually done in Peak Scanner V1.0 (Applied Biosystems, Carlsbad, CA).

Table 2
Detected and reported motif for sixteen candidate microsatellite loci, and the name of the scaffold hit with the query sequences in a Blast search against the Hi-C chromosome-length genome assembly of the Malayan tapir (dnazoo.org).

Locus	Reported motif	Detected motif	Accession no. of query sequence	Subject Scaffold	Bit score	Ref.
Tte12	$(AC)_{19}$	$(AC)_2AT(AC)_5$	KM275223	HiC_scaffold_7	174	Norton and Ashley (2004)
Tter3	$(CATT)_6$	$(CATT)_3CAAT(CATT)$	GQ141515	HiC_scaffold_3241	926	Sanches et al. (2009)
Tter4	$(TG)_{25}$	$(TG)_{12}CGCG(TG)_7$	GQ141516	HiC_scaffold_2	933	
Tter7	$(AATG)_5$	(AATG)3AATAAACG	GQ141518	HiC_scaffold_1	407	
Tter9	$(CAGG)_7$	$(TG)_6CC(TG)_{10}$	GQ141519	HiC_scaffold_24	357	
Tter13	$(CA)_{20}$	$(CA)_{14}(CTCA)_9$	GQ141521	HiC_scaffold_1	983	
Tter14	$(CA)_{22}$	$(CA)_{10}CG(CA)_{20}$	GQ141522	HiC_scaffold_1	841	
Tter17	$(TC)_{29}$	$(TC)_3TG(TC)_4CC$	GQ141523	HiC_scaffold_6	612	
Tter18	$(CA)_7$	$(CA)_7AA(CA)_5TCTA(CA)_4$	GQ141524	HiC_scaffold_1	667	
Tba23	$(AC)_{14}$	$(AC)_{15}$	KM275224	HiC_scaffold_24	291	
Tba25	$(AC)_{17}$	$(AC)_{15}AA(AC)_3$	KM275226	HiC_scaffold_2	211	
TtGT021	$(AC)_{13}$	$(AC)_{14}$	FJ904298	HiC_scaffold_5	492	Gonçalves da Silva et al. (2009)
TtGT048	$TT(GT)_{12}$	$(GT)_3CT(GT)_8$	FJ904301	HiC_scaffold_11	446	
TtGT070	$(CT)_2CATA(CA)_{16}$	$(CT)_2CATA(CA)_{13}$	FJ904304	HiC_scaffold_1	1029	
TtGT137	$(GT)_{17}$	$(GT)_6AT(GT)_6$	FJ904309	HiC_scaffold_6	686	
TtGT215	$(CA)_7CT(CA)_4CTCA$	$(CA)_5TA(CA)_2CTCA$	FJ904318	HiC_scaffold_3	527	

No duplicate hits with alignment length > 80 bp and bit score > 100 for any locus

2.4. Genotyping error

Error rate per allele was quantified as a measurement of genotyping error following the scheme of Pompanon et al. (2005). Genomic DNA of six samples were extracted for the second time as replicates for re-genotyping, and a separate five pairs of DNA samples (of different sample types) from five tapirs were genotyped to estimate the genotyping error. For these 11 individuals, two to five PCR replicates per locus were genotyped. A reference genotype (one with the highest frequency) was chosen among the candidate genotypes for each individual and locus. Unweighted mean error rate per allele for each locus was calculated by averaging the error rate per allele (calculated as the number of mismatched alleles against the reference genotype ÷ total number of alleles typed) across the 11 individuals. In addition, a serial dilution of a blood DNA sample to 20, 2.0, 0.2, 0.1, 0.02 ng/μL was prepared to examine the effect of DNA template amount to the genotyping error. Micro-checker 2.2.3 (van Oosterhout et al., 2004) was used to check for potential null allele, large allele dropout and scoring errors in all the groups.

2.5. Marker characterisation and genetic analyses

Allelic frequencies, number of alleles (Na), number of effective alleles (Ne), Shannon's information index (I), expected (He) and observed heterozygosity (He) were computed in GenAlEx 6.503 (Peakall and Smouse, 2012) for all polymorphic markers. Polymorphism information content (PIC), probability of identity (P_{ID}) and P_{ID} among siblings ($P_{ID\text{-}SIB}$) was computed using Cervus 3.0.7 (Kalinowski et al., 2007). The R package *hierfstat* 0.5–10 (Goudet, 2005), run in R version 4.0.2 in RStudio 1.1.447 (RStudio Team, 2021; R Core Team, 2021), was used to compute these genetic diversity indices: rarefied allelic richness (Ar; to account for variation in sample size), which was followed by a non-parametric Kruskal-Wallis Test to test the between-group difference; and inbreeding coefficients (Fis; Nei, 1987) with 1000 permutations in SpaGeDi 1.5 (Hardy and Vekemans, 2002). GenAlEx's G-Statistics was used to compute pairwise Fst (Nei, 1977), G″st (Meirmans and Hedrick, 2011), and Dest (Jost, 2008), permuted 999 times to test for genetic differentiation. These statistics were repeated for the genetic clusters identified from two clustering methods (see below). Pairwise relatedness (*r*) based on method of Lynch and Ritland (Lynch and Ritland, 1999) implemented in GenAlEx was computed. The R package *poppr* 2.9.2 (Kamvar et al., 2015, 2014) was used to plot a genotype accumulation curve, which assesses the power of a random set of markers in discriminating between multilocus genotypes (MLGs) and unique individuals. This was done by random sampling one to N-1 marker without replacement for 1000 iterations and counting the number of MLGs observed. Tests for Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were performed in Genepop 4.7.5 (online version available at http://genepop.curtin.edu.au/;Raymond and Rousset, 1995; Rousset, 2008), using probability tests based on Markov chain (MC) algorithm (dememorization = 1000, batches = 100, iterations per batch = 1000), followed by global tests using Fisher's method. Particularly, HWE was tested using the complete enumeration method, which is optional only for loci with not more than four alleles to compute an exact probability value (recommended by Genepop for sample with < 1000 individuals). The tests were performed in each group (*WILD, CAPTIVE* and *Unk*) and across all samples (*WILD*+*CAPTIVE*+*Unk*).

2.6. Population genetic structure

A few approaches were used to identify the population genetic structure of the Malayan tapir population in Peninsular Malaysia using only the *WILD* samples. First, principal coordinate analysis (PCoA) was performed on GenAlEx using genetic distance matrix generated by the 'distance-standardised' method to investigate the distribution of the *WILD* samples, and in addition, *CAPTIVE* and *Unk* samples for a view of their genetic distribution, on the first and second principal components (PCs).

Second, the number of genetic clusters (K) was inferred using the Bayesian clustering method implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000), using both no admixture and admixture models run separately under correlated and uncorrelated model assumptions. The admixture model was tested because the Malayan tapir population in the southern part of Malay Peninsula may have descended from two distinct populations in the past, as detected by the mtDNA control region (Lim et al., 2021). The parameter POPFLAG was set to 0. For runs with admixture models, an alternative ancestry prior or initial *alpha* = 0.1 was used to account for potential unbalanced sampling as suggested by (Wang, 2017). The models were run without prior information on sampling location. The analysis was performed for 20 iterations for each value of K from 1 to 15, with a 1×10^4 burn-in period and 1×10^5 Markov chain Monte Carlo (MCMC) repeats after the burn-in period. The other parameters were set to default. The R package *pophelper* (Francis, 2017) was used to calculate the distribution of delta K using Evanno's method to infer the number of K, average the membership coefficient across iterations, and to create the bar plots.

Third, genetic clusters were determined using K-means clustering method without prior assumption of subpopulations and analysed by the non-model-based discriminant analysis of principal component (DAPC) in the R package *adegenet* 2.1.3 (Jombart, 2008, 2011). Using the function *find.cluster*, ten PCs were retained. In the resulting graph of the Bayesian Information Criterion (BIC) curve, the optimal number of clusters, K, were chosen from around the elbow of the curve. However, the "elbow" can span a wide range of K values, and so to prevent data overfitting when a large K value was chosen, an Average Silhouette plot was generated using the function *fviz_cluster* in the R package *factoextra* 1.0.7 (Kassambara and Mundt, 2020). The input data matrix of pairwise Euclidean distances between samples was computed with the function *bitwise.dist* in the R package *poppr*, and the optimum K value was inferred when the cluster number maximised the Average Silhouette Width. The genetic clusters were then analysed by DAPC using the function *dapc* in *adegenet*, after retaining the optimum number of PCs determined by the function *xvalDapc* following the suggestions of Miller et al. (2020). To predict the memberships of tapirs in the *CAPTIVE* and *Unk* groups, the function *predict.dapc* was used.

Geographic distribution of the genetic clusters identified by the K-means clustering method was drawn in QGIS 3.20 (https://www.

qgis.org/) on a base map (downloaded from http://gadm.org/; Database of Global Administrative Areas) including land cover data retrieved from the European Space Agency Climate Change Initiative (ESA CCI) Land Cover project (2015 data; http://www.esa-landcover-cci.org). Land cover classes follows the System of Environmental-Economic Accounting (SEEA) 2012 classification but further collapsed to fewer classes: urban and associated areas, crop land (herbaceous, woody, layered crops), sparsely natural vegetated areas, natural vegetated areas (grassland and shrub-covered areas including aquatic or regularly flooded areas), tree-covered areas, mangroves, terrestrial barren land, and inland water bodies. In addition, 30 m Shutter Radar Topography Mission (SRTM) digital elevation models (DEMs) between longitudes 99.0–105.0 and latitudes 1.0–7.0 were downloaded from the NASA Earth Data server via the SRTM-Downloader plugin in QGIS to create an elevation map (Nasa, 2013). Sampling bias in the dataset may result from uneven sampling over spatial and temporal scales (Supplementary Fig. S1a). To examine the relationship between the sampling year and geographic distance, Moran's *I* was computed using the function *moran.test* and the significance was tested with 999 Monte-Carlo simulations using the function *moran.mc*, both implemented in the R package *spdep* (Bivand, 2022).

To delineate possible cryptic patterns of population structure, the spatial model using Bayesian clustering method implemented in GENELAND (Guillot and Santos, 2009; Guillot et al., 2005a, 2005b; Guillot, 2008) was used for the identification of subpopulations and geographic boundaries of the clusters, using both spatial coordinates and microsatellite genotypes of the *WILD* tapirs. Based on the results from K-means method, the range of K was narrowed down to K = 1–10 to save computation time to determine a K solution from 20 independent runs each consisting of 1×10^5 MCMC iterations thinning at every 100th iteration. The run with the highest average posterior probability after a 200 burn-in period was used as well as the inferred number of K. Correlated allele frequency model was assumed. Uncertainty on coordinate was set to 1 for the tapirs are mobile and use a large home range.

A spatial autocorrelation was also carried out in GenAlEx for the same sample set. An even distance class with a size of 15 km in 25 classes, which was selected considering the moving distance of the Malayan tapir (Mahathir et al., 2017). A positive spatial autocorrelation coefficient, *r*, as predicted under restricted dispersal, was inferred when: i) the probability of a *r*-value from 999 random permutations (under the assumption of no spatial structure) achieved a value greater or equal to the observed *r*-value was < 0.05 in a one-tailed test, and ii) when the 95% confidence interval (CI; from 1000 bootstrapping) did not straddle $r = 0$.

2.7. Sex-biased dispersal

The sex information of the *WILD* samples (14 male and 22 female tapirs) were retrieved from (Lim et al., 2020). A spatial autocorrelation analysis, as described above, treating the sexes as two different populations was carried out to test for sex-biased dispersal (Supplementary Fig. S2). Difference in spatial genetic structure between the two sexes for each distance class was tested with paired-sample t-tests. Heterogeneity between the spatial correlograms of the two sexes was predicted when $p < 0.01$ for the correlogram wide observed Omega $\geq$ random Omega (Banks and Peakall, 2012). In addition, mean corrected assignment indices (mAIc) were estimated in GenAlEx for both male and female individuals; the dispersive sex is predicted to show a lower value (Mossman and Waser, 1999). The significance of mAIc was then tested using a nonparametric Mann Whitney U-test in GenAlEx.

2.8. Effective population size

The effective population size (Ne) in *WILD* tapirs was estimated using the molecular co-ancestry (Nomura, 2008) and LDNe method with random-mating model (Waples and Do, 2008) implemented in the software NeEstimator v2.1 (Do et al., 2014), with or without omitting loci that are in LD that would therefore violate assumptions of the methods used and may cause bias in the Ne estimation (Wang et al., 2016).

Lastly, all plots other than those directly output from various standalone softwares and specific R packages were made using the R package *ggplot2* (Wickham, 2016). Adobe Illustrator CC 2019 (Adobe Inc.) was used for editing and arranging the graphics.

3. Results

3.1. DNA samples

DNA samples that consistently or frequently failed in PCR amplification were excluded from further analysis (n = 11). Concentrations of the remaining DNA samples ranged from 36.0 to 339 ng/μL for whole blood, 0.143–3.87 ng/μL for DBS, 4.272–230.4 ng/μL for tissue samples, and 0.246–3.10 ng/μL for hair samples. From studbooks and the Wildlife Genetic Resources Bank (WGRB) database, five of the whole blood samples collected from the zoos were identified as coming from the same tapirs from whom separately collected samples (all DBS) were stored in WGRB, i.e., TAP76, TAP80, TAP94, TAP105, and TAP107. These duplicated samples were used for estimating the genotyping error. Among the remaining 67 samples, eleven were collected from captive-born animals (*CAPTIVE*), 39 from wild animals (*WILD*), and 17 from animals of unknown origin (*Unk*). See Supplementary Table S1 for a summary of samples used in this study.

3.2. Marker characterisation

Sixteen candidate primer pairs, 14 from the lowland tapir and two from the Baird's tapir, were retained after screening for successful amplifications, repeat motifs that met the minimum criteria, and interpretable peak patterns. The seven primer pairs (prefix *TpGT*) designed in this study either yielded smear or no amplification and therefore were discarded. Many of the microsatellite motifs

in the candidate markers were interrupted or different from the reported ones, while no marker locus was found to be associated with the sex-linked chromosome-length scaffolds in the Hi-C genome assembly (HiC_scaffold_26 or HiC_scaffold_27) of the Malayan tapir (Table 2).

Nine of the microsatellite loci, *Tter4*, *Tter9*, *Tter13*, *Tter14*, *Tter17*, *TtGT021*, *TtGT070*, *Tba23*, and *Tba25* (Supplementary Table S4), genotyped in 67 tapirs carried at least two alleles and were considered polymorphic (for details and a genotype table, see Supplementary Table S5 **and** Table S6). Micro-checker found no evidence of null alleles, allele dropouts, or scoring errors in the samples (not in *CAPTIVE* or *WILD* groups), but suggested the presence of null alleles for the locus *TtGT021* in the *Unk* group. Two false allele drop-ins were recorded (each one in *TtGT021* and *Tter14*) at ≤ 0.02 ng gDNA of a serial dilution from a blood DNA sample, illustrating that the PCR protocols remained robust in amplifying a low amount of DNA (≥ 0.1 ng). In the eleven replicated samples (six re-extracted DNA and five duplicated samples from different biological materials) for which PCR were replicated at least twice, the unweighted means of error rate per allele calculated were 3.03% for *Tter9*, 2.42% for *TtGT021*, 3.79% for *Tter13*, and 2.27% for *Tter14*. No error was detected in the remaining markers.

More than 90% of MLGs (53 of 59) could be identified with eight or more microsatellite markers (Fig. 1). The distribution table of the 59 MLGs in the three sample groups is shown in Supplementary Fig. S3. Identical MLGs were found to be shared by three pairs of tapirs in *WILD*, one in *CAPTIVE*, and three in *Unk*. The three pairs in *WILD* (TAP104-TAP107, TAP59-TAP67, and TAP02-TAP10) were respectively identified as different individuals based on differences in microchip numbers, locality, and given name (information not shown). The only identical pair (TAP01-TAP31) in *CAPTIVE* were given different names. Only the first pair in the three pairs of identical samples in *Unk* (TAP05-TAP63, TAP06-TAP09, and TAP68-TAP69) can be differentiated by different sex. While P_{ID} shows that using nine microsatellite markers may be sufficient to distinguish different individuals (< 0.01), the same set of markers is not sufficient to correctly distinguish siblings ($P_{ID\text{-}SIB} > 0.01$), according to the threshold suggested by Waits et al. (2001) (Supplementary Table S7 **and** Fig. S4).

Deviations from HWE ($p < 0.05$) were only found in the loci *TtGT021* and *Tter14* for *Unk* group, and in *Tter14* when all groups were pooled (Supplementary Table S7). LD ($p < 0.05$), the non-random association of alleles from different loci, was detected in three marker pairs *Tter9-Tter14*, *TtGT070-Tba25*, and *Tter17-Tter14* for the *WILD* group; and in two marker pairs *TtGT021-Tter4* and *Tter9-Tter14* for the *Unk* group (Supplementary Table S8). However, a Blast search against the chromosome-length genome assembly (Table 2) indicated that none of the marker pairs in LD were located on the same chromosome-length scaffold, and neither was LD detected in marker pairs that are found on the same scaffold, so therefore all markers were assumed to be unlinked. Since deviations from HWE and occurrence of LD may indicate the occurrence of non-random mating in the *WILD* population, such as population structuring, the markers were retained for subsequent genetic analyses.

3.3. Genetic diversity

Genetic diversity indices across nine microsatellite loci for each group were tabulated in Table 3 (full list in Supplementary Table S7). Lower Ho to He (overall, ranged 0.295–0.333) was found across all samples and in groups except *CAPTIVE*, as well as low rarefied Ar (< 2.5 alleles) in all categories. Significantly positive Fis was found in the *WILD* group (0.115, $p < 0.05$). Both *WILD* and *CAPTIVE* groups were in Hardy-Weinberg equilibrium but not the *Unk* group. No significant difference in rarefied Ar was found between the groups ($X^2 = 1.006$, $p = 0.605$), although Ar was slightly higher in *WILD* and *Unk* groups compared to the *CAPTIVE* group. Private alleles were found in *WIlD* (n = 4) and *Unk* (n = 1). The most informative marker was *Tter14*, carrying seven alleles and 2.639

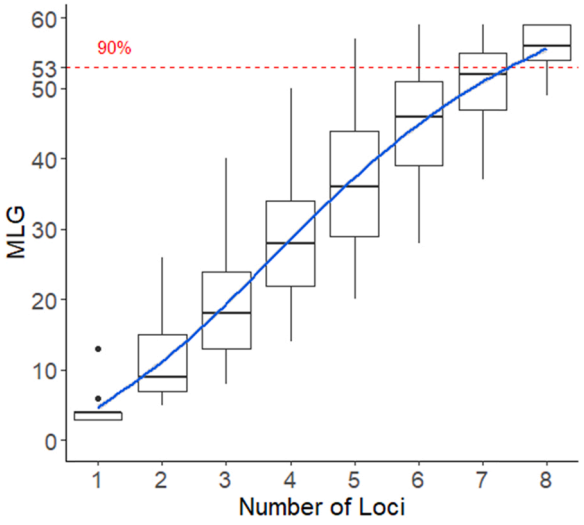


Fig. 1. Genotype cumulative curve for nine microsatellite markers characterised in 67 Malayan tapirs. More than 53 (90%) of the total number of multilocus genotypes (MLGs, n = 59) could be distinguished using eight or more random markers from the set.

Table 3
Genetic indices and characterisation of nine microsatellite loci in the *WILD*, *CAPTIVE*, and *Unk* (unknown) groups of Malayan tapirs and across all samples.

Group		n	Ar	Na	Pa	Ho	He	Hs	Fis	HW	PIC
WILD	**Mean**	39	2.316	3.000	4	0.299	0.333	0.337	**0.115***	0.307	0.286
	SE		0.267	0.527		0.070	0.071	0.072	(0.036)		0.029
CAPTIVE	**Mean**	11	2.111	2.111	0	0.333	0.295	0.309	-0.084	1.000	0.240
	SE		0.309	0.309		0.092	0.076	0.080	(0.448)		0.055
Unk	**Mean**	17	2.273	2.444	1	0.301	0.317	0.327	0.083	**0.046***	0.271
	SE		0.313	0.338		0.087	0.077	0.079	(0.336)		0.049
All	**Mean**	67	2.297	3.111	–	0.305	0.328	0.330	0.076	0.075	0.280
	SE		0.280	0.512		0.077	0.073	0.073	(0.059)		0.023
G-statistics			**Fst**	**P(rand ≥ data)**		**Gst″**	**P(rand ≥ data)**		**Dest**	**P(rand ≥ data)**	
WILD-CAPTIVE			0.016	0.424		0.003	0.376		0.001	0.386	
WILD-Unk			0.010	0.575		-0.005	0.559		-0.002	0.567	
CAPTIVE-Unk			0.017	0.570		-0.007	0.546		-0.002	0.554	

Ar, allelic richness per locus (number of alleles per locus) estimated and standardised to the smallest sample size (n = 11) by rarefaction; Na, number of alleles; Pa, private alleles; Ho, observed heterozygosity; He, expected heterozygosity; Hs, within population genetic diversity (Nei, 1987); Fis, inbreeding coefficient, 1−(mean Ho/mean Hs), significant difference from Fis = 0 tested by 1000 permutations (*p*-value shown in brackets); HW, *p*-value of Hardy-Weinberg test; PIC, polymorphism information content; Fst, measure of genetic differentiation (Nei, 1987); G″st, Hedrick's standardised Gst (Meirmans and Hedrick, 2011); Dest, Jost's estimate of differentiation (Jost, 2008); P(rand ≥ data): probability of random data ≥ observed data based on 999 permutations
*p-values significant at 95% confidence level ($p < 0.05$)

effective alleles, with the highest I value of 1.200 and PIC value of 0.558. The least informative marker was *Tba23*, carrying three alleles and 1.030 effective alleles, with the lowest I value of 0.088 and PIC value of 0.029.

Type of kinships namely full-sibs ($0.25 \leq r < 0.5$) and half-sibs ($0.125 \leq r < 0.25$) for each sample pair was determined based on the mean values of pairwise genetic relatedness (r). The identical twins ($r = 0.5$) were considered as full-sibs (identical full-sibs) since the estimated $P_{ID\text{-}SIB}$, the probability that two siblings drawn at random share the same MLGs, was higher than 0.01, and that tapir twins are extremely rare. Results are summarised in the Supplementary Table S9. In the full dataset, out of 2211 pairwise comparisons, 18.09% might share kinships. The trends of kindships proportions were similar for the *WILD* and *Unk* groups (identical full-sibs < full-sibs < half-sibs). In *CAPTIVE*, full-sibs constituted the highest proportions (7.27%) followed by half-sibs (5.46%), and identical full-sibs (1.82%); there were three full-sibs tapirs sharing the same parents (identified by individual number, not shown). In *WILD*, 17.41% of 741 pairwise comparisons might share possible kinships (11.74% half-sibs, 5.26% full-sibs, and 0.41% identical full-sibs).

3.4. Population genetic structure

The first two PCs in the PCoA plot explained a total of 36.9% of the genetic variation in the samples (n = 67). Although a clear manifestation of genetic isolation was not apparent, subgroups might be differentiating along the first PC axis (Supplementary Fig. S5). Genetic variation in the *Unk* group was fairly evenly distributed around the *WILD* group, while the *CAPTIVE group* was more concentrated in the middle along the PC1 axis. A circular bar plot of membership probabilities (K = 3; identified using K-means method. See below) for the whole dataset, in which memberships of *CAPTIVE* and *Unk* were predicted based on the *WILD* dataset ("training data"), is shown in Supplementary Fig. S6. No apparently unique genetic cluster was detected within the *CAPTIVE* and *Unk* groups, all three clusters from the *WILD* group were represented, and no significant genetic differentiation based on the G-statistics (Fst, G″st, and Dest; Table 3). Therefore, for the subsequent analyses, only the *WILD* dataset (n = 39) was used.

In STRUCTURE, the number of genetic clusters, K = 2 to K = 7, were inferred from a combination of admixture/no admixture and correlated/uncorrelated models. Graphs of delta K and bar plots of membership coefficient to ancestral populations are shown in Supplementary Fig. S7. However, no population genetic structure could be inferred from the bar plots.

The BIC curve for K-means clustering (Fig. 2**a**) did not show a clear elbow; the lowest BIC value was at K = 15, which was large and probably not a realistic K value for a sample size of 39 wild tapirs. Instead, the optimum cluster number (K = 3) suggested by Average Silhouette plot was used (Fig. 2**a**). DAPC analysis for K = 3 was performed, keeping all DA eigenvalues and three PCs respectively, as determined by the cross-validation steps (Supplementary Fig. S8). The number of PCs kept in the DAPC analysis accounted for 54.2% of the total genetic variance in the data, achieving 96.4% mean successful assignment (Supplementary Fig. S9). Thirty-six *WILD* tapirs (the other three missing location information; see Supplementary Table S1) were extracted to examine the spatial genetic structure in Peninsular Malaysia (Fig. 2**a**). The genetic clusters (K-C1, K-C2, and K-C3) in the DAPC plots were distinctly separated.

Bayesian clustering method implemented in GENELAND, using the same set of 36 *WILD* tapirs and based on spatial model, detected K = 4 in 15 out of 20 independent runs, including the run with the highest average posterior probability (log probability = −220.19). Probability density of the cluster solution (K = 4) was 31% after removing the first 200 × 100th iterations from the trimmed MCMC chains that has converged at around 10,000th (Fig. 2**b**). Across the 20 runs, G-C1 and G-C2 were more consistently detected, although there could be slight changes in the cluster members; posterior memberships of G-C3 and G-C4, on the other hand, changed more readily and so did the distribution, although general pattern remains similar. When compared to the cluster distribution inferred from the K-means clusters (Fig. 2**a**), an additional cluster G-C2 could be identified to the south of G-C1, a cluster resembling K-C1 (Fig. 2**b**).

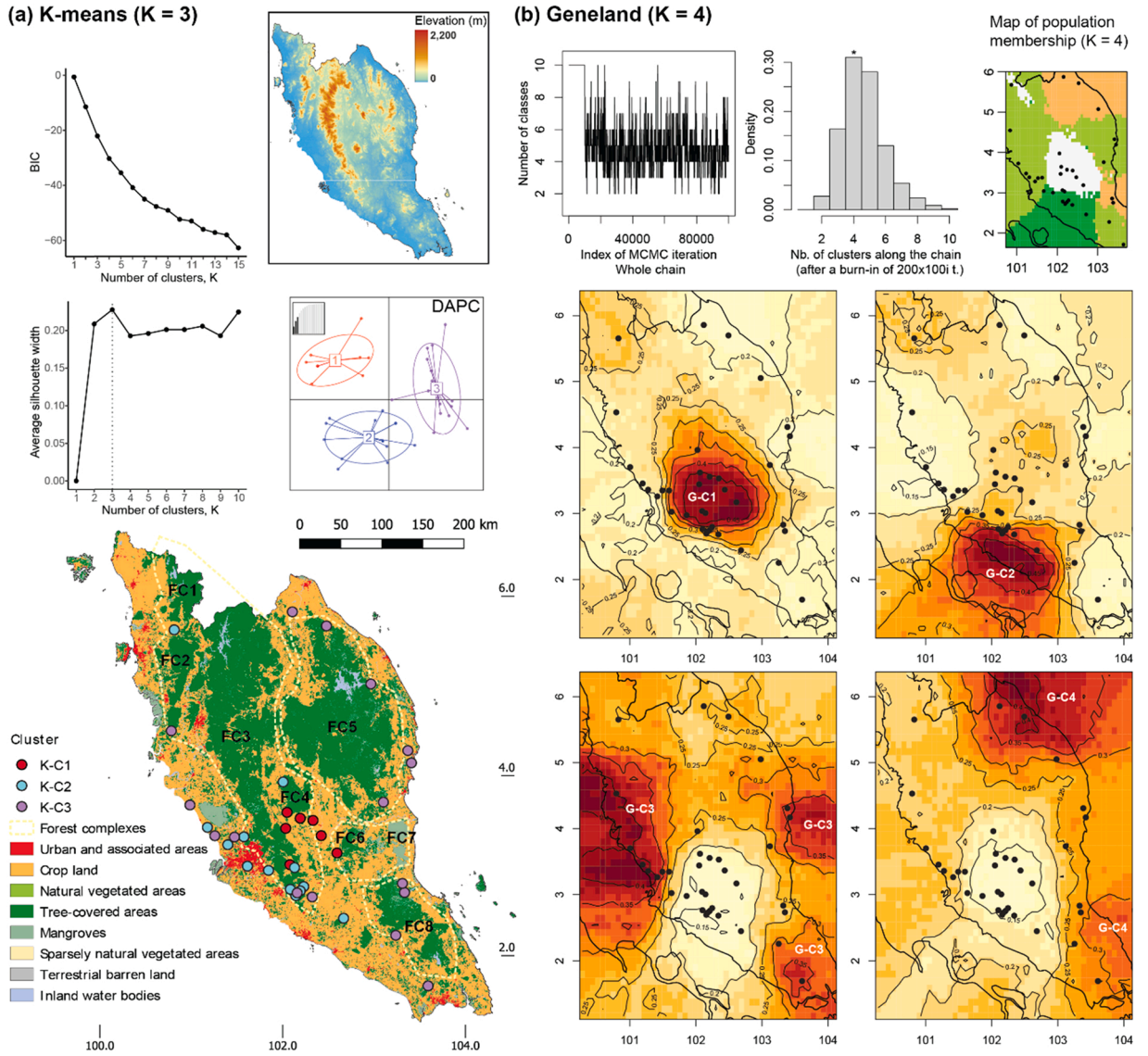


Fig. 2. Genetic clusters of the wild Malayan tapirs identified by nine microsatellite loci using K-means clustering method implemented in the R package *adegenet* and Bayesian-based spatial model implemented in the software GENELAND. (a) K-means clustering algorithm identified K = 3 in 39 wild tapirs based on the curve of Bayesian Information Criterion (BIC) and Average Silhouette Width. Discriminant analysis of principal component (DAPC) shows three distinct genetic clusters (K-C1 to K-C3), for which spatial distribution of 36 tapirs was plotted on a map of Peninsular Malaysia. (b) Spatial Bayesian clustering model in GENELAND with an optimum number of clusters (K = 4) was determined in 36 wild tapirs after a burn-in of 200 × 100th iteration from a total of 100,000 Markov chain Monte Carlo (MCMC) chains. Higher posterior memberships to each of the four clusters (G-C1 to G-C4) are represented by darker colours. The four main complexes (yellow dotted lines, from top left to bottom) are the Banjaran Titiwangsa-Banjaran Bintang-Banjaran Nakawan, Taman Negara-Banjaran Timur, South-East Pahang-Cini and Bera Wetlands, and Endau-Rompin National Park-Kluang Wildlife Reserve. The eight forest sub-complexes are the Kedah Singgora (FC1), Bintang Hijau (FC2), Main Range (FC3), Benom (FC4), Greater Taman Negara (FC5), Cini-Bera (FC6), South-East Pahang (FC7), and Endau-Rompin-Sedili (FC8) Forest Complex. A simple map coloured by elevation is also provided in the box on top.

The other two clusters, G-C3 and G-C4, were more widespread but distribution pattern disrupted and restricted from the boundaries of G-C1 and G-C2. In between G-C1 and G-C2, a strong gene flow barrier over a relatively short distance could also be inferred.

A correlogram of spatial autocorrelation (Fig. 3**a**) found a positive correlation between the genetic distance and geographic distance ($p < 0.05$) in distance classes 15–30 km (n = 21), 30–45 km (n = 24), and 45–60 km (n = 31). The first x-intercept (82 km) provided an estimate of the extent of the positive autocorrelation. Average sample size in each of the distance classes was 24.72 ± 11.41. The Moran's *I* found no significant correlation (Moran's $I = -0.093$, $p = 0.986$) between the sampling year and spatial distribution (Supplementary Fig. S1b) of the 32 *WILD* tapirs (whose sampling year information were available), suggesting limited effect of spatiotemporal correlation on the observed spatial genetic structure.

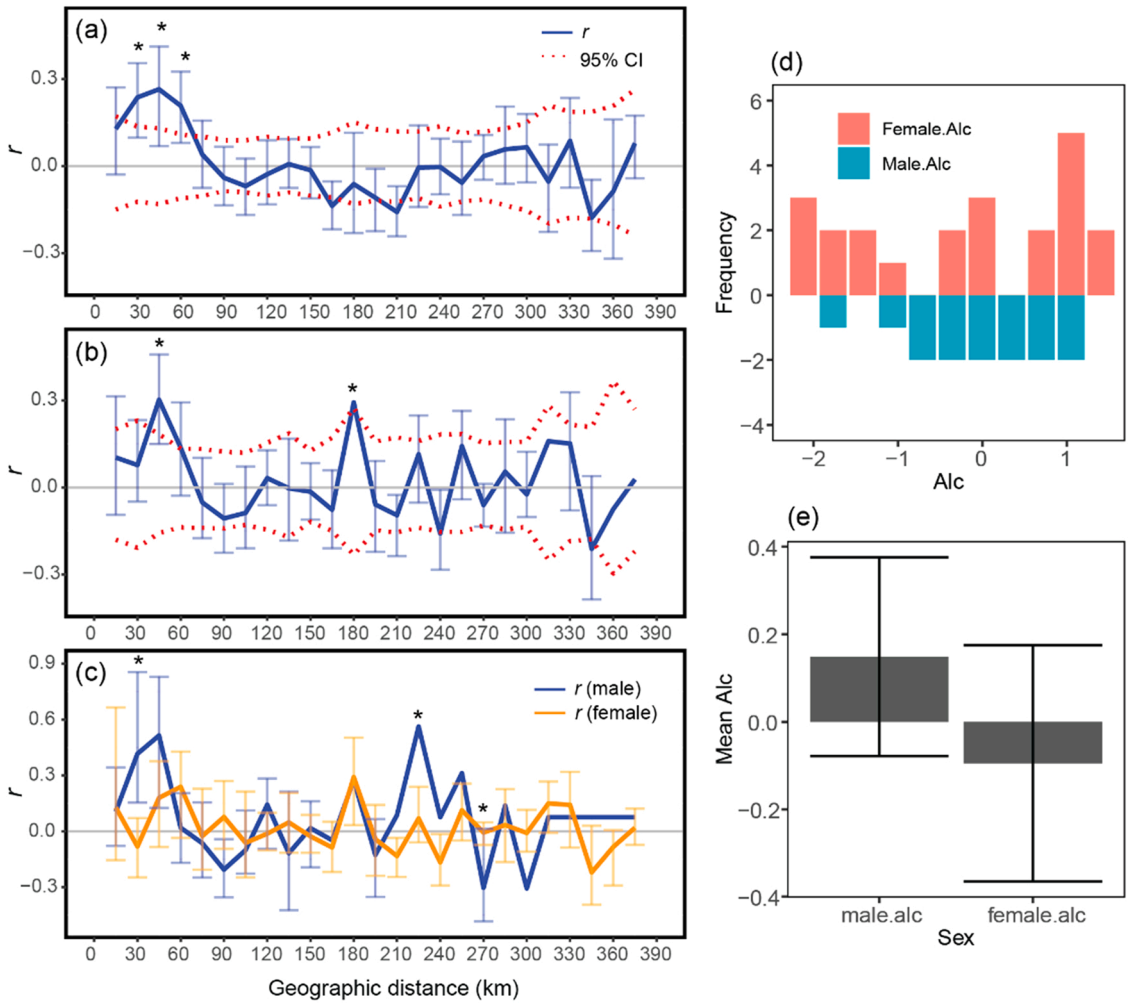


Fig. 3. Spatial autocorrelation analysis and assignment index for 36 wild Malayan tapirs consisting of 14 males and 22 females. (a) Correlogram of the average autocorrelation coefficient, *r,* for 25 distance classes each spans 15 km showing spatial autocorrelation in 36 samples (upper); (b) Combined correlogram and (c) separate correlograms of males and females for testing sex-biased dispersal. The dashed lines U and L represent the upper boundary and lower boundary of the distribution of *r* under the assumption of no spatial structure (999 permutations), respectively. The error bars represent the 95% confidence limits of *r* bootstrapped 1000 times (not shown for classes with ≤ 5 samples). The x-intercept of line *r* is 82.469 km; (d) Frequency distribution of AIc (corrected assignment index) values for males and females (upper); (e) mean AIc (mAIc) with error bars representing the 95% confidence limits (lower). Asterisks indicate significant positive spatial autocorrelation (see methods) or significant paired-sample t test between sexes.

Genetic indices for the genetic clusters are shown in Table 4. First, the genetic diversity based on rarefied Ar was low (< 2.5 alleles) and the clusters was not significantly different from each other ($X^2 = 0.231$, $p = 0.891$ for K-means clusters; $X^2 = 1.639$, $p = 0.651$ for GENELAND clusters). The polymorphism levels were also low for this set of markers (PIC < 0.5). Permutations on Fis in K = 3 (−0.202 to 0.063) and K = 4 (−0.174 to 0.081) did not support significant inbreeding or outbreeding in any of the genetic clusters identified (Table 4). The clusters, however, were found to be significantly different from each other based on either Fst, G′′st, or Dest ($p < 0.05$), indicating possible population sub-structuring. Levels of genetic diversity in the clusters remained low based on observed heterozygosity (0.299–0.321) and PIC (0.214–0.265). Clusters K-C1, G-C1, and G-C2, shown with limited dispersal (Fig. 2), also possessed a lower rarefied Ar (1.743–1.889) and fewer number of private alleles (1–2 alleles) as compared to the other clusters that have a higher range of rarefied Ar (1.968–2.249) and private alleles (3–5 alleles).

3.5. Sex-biased dispersal

Results of sex-biased dispersal tests for the two sex classes are shown in Figs. 3**b** and 3**c** for the combined correlograms and separate correlograms of the two sexes, respectively. Average sample size in each of the distance classes was 12.52 ± 5.58. Paired-sample t-test was significant only at 15–30 km (excluding classes with ≤ 5 samples). The heterogeneity test for the correlogram wide Omega

Table 4
Genetic indices of the genetic clusters identified by K-means clustering and Bayesian spatial model in GENELAND software, among the *WILD* tapirs.

Software		K	Cluster		n	Ar	Na	Pa	Ho	He	Hs	Fis	PIC
adegenet		**3**	K-C1	**Mean**	9	1.889	1.889	1	0.321	0.255	0.267	-0.202	0.214
(K-means, non-spatial)				**SE**		0.309	0.309		0.109	0.085	0.089	(0.116)	
n = 39			K-C2	**Mean**	14	2.154	2.333	2	0.286	0.264	0.273	-0.046	0.233
				SE		0.264	0.289		0.098	0.074	0.076	(0.596)	
			K-C3	**Mean**	16	2.249	2.556	5	0.299	0.308	0.319	0.063	0.265
				SE		0.191	0.242		0.072	0.067	0.069	(0.418)	
GENELAND		**4**	G-C1	**Mean**	8	1.857	2.000	1	0.306	0.273	0.291	-0.055	0.229
(Bayesian, spatial)				**SE**		0.257	0.289		0.096	0.082	0.087	(0.722)	
n = 36			G-C2	**Mean**	9	1.743	2.000	1	0.235	0.214	0.227	-0.038	0.184
				SE		0.265	0.333		0.095	0.083	0.087	(0.734)	
			G-C3	**Mean**	14	1.968	2.222	3	0.310	0.324	0.336	0.081	0.272
				SE		0.132	0.147		0.065	0.053	0.054	(0.387)	
			G-C4	**Mean**	5	2.111	2.111	3	0.356	0.278	0.309	-0.174	0.242
				SE		0.351	0.351		0.104	0.081	0.088	(0.279)	
Cluster	**Pair**	**Fst**	**P(rand ≥ data)**			**Gst''**	**P(rand ≥ data)**			**Dest**	**P(rand ≥ data)**		
K = 3	**C1-C2**	0.136	0.001			0.286	0.001			0.098	0.001		
(Prefix K-)	**C1-C3**	0.150	0.001			0.324				0.123	0.001		
							0.001						
	C2-C3	0.110	0.001			0.242				0.086	0.001		
							0.001						
K = 4	**C1-C2**	0.157	0.001			0.308	0.001			0.103	0.001		
(Prefix G-)	**C1-C3**	0.136	0.001			0.289				0.113	0.001		
							0.001						
	C2-C3	0.083	0.001			0.154				0.049	0.001		
							0.001						
	C1-C4	0.164	0.002			0.322				0.123	0.002		
							0.002						
	C2-C4	0.147	0.003			0.275				0.091	0.002		
							0.003						
	C3-C4	0.072	0.029			0.108				0.037	0.028		
							0.027						

Ar, allelic richness per locus (number of alleles per locus) estimated and standardised to the smallest sample sizes by rarefaction; Na, number of alleles; Pa, private alleles; Ho, observed heterozygosity; He, expected heterozygosity; Hs, within population genetic diversity (Nei ,1987); Fis, inbreeding coefficient, 1−(mean Ho/mean Hs), significant difference from Fis = 0 tested by 1000 permutations (*p*-value shown in brackets); HW, *p*-value of Hardy-Weinberg test; PIC, polymorphism information content; Fst, measure of genetic differentiation (Nei, 1987); G''st, Hedrick's standardised Gst (Meirmans and Hedrick, 2011); Dest, Jost's estimate of differentiation (Jost, 2008);
P(rand ≥ data): probability of random data ≥ observed data based on 999 permutations
*p-values significant at 95% confidence level ($p < 0.05$)

computed based on the *p*-values of the paired sample t tests predicted homogeneity between the spatial correlograms of the two sexes, therefore suggesting no sex-biased dispersal in this dataset ($p = 0.186$). Fig. 3**d** shows the frequency distribution of AIc (corrected assignment index) values and Fig. 3**e** the bar plots of mAIc values in the male and female samples. A lower value of mAIc in the female group might be indicative of a dispersive sex, however, the difference between the sexes and therefore sex-biased dispersal was also not significantly supported ($p = 0.783$).

Table 5
Effective population size (Ne) and jackknife 95% confident interval (in brackets) estimated in the wild tapirs (n = 39) using either LDNe or molecular ancestry, including all nine microsatellite markers or removing marker not in linkage equilibrium (*Tter14*, or *Tter9*+*Tter17*).

Method	All (n = 9)	*Tter14* removed (n = 8)	*Tter9* and *Tter17* removed (n = 7)
LDNe	47.8	25.2	34.3
	(4.7-inf)	(1.8-inf)	(2.3-inf)
Molecular coancestry	11.1	11.1	10.7
	(0.0–55.8)	(0.1–55.7)	(0.0–53.9)
Harmonic mean Ne	18.0	15.4	16.3
Harmonic mean Ne/Nc	0.46	0.39	0.42

Critical value set to 0.02. Alleles with frequencies lower than this value are screened out. This is ignored by molecular coancestry method.
Harmonic mean Ne, harmonic mean of Ne estimated from LDNe and molecular ancestry methods. Ne/Nc, ratio of the harmonic mean of Ne to sample census size (n = 39).
Inf, infinity

3.6. Effective population size

Lastly, Ne estimates were computed using two methods molecular coancestry and LDNe, by including all markers or after discarding one of the markers in pair that were in LD (either removing *Tter9* and *Tter17*, or removing *Tter14*), as shown in Table 5. The harmonic means of Ne estimated across the two methods for three marker sets ranged from 15.4 to 18.0 tapirs, or 0.39–0.46 effective to census population ratio (Ne/Nc).

4. Discussion

4.1. Genetic diversity

Genetic diversity (Ho, He, and rarefied Ar) and the genetic structure of the *WILD* and *Unk* groups of Malayan tapirs assessed by the nine microsatellite markers were comparable (Table 3 and Supplementary Fig. S7). This suggests that the *Unk* (unknown origin) tapirs are largely representing the genetic makeup of the wild tapirs, although together with the *CAPTIVE* samples, they had to be removed from the population genetic structure analysis to avoid introducing errors (violations of HWE and non-random mating). These samples may be used for kinships analysis in the future with a more powerful set of nuclear genetic markers ($P_{ID\text{-}SIB} < 0.01$).

An overall low heterozygosity (except *CAPTIVE*) and low rarefied Ar in all categories indicates low genetic variability in the Malayan tapir populations in Peninsular Malaysia. The observed (0.299) and expected heterozygosity (0.333) of the wild tapirs in Peninsular Malaysia was comparable to the Baird's tapir but lower than the lowland tapir, putting the Malayan tapir in a lower range of genetic variability among the large mammals (Norton and Ashley, 2004; Pinho et al., 2014; de Thoisy et al., 2006). The allelic richness was also found to be very low (Ar = 2.316) when compared to the lowland tapirs (5.83–7.66), the white-tailed deer population started with four founders (5.36), the Bornean tree shrew in a fragmented landscape (2.67–7.11), and the endangered tigers (3.51) (Saranholi et al., 2022; Kekkonen et al., 2012; Brunke et al., 2020; Thapa et al., 2018). Significantly positive Fis (0.115) in the wild tapirs (population in HWE) may suggest inbreeding given 17.41% of individual pairs among the wild tapirs may share kinships ($r \geq 0.125$). However, in this case, it may also be caused by the Wahlund effect from population sub-structuring (see below).

On the other hand, captive breeding between genetically distant tapirs has likely produced an excess of heterozygotes (Fis = −0.084, but not significant)—also suggested by the genetic distribution on the PCoA plot where most captive tapirs aggregate in the middle (Supplementary Fig. S5). Much of the genetic variation in the captive tapirs are found in the wild tapirs, and their allelic richness is not significantly lower. Allelic richness is a more direct measure of genetic diversity, as having a higher observed heterozygosity level (*CAPTIVE* > *WILD*) does not warrant a more diverse pool of alleles including the presence of private alleles (four in *WILD* and none in *CAPTIVE*). Therefore, future captive programmes can benefit from taking both allelic richness (Ar) and heterozygosity (Ho and He) reported here as the baseline to evaluate the effectiveness of future captive breeding strategies.

4.2. Population genetic structure

Overall, results of DAPC (K = 3) and GENELAND (K = 4) suggested a possible genetic structure in the wild tapir population that would otherwise go undetected using the STRUCTURE analysis (see Supplementary Fig. S7). The cryptic pattern illustrated by overlaying the geographic distribution of the K-means clusters was further supported by the genetic analysis with a spatial model in GENELAND (Fig. 2). The better performance of DAPC may be attributed to its ability to detect fine-scale genetic structures, as it maximizes between-group variation while minimising within-group variation (Jombart et al., 2010), and of GENELAND that carries out Bayesian clustering with a spatial model that can be particularly useful in the case of sparse sampling (Ball et al., 2010). In, addition, the Bayesian method employed in STRUCTURE for differentiating genetically separate groups suffers from a limited number of samples and markers (Corander and Marttinen, 2006). Microsatellite markers used in this study was able to describe the genetic structure that was previously undetected using the maternally inherited mtDNA control region and employing Mantel test (Lim et al., 2021).

The cryptic pattern of genetic structure in the wild tapir population may be explained by the differing degree of connections between the forest complexes and sub-complexes. The heterogenous landscape of Peninsular Malaysia is characterised by four major forest complexes that can be further broken down to eight sub-complexes surrounded by agricultural and urbanised lands (Fig. 2**a**). The forest complexes are the backbone of the increasingly fragmented, nation-wide forests that are connected by primary and secondary linkages designated under the CFS Master Plan scheme to secure landscape connectivity (UNDP, G.E.F., G.O.M., 2015). However, not all linkages function at equal efficiency (Torre et al., 2019), and gene flow between populations would also depend on the dispersal behaviours of the animals as well as presence of natural and anthropogenic genetic barriers.

On the map, only K-C1 shows the least overlapping with other clusters, suggesting it as a distinct cluster with restricted dispersal, while K-C2 and K-C3 may suggest a longitudinal genetic cline (Fig. 2**a**). Spatial boundaries of the genetic clusters G-C1 to G-C4 (Fig. 2**b**) also indicate limited dispersion and gene flow between the populations in the southern part of Banjaran Titiwangsa-Banjaran Bintang-Banjaran Nakawan forest complex and the Benom Forest sub-complex with the rest of the sub-complexes (see Fig. 2**a**). Limited dispersal of the tapirs in this landscape was shown by positive correlation between the genetic and geographic distances in the distance classes ranging 15–60 km (Fig. 3**a**), which suggests that individuals within these distance classes are genetically closer than randomly expected. The extent of limited dispersion (x-intercept = 82 km) of the spatial autocorrelation analysis (Fig. 3**a**) approximate the extent of cluster spatial boundaries in Fig. 2**b** (darker region) and the description of tapir (male) home range of over 12 km^2 and up to 62 km^2 (Williams, 1979; Mohamad et al., 2019). The magnitude of pairwise F-/G-statistics (Table 4) between the *de novo* assigned

genetic clusters can be used as an indication for degree of genetic differentiation and for interpreting hypothetical geographical borders (Meirmans, 2015). It could also be inferred from the cluster memberships of the two methods with different K solution (Supplementary Fig. S10), in which members of K-C1 and G-C1 remains largely the same, G-C4 was derived from within K-C3, and more than half if not most of the members from K-C2 and K-C3 were retained in G-C2 and G-C3 while some others 'jump' in between the two clusters. Collectively, these observations and the cluster pattern produced by GENELAND (Fig. 2**b**) hints at a continuous genetic gradation across the Peninsular Malaysia forests (genetic clines) or admixture groups instead of distinct, isolated clusters (Blair et al., 2012; François and Durand, 2010) for G-C3, and G-C4 (and to a lesser degree G-C2). In contrast, the largest estimates of pairwise Fst are found between K-C1/G-C1 with other clusters and suggest moderate differentiation (Fst > 0.1). It reiterates the presence of true barriers in this region that may have been caused by low forest connectivity and anthropogenic dispersal barriers—consequences of habitat fragmentation and landscape conversion (urbanisation), that could not be explained solely by the effect of isolation-by-distance detected with spatial autocorrelation that breaks down at distances beyond 82 km (Fig. 3**a**). By contrast, the widespread K-C3/G-C3 on the east and west coasts of the peninsula might have experienced less of these effects due to better gene flow conferred by adequate forest linkages. In fact, tapirs have been observed at elevations as high as 1700 m (PERHILITAN, 2012), so altitude may not be an absolute barrier to gene flow for this species. This could explain the wide distribution of K-C3/G-C3/G-C4 along or across the east and west coasts, despite the existence of several mountain ranges in between (elevation map in Fig. 2).

The central and south-western clusters (i.e., K-C1, G-C1, and G-C2) are entrapped within or in proximity to the boundaries of the four forest sub-complexes i.e., the Main Range Forest Complex, Benom Forest Complex, Greater Taman Negara Forest Complex, and Chini-Bera Forest Complex. From the perspective of genetic diversity, it was observed that these clusters, what often with smaller sample sizes ($n = 8–9$) and restricted spatial boundaries (except G-C4), tend to have a lower allelic richness (1.743–1.889; but not significantly different from the other clusters) but demonstrate heterozygotes excess (Fis > 0 but insignificant). This may indicate possible inbreeding avoidance strategies (e.g., disassortative mating; Galaverni et al., 2016) within these geographically restricted populations, despite the contribution of kinship relationships (Supplementary Fig. S11) especially within the cluster K-C1(similar to G-C1). Further genetic monitoring will be required to validate this.

Period of sampling year within either K-cluster ($K = 3$, $n = 39$) ranged 7–12 years, while within G-cluster ($K = 4$, $n = 36$) was 3–11 years. In particular, the clusters seemingly with limited dispersal (K-C1, G-C1, and G-C2) were sampled across 6–9 years. Additionally, Moran's *I* found no spatial autocorrelation with sampling year (Supplementary Fig. S1), therefore false positive genetic structure due to spatial-temporal sampling biases without accounting for dynamic changes in allele frequency and its distribution is likely non-existent.

4.3. Sex-biased dispersal

Malayan tapirs are solitary, territorial, and mainly monogamous animals (Williams, 1978). Female-biased dispersal in mammals may be more likely found on monogamous branches of phylogeny, and instances of female-biased dispersal in solitary, monogamous mammals based on genetic evidence has been found, although the association between mating system and sex-biased dispersal is not completely clear (Munshi-South, 2008; Favre et al., 1997; Mabry et al., 2013). Sexual selection on body size in female tapirs and thus sex-biased dispersal was expected, however, this hypothesis was not supported by neither the spatial autocorrelation analysis nor assignment index method (Fig. 3). Nonetheless, the lower value of mAIc in the female group, although insignificant, may be suggestive of a dispersive sex (Fig. 3**e**). This was also unempirically shown by the wider sampling range of female compared to male tapirs (Supplementary Fig. S2), though there may have been a sampling bias and female-biased sex ratio (Lim et al., 2020).

4.4. Effective population size

For species with overlapping generations, Waples et al. (2014) has shown that while random samples of adults consistently underestimated Ne, drawing a number of cohorts of samples closely approximating the generation length (for Malayan tapir, 12 years; Traeholt et al., 2016) tended to produce an estimate with less downward bias in LDNe. The sample years of this study ranged 2002–2017 (Supplementary Table 1) and could approximate a generation. Ne estimated by the molecular coancestry method under two marker removal schemes yielded less disparity than what estimated by the LDNe method. Taken together, the harmonic means of Ne/Nc ratios (0.39–0.46) was higher than the median value (0.231) of 31 correctly-linked Ne/Nc estimates (Palstra and Fraser, 2012). The Ne/Nc ratios extrapolates to an Ne of 507–690 assuming a population of 1300–1500 tapirs in Peninsular Malaysia.

4.5. Conservation implications

Although only a pilot study, this study has provided an overview of genetic diversity, population genetic structure and effective population size of the wild tapir population in Peninsular Malaysia using biparentally inherited nuclear DNA markers. Across the sampling years of 2002–2017, overall wide distribution of tapir genetic clusters may be attributed to adequate forest connectivity implemented by the CFS project between the forest complexes (Torre et al., 2019). At least over the past decade, there has been substantial gene flow between the main forest complexes which has ensured the movement and gene flow between tapir populations. This is to be expected since tapirs are a species with wide home ranges and the individuals tend to disperse, e.g., the lowland tapir (Pinho et al., 2014). However, the observation of genetic barriers around the boundaries of the forest complexes in the central and south-western region indicates the possible ongoing evolution of allele frequencies. This may lead to undesired effects including genetic drift and inbreeding depression in small populations experiencing habitat isolation or other barriers to gene flow as a result of anthropogenic activities, e.g., urbanisation. The efforts of the Department of Wildlife and National Parks (PERHILITAN) to manage the

wild tapir populations and to increase the number of Wildlife Genetic Resources Bank (WGRB) samples via opportunistic collection of tapirs' biomaterials will also play an important role in contributing to the deepening of our knowledge on the tapir population genetic structure in the country.

4.6. Limitations and recommendations

Nine polymorphic microsatellite markers were successfully cross amplified to the Malayan tapir from the lowland tapir and the Baird's tapir. However, it is important to note that most of these markers carry imperfect, interrupted, and in some cases, compound microsatellites, except for *TtGT021* and *Tba23* (Table 2 **and** Supplementary Table S4). Further analysis using the microsatellite markers with perfect motifs from a genome assembly would guarantee a reduced risk of homoplasy (Chantra et al., 2021). The cumulative genotypic curve (Fig. 1) did not reach a plateau; therefore, the addition of more markers would also confer a better discrimination power between the unique MLGs and improvements over the current analysis. The combined P_{ID} and $P_{ID\text{-}SIB}$ of the nine markers (Supplementary Fig. S4 and Table S7) present limitations to identify individuals and siblings sufficiently, which was demonstrated as some individuals that could be (at best) identified from recorded information as unique individuals but shared the same MLGs with another individuals, even if assuming absence of null alleles. Furthermore, most of the wild tapir samples in this study were collected over more than 10 years under opportunistic events, i.e., rescue and translocation operations, road-kill incidents, etc. Therefore, the sampling locations as shown in Fig. 2 are mainly concentrated in areas with high rates of such operations or road-kill incidents, and most of these are nearby disturbed areas. Not all biological materials of encountered tapirs were sampled and stored in WGRB, and some of those which were stored contained improper records which led some samples to be assigned to the *Unk* group due to missing location information (Supplementary Table S1). Due to these factors, the presence of sampling bias in this study was inevitable. Future studies should focus on *in-situ* sample collection from the main tapir habitats within protected forests areas and forest reserves to enable a more comprehensive understanding of the historical and contemporary distribution of the tapirs' gene pool in Malaysia. Lastly, the exact mechanisms driving the microevolution of allele frequency in the central and south-western region compared to the outer region of the forest complexes require further investigation with a larger sample size and sampling area in order to test the relationship between genotype and landscape.

Ethical approval

All sampling procedures used for the whole blood samples were approved by the Institutional Animal Care and Use Committee, Universiti Putra Malaysia (Ethical Approval Ref.: UPM/IACUC/AUP-R033/2016). All methods were performed in accordance with the Universiti Putra Malaysia Code of Practice for the Care and Use of Animals for Scientific Purposes.

Funding

This project was supported by the Ministry of Education Malaysia's Fundamental Research Grant Scheme (FRGS), project code: FRGS/1/2014/SG03/UPM/02/8; project vote no: 5524552, and by the Malaysian Nature Society Tapir Project Grant awarded to Geetha Annavi.

CRediT authorship contribution statement

Q.L.L., G.A. and A.I. conceptualized the research work. **Q.L.L.** carried out the data collection and analysis, and drafted the manuscript. **W.L.N., C.S.Y.Y., A.I., J.J.R.-R., N.R., and G.A.** supervised the project. **M.I.-M.** provided laboratory supports for the bioinformatic analysis. All authors contributed to the revision of the manuscript. **G.A.** approved the final version of the manuscript to be published.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

We would like to thank Donny Yawah (PERHILITAN), Mat Naim Bin Haji Ramli (National Zoo of Malaysia), and Kavitha Jayaseelan (National Zoo of Malaysia) for their assistance in collecting blood samples at SDWCC and the National Zoo of Malaysia, respectively. We also thank Mdm Noor Azleen binti Mohd Kulaimi (PERHILITAN) for contributing to the collection of tapir samples from the National Wildlife Forensic Lab, PERHILITAN, as well as the staff from PERHILITAN, National Zoo of Malaysia, and SDWCC for providing sample information. We would like to thank the three anonymous reviewers who have provided valuable feedback on the

manuscript. All scientific activities were conducted under a permit from PERHILITAN (permit reference number: NRE 600-2/2/21 JILID 2(42)). Unpublished chromosome-length Hi-C genome assembly for the Malayan tapir was used with permission from the DNA Zoo Consortium (dnazoo.org). Q.L. Lim would like to thank the Ministry of Higher Education (MyBrainSc Scholarship Program) and Universiti Putra Malaysia (Graduate Research Fellowships) for the financial support received during her master studies at Universiti Putra Malaysia when most of the research in this study took place. She would also like to give thanks to Wildlife Research Center of Kyoto University for the bioinformatic resources received to complete the final analyses presented in this study.

Data availability

All datasets generated and/or analysed during the current study that are not included in this manuscript and the supplementary documents are available from the corresponding author on reasonable request.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.gecco.2022.e02321.

References

Allendorf, F.W., Luikart, G., Aitken, S.N., 2013. Conservation and the Genetics of Populations, second ed. Wiley-Blackwell, UK.

Ball, M.C., Finnegan, L., Manseau, M., Wilson, P., 2010. Integrating multiple analytical approaches to spatially delineate and characterize genetic population structure: an application to boreal caribou (*Rangifer tarandus caribou*) in central Canada. Conserv. Genet. 11, 2131–2143. https://doi.org/10.1007/s10592-010-0099-3.

Banks, S.C., Peakall, R., 2012. Genetic spatial autocorrelation can readily detect sex-biased dispersal: genetic signals of sex-biased dispersal. Mol. Ecol. 21, 2092–2105. https://doi.org/10.1111/j.1365-294X.2012.05485.x.

Barongi, R., 1986. Tapirs in captivity and their management at Miami Metrozoo. AAZPA Annual Conference Proceedings 96–108.

Bivand, R., 2022. R packages for analyzing spatial data: a comparative case study with areal data. Geogr. Anal. 54, 488–518. https://doi.org/10.1111/gean.12319.

Blair, C., Weigel, D.E., Balazik, M., Keeley, A.T.H., Walker, F.M., Landguth, E., et al., 2012. A simulation-based evaluation of methods for inferring linear barriers to gene flow. Mol. Ecol. Resour. 12, 822–833. https://doi.org/10.1111/j.1755-0998.2012.03151.x.

Brodie, J.F., Paxton, M., Nagulendran, K., Balamurugan, G., Clements, G.R., Reynolds, G., et al., 2016. Connecting science, policy, and implementation for landscape-scale habitat connectivity. Conserv. Biol. 30, 950–961. https://doi.org/10.1111/cobi.12667.

Brunke, J., Russo, I.-R.M., Orozco-terWengel, P., Zimmermann, E., Bruford, M.W., Goossens, B., et al., 2020. Dispersal and genetic structure in a tropical small mammal, the Bornean tree shrew (*Tupaia longipes*), in a fragmented landscape along the Kinabatangan River, Sabah, Malaysia. BMC Genet. 21, 43. https://doi.org/10.1186/s12863-020-00849-z.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al., 2009. BLAST+: architecture and applications. BMC Bioinform. 10, 421. https://doi.org/10.1186/1471-2105-10-421.

Campos-Arceiz, A., Traeholt, C., Jaffar, R., Santamaria, L., Corlett, R.T., 2012. Asian tapirs are no elephants when it comes to seed dispersal. Biotropica 44, 220–227. https://doi.org/10.1111/j.1744-7429.2011.00784.x.

Chantra, R., Dai, Y., Inoue-Murayama, M., Kittiwattanawong, K., Lim, Q.L., Rovie-Ryan, J.J., et al., 2021. Microsatellite records for volume 13, issue 4. Conserv. Genet. Resour. https://doi.org/10.1007/s12686-021-01243-2.

Corander, J., Marttinen, P., 2006. Bayesian identification of admixture events using multilocus molecular markers. Mol. Ecol. 15. https://doi.org/10.1111/j.1365-294X.2006.02994.x.

Crooks, K.R., Burdett, C.L., Theobald, D.M., King, S.R.B., di Marco, M., Rondinini, C., et al., 2017. Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. Proc. Natl. Acad. Sci. USA 114, 7635–7640. https://doi.org/10.1073/pnas.1705769114.

Culley, T.M., Stamper, T.I., Stokes, R.L., Brzyski, J.R., Hardiman, N.A., Klooster, M.R., et al., 2013. An efficient technique for primer development and application that integrates fluorescent labelling and multiplex PCR. Appl. Plant Sci. 1, 1300027. https://doi.org/10.3732/apps.1300027.

De, R., Sharma, R., Davidar, P., Arumugam, N., Sedhupathy, A., Puyravaud, J.-P., et al., 2021. Pan-India population genetics signifies the importance of habitat connectivity for wild Asian elephant conservation. Glob. Ecol. Conserv 32, e01888. https://doi.org/10.1016/j.gecco.2021.e01888.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., Ovenden, J.R., 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. Mol. Ecol. Resour. 14, 209–214. https://doi.org/10.1111/1755-0998.12157.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., et al., 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356, 92–95. https://doi.org/10.1126/science.aal3327.

Ernest, E.M., Haanes, H., Bitanyi, S., Fyumagwa, R.D., Msoffe, P.L., Bjørnstad, G., et al., 2012. Influence of habitat fragmentation on the genetic structure of large mammals: evidence for increased structuring of African buffalo (*Syncerus caffer*) within the Serengeti ecosystem. Conserv. Genet. 13, 381–391. https://doi.org/10.1007/s10592-011-0291-0.

Favre, L., Balloux, F., Goudet, J., Perrin, N., 1997. Female-biased dispersal in the monogamous mammal *Crocidura russula*: evidence from field data and microsatellite patterns. Proc. R. Soc. Lond. B Biol. Sci. 264. https://doi.org/10.1098/rspb.1997.0019.

Francis, R.M., 2017. pophelper: an R package and web app to analyse and visualize population structure. Mol. Ecol. Resour. 17, 27–32. https://doi.org/10.1111/1755-0998.12509.

François, O., Durand, E., 2010. Spatially explicit Bayesian clustering models in population genetics. Mol. Ecol. Resour. 10. https://doi.org/10.1111/j.1755-0998.2010.02868.x.

Galaverni, M., Caniglia, R., Milanesi, P., Lapalombella, S., Fabbri, E., Randi, E., 2016. Choosy wolves? Heterozygote advantage but no evidence of MHC-based disassortative mating. J. Hered. 107, 134–142. https://doi.org/10.1093/jhered/esv090.

Gonçalves da Silva, A., Lalonde, D.R., Russello, M.A., 2009. Isolation and characterization of microsatellite loci in a Neotropical ungulate, the lowland tapir (*Tapirus terrestris*). Conserv. Genet. Resour. 1, 39–41. https://doi.org/10.1007/s12686-009-9009-8.

Goossens, B., Chikhi, L., Ancrenaz, M., Lackman-Ancrenaz, I., Andau, P., Bruford, M.W., 2006. Genetic signature of anthropogenic population collapse in orang-utans. PLoS Biol. 4, e25 https://doi.org/10.1371/journal.pbio.0040025.

Goudet, J., 2005. hierfstat, a package for r to compute and test hierarchical F-statistics. Mol. Ecol. Notes 5, 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x.

Griciuvienė, L., Janeliūnas, Ž., Jurgelevičius, V., Paulauskas, A., 2021. The effects of habitat fragmentation on the genetic structure of wild boar (*Sus scrofa*) population in Lithuania. BMC Genom. Data 22, 53. https://doi.org/10.1186/s12863-021-01008-8.

Gros, A., Hovestadt, T., Poethke, H.J., 2008. Evolution of sex-biased dispersal: the role of sex-specific dispersal costs, demographic stochasticity, and inbreeding. Ecol. Model. 219. https://doi.org/10.1016/j.ecolmodel.2008.08.014.
Guillot, G., 2008. Inference of structure in subdivided populations at low levels of genetic differentiation—the correlated allele frequencies model revisited. Bioinformatics 24, 2222–2228. https://doi.org/10.1093/bioinformatics/btn419.
Guillot, G., Santos, F., 2009. A computer program to simulate multilocus genotype data with spatially autocorrelated allele frequencies. Mol. Ecol. Resour. 9, 1112–1120. https://doi.org/10.1111/j.1755-0998.2008.02496.x.
Guillot, G., Mortier, F., Estoup, A., 2005. Geneland: a computer package for landscape genetics. Mol. Ecol. Notes 5, 712–715. https://doi.org/10.1111/j.1471-8286.2005.01031.x.
Guillot, G., Estoup, A., Mortier, F., Cosson, J.F., 2005. A spatial statistical model for landscape genetics. Genetics 170, 1261–1280. https://doi.org/10.1534/genetics.104.033803.
Hardy, O.J., Vekemans, X., 2002. spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2, 618–620. https://doi.org/10.1046/j.1471-8286.2002.00305.x.
Jain, A., Chong, K.Y., Chua, M.A.H., Clements, G.R., 2014. Moving away from paper corridors in Southeast Asia. Conserv. Biol. 28, 889–891. https://doi.org/10.1111/cobi.12313.
Jenkins, D.G., Brescacin, C.R., Duxbury, C.V., Elliott, J.A., Evans, J.A., Grablow, K.R., et al., 2007. Does size matter for dispersal distance. Glob. Ecol. Biogeogr. 16. https://doi.org/10.1111/j.1466-8238.2007.00312.x.
Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.
Jombart, T., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071.
Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11, 94.
Jost, L., 2008. G. S.T. and its relatives do not measure differentiation. Mol. Ecol. 17, 4015–4026. https://doi.org/10.1111/j.1365-294X.2008.03887.x.
Kalinowski, S.T., Taper, M.L., Marshall, T.C., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16, 1099–1106. https://doi.org/10.1111/j.1365-294x.2007.03089.x.
Kamvar, Z.N., Tabima, J.F., Grünwald, N.J., 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2, e281. https://doi.org/10.7717/peerj.281.
Kamvar, Z.N., Brooks, J.C., Grünwald, N.J., 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genet. 6, 208. https://doi.org/10.3389/fgene.2015.00208.
Kassambara, A., Mundt, F., 2020. factoextra: extract and visualize the results of multivariate data analyses. In: R package version 1.0.7 [Internet]. 2020 [cited 10 Sep 2021]. Available: ⟨https://cran.r-project.org/web/packages/factoextra/index.html⟩.
Kawanishi, K., Sunquist, M., Othman, S., 2002. Malayan Tapirs (*Tapirus indicus*): far from extinction in a Malaysian rainforest. In: Tapir Conserv., 11, pp. 23–27.
Kekkonen, J., Wikström, M., Brommer, J.E., 2012. Heterozygosity in an isolated population of a large mammal founded by four individuals is predicted by an individual-based genetic model. PLoS One 7, e43482. https://doi.org/10.1371/journal.pone.0043482.
Kilanowski, A.L., Koprowski, J.L., 2016. Female-biased sexual size dimorphism: ontogeny, seasonality, and fecundity of the cliff chipmunk (*Tamias dorsalis*). J. Mammal. https://doi.org/10.1093/jmammal/gyw172.
Li, X.-Y., Kokko, H., 2019. Intersexual resource competition and the evolution of sex-biased dispersal. Front. Ecol. Evol. 7. https://doi.org/10.3389/fevo.2019.00111.
Lim, Q.L., Ismail, N.A., Arumugam, R., Ng, W.L., Yong, C.S.Y., Ismail, A., et al., 2019. A revisit to a low-cost method for the isolation of microsatellite markers: the case of the endangered Malayan tapir (*Tapirus indicus*). Malayan Nat. J. 71, 423–437.
Lim, Q.L., Tan, Y.L., Ng, W.L., Yong, C.S.Y., Ismail, A., Rovie-Ryan, J.J., et al., 2020. Molecular sexing and preliminary assessment of population sex ratio of the endangered Malayan tapir (*Tapirus indicus*) in Peninsular Malaysia. Sci. Rep. 10, 3973. https://doi.org/10.1038/s41598-020-60552-y.
Lim, Q.L., Yong, C.S.Y., Ng, W.L., Ismail, A., Rovie-Ryan, J.J., Rosli, N., et al., 2021. Genetic diversity and phylogenetic relationships of Malayan tapir (*Tapirus indicus*) populations in the Malay Peninsula based on mitochondrial DNA control region. Biodivers. Conserv. 30, 2433–2449. https://doi.org/10.1007/s10531-021-02202-x.
Linkie, M., Guillera-Arroita, G., Smith, J., Ario, A., Bertagnolio, G., Cheong, F., et al., 2013. Cryptic mammals caught on camera: assessing the utility of range wide camera trap data for conserving the endangered Asian tapir. Biol. Conserv. 162, 107–115. https://doi.org/10.1016/j.biocon.2013.03.028.
Lino, A., Fonseca, C., Rojas, D., Fischer, E., Ramos Pereira, M.J., 2019. A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. Mamm. Biol. 94, 69–76. https://doi.org/10.1016/j.mambio.2018.09.006.
Lynch, M., Ritland, K., 1999. Estimation of pairwise relatedness with molecular markers. Genetics 152, 1753–1766. https://doi.org/10.1093/genetics/152.4.1753.
Mabry, K.E., Shelley, E.L., Davis, K.E., Blumstein, D.T., van Vuren, D.H., 2013. Social mating system and sex-biased dispersal in mammals and birds: a Phylogenetic Analysis. PLoS One 8. https://doi.org/10.1371/journal.pone.0057980.
Magintan, D., Traeholt, C., 2012. Karuppanannan K v. Displacement of the Malayan Tapir (*Tapirus indicus*) in Peninsular Malaysia from 2006 to 2010. Tapir Conserv. 21, 13–17.
Magintan, D., Abdul Rahman, T., Jiliun, E.A., Adib, Y., Abd Aziz, A.A.H., Mohd Suri, M.S., et al., 2021. Malayan tapir roadkill in Peninsular Malaysia from 2006 to 2019. J. Wildl. Parks 36, 19–37.
Mahathir, M., Donny, Y., Noor, J.N.J., Magintan, D., Anuar, I., Shahril, E.J., et al., 2017. Movement patterns of a translocated Malayan tapir in Senaling Inas Forest Reserve, Negeri Sembilan. J. Wildl. Parks 32, 13–21.
Medici, E.P., Lynam, A., Boonratana, R., Kawanishi, K., Yatim, S.H., Traeholt, C., et al., 2003. Malay Tapir (Tapirus indicus): Conservation Workshop. IUCN/SSC Conservation Breeding Specialist Group; 1–103. Available: http://www.tapirs.org/Downloads/action-plan/malay-tapir-workshop-report.pdf.
Meirmans, P.G., 2015. Seven common mistakes in population genetics and how to avoid them. Mol. Ecol. 24, 3223–3231. https://doi.org/10.1111/mec.13243.
Meirmans, P.G., Hedrick, P.W., 2011. Assessing population structure: F_{ST} and related measures. Mol. Ecol. Resour. 11, 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x.
Miller, J.M., Cullingham, C.I., Peery, R.M., 2020. The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. Heredity 125, 269–280. https://doi.org/10.1038/s41437-020-0348-2.
Mohamad, M., Yawah, D., Magintan, D., Traeholt, C., Jemali, N.J.N., 2019. Habitat utilization of a translocated Malayan Tapir in Senaling Inas forest reserve, Negeri Sembilan. J. Sustain. Sci. Manag. 14, 65–70.
Mossman, C.A., Waser, P.M., 1999. Genetic detection of sex-biased dispersal. Mol. Ecol. 8, 1063–1067 (Available). ⟨http://www.ncbi.nlm.nih.gov/pubmed/10434424⟩.
Muangkram, Y., Wajjwalku, W., Salakij, C., Kaolim, N., Siriaroonrut, B., Kamolnorranath, S., et al., 2013. Use of mitochondrial cytochrome b sequences to determine the origin of captive Asian tapirs *Tapirus indicus*: implications for conservation. Endanger. Species Res. 21, 97–103. https://doi.org/10.3354/esr00509.
Muangkram, Y., Amano, A., Wajjwalku, W., Pinyopummintr, T., Thongtip, N., Kaolim, N., et al., 2017. Genetic diversity of the captive Asian tapir population in Thailand, based on mitochondrial control region sequence data and the comparison of its nucleotide structure with Brazilian tapir. Mitochondrial DNA DNA Mapp. Seq. Anal. 28, 597–601. https://doi.org/10.3109/24701394.2016.1149828.
Munshi-South, J., 2008. Female-biased dispersal and gene flow in a behaviorally monogamous mammal, the Large Treeshrew (*Tupaia tana*). PLoS One 3. https://doi.org/10.1371/journal.pone.0003228.
Nasa, J.P.L., 2013. NASA Shuttle Radar Topography Mission Global 1 arc second [Data set]. NASA EOSDIS Land Processes DAAC. Available: https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003.
Nei, M., 1977. F-statistics and analysis of gene diversity in subdivided populations. Ann. Hum. Genet. 41, 225–233. https://doi.org/10.1111/j.1469-1809.1977.tb01918.x.
Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press.

Nomura, T., 2008. Estimation of effective number of breeders from molecular coancestry of single cohort sample. Evol. Appl. 1, 462–474. https://doi.org/10.1111/j.1752-4571.2008.00015.x.
Norton, J.E., Ashley, M. v, 2004. Genetic variability and population structure among wild Baird's tapirs. Anim. Conserv. 7, 211–220. https://doi.org/10.1017/S1367943004001295.
Palstra, F.P., Fraser, D.J., 2012. Effective/census population size ratio estimation: a compendium and appraisal. Ecol. Evol. 2, 2357–2365. https://doi.org/10.1002/ece3.329.
Peakall, R., Smouse, P.E., 2012. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Bioinformatics 6, 288–295.
PERHILITAN, Tapir Information Sheet. Department of Wildlife and National Parks (PERHILITAN); 2012.
Pinho, G.M., Gonçalves da Silva, A., Hrbek, T., Venticinque, E.M., Farias, I.P., 2014. Kinship and social behavior of Lowland Tapirs (*Tapirus terrestris*) in a central Amazon Landscape. PLoS One 9, 1–10. https://doi.org/10.1371/journal.pone.0092507.
Pompanon, F., Bonin, A., Bellemain, E., Taberlet, P., 2005. Genotyping errors: causes, consequences and solutions. Nat. Rev. Genet. 6, 847–859.
Pritchard, J., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
R Core Team, R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available: ⟨https://www.r-project.org/⟩.
Raymond, M., Rousset, F., 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Hered. 86, 248–249.
Rousset, F., 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.
Rovie-Ryan, J.J., Traeholt, C., Zainuddin, E.M.-J., Mohd Shariff, Z.Z., Elagupillay S, K., et al., 2008. Sequence variation in Malayan Tapir (Tapirus indicus) inferred using partial sequences of the cytochrome b segment of the mitochondrial DNA. J. Wildl. Parks 25, 16–18.
RStudio Team, RStudio: Integrated Development for R. Boston, MA: RStudio, Inc.; 2021. Available: ⟨http://www.rstudio.com/⟩.
Samantha, L.D., Tee, S.L., Kamarudin, N., Lechner, A.M., Azhar, B., 2020. Assessing habitat requirements of Asian tapir in forestry landscapes: implications for conservation. Glob. Ecol. Conserv. 23, e01137 https://doi.org/10.1016/j.gecco.2020.e01137.
Sanches, A., de Figueiredo, M.G., Hatanaka, T., de Paula, F.F.P., Silveira, L., Jácomo, A.T.A., et al., 2009. Microsatellite loci isolated from the lowland tapir (*Tapirus terrestris*), one of the largest Neotropical mammal. Conserv. Genet. Resour. 1, 115.
Sanches, A., de Figueiredo, M.G., Hatanaka, T., de Paula, F.F.P., Silveira, L., Jácomo, A.T.A., et al., 2009. Microsatellite loci isolated from the lowland tapir (*Tapirus terrestris*), one of the largest neotropical mammal. Conserv. Genet. Resour. 1, 115–117. https://doi.org/10.1007/s12686-009-9028-5.
Saranholi, B.H., Sanches, A., Moreira-Ramírez, J.F., Carvalho, C., da, S., Galetti, M., Galetti Jr, P.M., 2022. Long-term persistence of the large mammal lowland tapir is at risk in the largest Atlantic forest corridor. Perspect. Ecol. Conserv. https://doi.org/10.1016/j.pecon.2022.02.002.
Schlaepfer, D.R., Braschler, B., Rusterholz, H.-P., Baur, B., 2018. Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere 9, e02488. https://doi.org/10.1002/ecs2.2488.
Thapa, K., Manandhar, S., Bista, M., Shakya, J., Sah, G., Dhakal, M., et al., 2018. Assessment of genetic diversity, population structure, and gene flow of tigers (*Panthera tigris tigris*) across Nepal's Terai Arc Landscape. PLoS One 13, e0193495. https://doi.org/10.1371/journal.pone.0193495.
de Thoisy, B., Richard-Hansen, C., Catzeflis, F., Lavergne, A., 2006. Population dynamics and DNA microsatellite survey in the lowland tapir. Tapir Conserv. 15, 14–16.
Torre, J.A., Lechner, A.M., Wong, E.P., Magintan, D., Saaban, S., Campos-Arceiz, A., 2019. Using elephant movements to assess landscape connectivity under Peninsular Malaysia's central forest spine land use policy. Conserv. Sci. Pract. 1. https://doi.org/10.1111/csp2.133.
Traeholt, C., Novarino, W., bin Saaban, S., Shwe, N.M., Lynam, A.J., Zainuddin, Z.Z., et al. Tapirus indicus. The IUCN Red List of Threatened Species 2016: e. T21472A45173636. 2016. Available: ⟨http://www.iucnredlist.org/details/21472/0⟩.
UNDP, G.E.F., G.O.M., Improving Connectivity in the Central Forest Spine (CFS) Landscape. IC-CFS Project Inception Report (PIMS 4594). 2015; 77. Available: http://www.my.undp.org/content/malaysia/en/home/operations/projects/environment_and_energy/improving-connectivity-in-the-central-forest-spine–cfs–landsca.html.
Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M., 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 35, W71–W74. https://doi.org/10.1093/nar/gkm306.
Vallone, P.M., Butler, J.M., 2004. AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques 37, 226–231.
van Oosterhout, C., Hutchinson, B., Wills, D., Shipley, P., 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4, 535–538.
Vartia, S., Collins, P.C., Cross, T.F., Fitzgerald, R.D., Gauthier, D.T., McGinnity, P., et al., 2014. Multiplexing with three-primer PCR for rapid and economical microsatellite validation. Hereditas 151, 43–54. https://doi.org/10.1111/hrd2.00044.
Waits, L.P., Luikart, G., Taberlet, P., 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol. Ecol. 10, 249–256. https://doi.org/10.1046/j.1365-294X.2001.01185.x.
Wang, J., 2017. The computer program STRUCTURE for assigning individuals to populations: easy to use but easier to misuse. Mol. Ecol. Resour. 17, 981–990. https://doi.org/10.1111/1755-0998.12650.
Wang, J., Santiago, E., Caballero, A., 2016. Prediction and estimation of effective population size. Heredity 117, 193–206. https://doi.org/10.1038/hdy.2016.43.
Waples, R.S., Do, C., 2008. ldne: a program for estimating effective population size from data on linkage disequilibrium. Mol. Ecol. Resour. 8, 753–756. https://doi.org/10.1111/j.1755-0998.2007.02061.x.
Waples, R.S., Antao, T., Luikart, G., 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics 197, 769–780. https://doi.org/10.1534/genetics.114.164822.
Whitmee, S., Orme, C.D.L., 2013. Predicting dispersal distance in mammals: a trait-based approach. J. Anim. Ecol. 82. https://doi.org/10.1111/j.1365-2656.2012.02030.x.
Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis, second ed. Use R!. Springer, Cham. Available: http://www.gbv.de/dms/ilmenau/toc/846024217.PDF.
Williams, K.D., 1978. Aspects of the ecology and behavior of the Malayan tapir (Tapirus indicus Dermarest) in the National Park of West Malaysia.
Williams, K.D., 1979. Radio-tracking tapirs in the rain forest of West Malaysia. Malay. Nat. J. 32, 253–258.
Williams, K.D., Petrides, G.A., 1980. Browse use, feeding behavior and management of Malayan Tapir. J. Wildl. Manag. 44, 489–494.