



# Phylogenetic divergence associated with climate oscillations and topology illustrates the dispersal history of Formosan sambar deer (*Rusa unicolor swinhoii*) in Taiwan

Kuan-Yi Li<sup>1</sup> · Chen Hsiao<sup>1</sup> · Shih-Ching Yen<sup>2</sup> · Chien-Yi Hung<sup>1</sup> · You-Zhu Lin<sup>1</sup> · Sheng-Wun Jheng<sup>1</sup> · Pei-Ju Yu<sup>1</sup> · Mei-Hsiu Hwang<sup>3</sup> · Guo-Jing Weng<sup>3</sup> · Ku-Lin Chen<sup>4</sup> · Shu-Fen Lin<sup>5</sup> · Shih-Wei Chang<sup>6</sup> · Ying Wang<sup>7</sup> · Yu-Ten Ju<sup>1</sup>

Received: 30 April 2022 / Accepted: 3 March 2023 / Published online: 20 March 2023

© The Author(s) 2023

## Abstract

The island of Taiwan represents an ideal context for studying the effects of climatic oscillations and topographic variation on large herbivores due to its varied tropical to sub-tropical climate zones at different elevational ranges. We explored the phylogenetics of Formosan sambar deer (*Rusa unicolor swinhoii*) using the control region of the mitochondrial genome. We detected 18 haplotypes among 454 sequences across the island and grouped them into six regions based on SAMOVA, with 68.78% variance among regions. A Bayesian phylogenetic dendrogram revealed two spatially segregated genetic clades. Neutrality tests and Bayesian skyline plots uncovered different demographic expansion histories for the two clades. We further tested divergence times and chronology to propose potential phylogenetic scenarios, which were examined using approximate Bayesian computation. Finally, we present a credible hypothesis for a glacial refugium in the northern part of the Central Mountain Range. Subsequent secondary contact between the two clades during interglacial periods has led to the extant genetic structure of Formosan sambar deer.

**Keywords** Phylogeography · Mitochondrial control region DNA · Glacial refugium · Secondary contact

---

Communicated by: Astrid V. Stronen

---

Kuan-Yi Li and Chen Hsiao equally contributed to this work.

---

✉ Yu-Ten Ju  
ytju@ntu.edu.tw

<sup>1</sup> Department of Animal Science and Technology, National Taiwan University, Taipei 10673, Taiwan

<sup>2</sup> Center for General Education, National Tsing Hua University, Hsinchu 30013, Taiwan

<sup>3</sup> Institute of Wildlife Conservation, College of Veterinary Medicine, National Pingtung University of Science and Technology, Pingtung 91201, Taiwan

<sup>4</sup> Taroko National Park, Hualien 97253, Taiwan

<sup>5</sup> Institute of History and Philology, Academia Sinica, Taipei 11529, Taiwan

<sup>6</sup> Division of Zoology, Endemic Species Research Institute, Nantou 55244, Taiwan

<sup>7</sup> Department of Life Science, National Taiwan Normal University, Taipei 40227, Taiwan

## Introduction

Population dispersal and vicariance has become the subject of an increasing number of phylogeographic studies (Chan et al. 2011; Chavez et al. 2014; Valente et al. 2014). Integrating the molecular phylogenies of extant taxa with topographic factors and known climatic oscillations contributes to our understanding of the evolutionary processes that shape genetic patterns, historic colonization routes, and species evolution (Avisé 2009; Chavez et al. 2014). Environmental change and climatic oscillations drive species range shifts along geographic gradients, in terms of both latitude and altitude, and have been linked to evolutionary events (Epps et al. 2006; Gorelick 2008; Angert et al. 2011; Liang et al. 2017).

The succession of warm and cold periods during the Pleistocene (starting 2.5 million years ago) has greatly contributed to intra-specific diversification, especially arising from populations persisting in allopatric refugia (Avisé 2007; Husemann et al. 2014). Refugia represent areas in which a population can survive during unfavorable environmental conditions (in this case glacial periods). During the cold episodes of the Pleistocene, populations became isolated in refugia as their habitats

fragmented (Bennett and Provan 2008; Sommer and Zachos 2009; Stewart et al. 2010). This scenario is particularly relevant to mountainous terrain, where habitat fragmentation gives rise to multiple refugia in valleys, greatly facilitating species diversification (Potts 2013).

Taiwan features a tropical to sub-tropical climate. The island has six steep mountain ranges, encompassing 268 peaks > 3000 m, deep gorges, and wide valleys, together producing diverse altitudinal climate zones and vegetation types at a very fine topographical scale. These features render Taiwan an ideal locale to study refugia and biodiversification. Significant climatic oscillations have occurred in Taiwan throughout its geological history. Alpine cirque glaciers have been identified as major features of the high mountains in northern Taiwan (Böse 2004; Böse and Hebenstreit 2011). Moreover, its low- and mid-altitude areas display evidence of a shift of highland forests to lower elevations (Böse 2004). Vertical distributions of vegetation zones, ranging from cryoflora (alpine tundra and subalpine coniferous woodland) to sub-mountainous evergreen broadleaf forest (tropical rain forest), shape faunal distribution patterns (Böse 2000; Li et al. 2013). Dramatic climate change during the transition of interglacial periods to glacial periods can result in niche fragmentation and produce barriers potentially restricting species dispersal. The effects of climatic oscillations on the genetic structure and demographic history of taxa such as amphibians, freshwater fishes, and small mammals during Pleistocene glaciations have been described for Taiwan (Yuan et al. 2006; Lin et al. 2008, 2012; Oshida et al. 2011). Large herbivores like ruminants are particularly suitable for studying how mammals respond to changes in terms of adaptive strategies because they are highly sensitive to vegetation modification during climate changes (DeMiguel et al. 2010; Strani et al. 2018). However, the ramifications of climate fluctuations for large herbivores in Taiwan have remained unexplored.

Formosan sambar deer (*Rusa unicolor swinhoii*) is a large herbivore endemic to Taiwan that is found in mountainous areas (1500 to 3500 m). It occurs in both forested and grassland habitats, and areas of rugged terrain (less than 60° slope). It exhibits altitudinal migratory behavior, preferring lower elevations during colder weather (Yen et al. 2013, 2019). Based on its island-wide distribution and migratory behavior, the dispersal history of the Formosan sambar deer represents an ideal model for examining the impact of Pleistocene climatic oscillations on a large herbivore.

Here, we use mitochondrial control region sequences to examine the genetic patterns, lineage divergence, and demographic history of the Formosan sambar deer. We link our findings to climate change during the Pleistocene and hypothesize a possible scenario for how topology and climatic oscillations have influenced sambar dispersal in Taiwan.

## Materials and methods

### Sample collection

We collected 877 fecal samples from 25 sampling locations in mountainous areas throughout Taiwan. Samples from these sampling sites were divided into nine subpopulations according to geographic location and connectivity, i.e., sampling locations located in adjacent mountain ranges without obvious geographic barriers to dispersal were classified as a subpopulation. Fecal samples were preserved in 95% EtOH and stored at –20 °C. All sampling protocols were approved by the administration of Taroko National Park, the Forestry Bureau of Taiwan, and the Institutional Animal Care and Use Committee of National Taiwan Normal University (approval no. 101004), and were performed in accordance with relevant guidelines and regulations.

### Fecal DNA extraction, PCR amplification, and mitochondrial control region sequencing

Genomic DNA was extracted from fecal samples using the FavorPrep Stool DNA Isolation Mini Kit (Favorgen Biotech Corp., Pingtung, Taiwan). The full-length mitochondrial control region (D-loop) was amplified and sequenced using polymerase chain reaction (PCR) primers designed according to the complete mitochondrial DNA sequence of Formosan sambar deer (GenBank accession number NC\_008414): forward primer D-loop F1, 5'-GAAAAG GAGAGCAACCAACC-3', and reverse primer D-loop R1, 5'-GGCTGGGACCAAACCTGTATG-3'. The highly variable tandem repeats (5'-TATAATAGTACATAAAATTAA TGTATTAGGACATATCATG-3') and cytosine-thymine repeats (5'-ATATTTCCCCCCCCCTCCATTAATTTTT CCCCC-3') within the control region may affect sequencing accuracy. To avoid sequencing errors arising from these regions, we used a nested primer pair to acquire internal sequences: D-loop F2, 5'-TGCCCCATGCTTATAAGC-3'; D-loop R2, 5'-CATTGATGAGTCCAAGCATA-3'. PCR was performed using the Advantage 2 PCR enzyme system (TAKARA Bio, Inc., Shiga, Japan). Thermal cycling conditions were 94 °C for 5 min, 39 cycles at 94° C for 45 s, 60 °C for 30 s, and 68 °C for 80 s, with a final extension at 68 °C for 10 min. The PCR products were purified with the GenepHlow Gel/PCR Kit (Geneaid Biotech Ltd., New Taipei City, Taiwan), and sequenced with an Applied Biosystems 3730 DNA Analyzer (Applied Biosystems, Inc., Foster City, CA). Partial control region sequences were obtained by merging sequence fragments in the EditSeq software (DNASTAR, Inc., Madison, WI).

## Haplotype network analysis, Bayesian phylogenetic tree, and divergence time estimation

The highly variable tandem repeats and cytosine-thymine repeats within the control region were excluded from phylogenetic analysis, leaving 889 base pairs (bp) for phylogenetic analysis. Out of 877 fecal samples, a total of 454 bp sequences were successfully acquired, which have been deposited into the National Center for Biotechnology Information (NCBI) GenBank with accession numbers KU531437–KU531440, KU531442, KU531443, and KU531445–KU531456. Multiple sequence alignment was executed in the MegAlign software (DNASTAR Inc.) to identify haplotypes. A haplotype network with 95% probability of parsimony connections was constructed in TCS v1.21 software (Clement et al. 2000) to clarify the relationship among haplotypes. In addition, control region sequences from *Rusa unicolor hainana* (GQ304777), *Muntiacus reevesi* (NC\_004069), and *Muntiacus crinifrons* (NC\_004577) were obtained from NCBI GenBank. A phylogenetic tree of haplotypes using Bayesian inference was constructed in the BEAST software package v2.3.0 (Bouckaert et al. 2014) using the HKY + I model, which was selected as the best-fitting model in jModelTest v2.1.5 (Bouckaert et al. 2014), and four times Markov chain Monte Carlo (MCMC) sampling was executed for 100 million generations and sampled every 1000 generations. A maximum clade credibility tree was generated after the first 10% of tree samples had been discarded as burn-in using TreeAnnotator v2.3.0 (Bouckaert et al. 2014). Divergence time was estimated by applying a Bayesian strict clock model with a clock rate assuming 0.04 substitutions/million years (Randi et al. 1998) and a calibrated Yule tree prior (Drummond and Suchard 2010; Heled and Drummond 2011). To calibrate divergence times on our phylogenetic tree against the fossil record, we specified a normal prior distribution of  $8 \pm 1$  million years ago (MYA) for the node of the *Cervinae* and *Muntiacinae* subfamily (Dong et al. 2004; Gilbert et al. 2006). The tree was visualized and edited in FigTree v1.4.2 (Rambaut 2014). Values for effective sample size ( $> 200$ ) for all parameter estimates were verified in the Tracer v1.6 software (Rambaut et al. 2013).

## Analyses of population structure

A spatial analysis of molecular variance (SAMOVA) was conducted using SAMOVA 1.0 (Dupanloup et al. 2002) to define groups of subpopulations that are genetically homogeneous and maximally differentiated. SAMOVA was conducted with the number of groups ( $K$ ) ranging from 2 to 8. To verify output consistency, we ran the analysis five times for each  $K$  value with 1000 independent iterations, starting from 100 random initial conditions. We determined optimal

$K$  as that having the greatest percentage of variation and being most significant. To investigate genetic diversity, the DnaSP v 6.12 software (Librado and Rozas 2009) was used to calculate haplotype diversity ( $h$ ) and nucleotide diversity ( $\pi$ ) in Formosan sambar deer. In addition, pairwise genetic differentiation ( $F_{ST}$ ) and genetic distance ( $D_{xy}$ ) among our a priori nine subpopulations were assessed in the Arlequin v3.5.1.2 (Excoffier and Lischer 2010) and MEGAXI (Tamura et al. 2021) software, respectively.

## Phylogeographic and historical demographic analyses

To investigate the history of divergence among SAMOVA-identified regions (NSP, SSP, NTR, STR, NDD and CSR; see “Results” and Table 1 for details), we tested 10 scenarios by means of approximate Bayesian computation (ABC) in the DIYABC v2.1 software (Cornuet et al. 2014). We generated CSR-origin (scenarios 1–5) and SSP-origin (scenarios 6–10) scenarios to test the direction of migration and the origin of Formosan sambar. Since the NDD, NTR, and STR regions possessed haplotypes from both genetic clades, our tests also explored different divergence orders for these subpopulations and the existence of secondary contact events. Scenarios 1 and 2 assume that the NDD subpopulation was the first to split from the CSR subpopulation. A secondary contact then occurred between the CSR and NDD subpopulations to give rise to the NTR + STR subpopulation. Scenarios 3 and 4 assume that the earliest divergence occurred between the CSR and STR subpopulations, followed by secondary contact between them. The timing of the split of the SSP and NSP or STR and NTR subpopulations differs within these two pairs of scenarios (scenario 1 versus 2 or scenario 3 versus 4, respectively). Scenarios 6 and 7 predict different origins of the CSR, i.e., from the SSP (scenario 6) or NDD (scenario 7) subpopulations. Both scenarios 8 and 9 allowed us to test if NTR initially split from NSP, but with the order of SSP diverging from NSP and STR diverging from NTR differing. Scenarios 5 and 10 tested the hypothesis that no secondary contact had occurred. Neutrality tests, including Tajima’s  $D$  and Fu’s  $F_s$  tests, were performed in the Arlequin v3.5.1.2 software (Excoffier and Lischer 2010). Mismatch distributions were calculated and depicted using DnaSP v 6.12. A Bayesian skyline plot (BSP) was generated in BEAST v2.3.0 (Bouckaert et al. 2014) to date changes in population size over time. For that analysis, we applied estimated clock rates from our Bayesian phylogenetic tree using fossil calibrations. Under a coalescent Bayesian skyline model, four MCMC iterations were executed for 300 million generations and then sampled every 1000 generations, followed by removal of 10% of samples as burn-in. BSP reconstructions were visualized in the Tracer v1.6 software (Rambaut et al. 2013).

**Table 1** SAMOVA on nine Formosan sambar deer subpopulations

Groups (K)	Grouping pattern	Percentage variance among groups (%)	$\Phi_{CT}$
2	Region 1 (NSP and SSP) and region 2 (NTR, STR, NDD, SDD, NYS, SYS, and SGL)	66.38	0.6638*
3	Region 1 (NSP), region 2 (SSP), and region 3 (NTR, STR, NDD, SDD, NYS, SYS, and SGL)	65.31	0.6531*
4	Region 1 (NSP), region 2 (SSP), region 3 (NDD), and region 4 (NTR, STR, SDD, NYS, SYS, and SGL)	64.25	0.6425*
5	Region 1 (NSP), region 2 (SSP), region 3 (NTR and STR), region 4 (NDD), and region 5 (SDD, NYS, SYS, and SGL)	65.08	0.6508***
6	Region 1 (NSP), region 2 (SSP), region 3 (NTR), region 4 (STR), region 5 (NDD), and region 6 (SDD, NYS, SYS, and SGL)	68.78	0.6878**
7	Region 1 (NSP), region 2 (SSP), region 3 (NTR), region 4 (STR), region 5 (NDD), region 6 (NYS), and region 7 (SDD, SYS, and SGL)	68	0.6800*
8	Region 1 (NSP), region 2 (SSP), region 3 (NTR), region 4 (STR), region 5 (NDD), region 6 (NYS), region 7 (SYS), and region 8 (SDD and SGL)	67.3	0.6730*

$\Phi_{CT}$ , \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.0001$

## Results

### Population structure and genetic differentiation

We generated 454 control region sequences from 25 sampling sites and divided them into nine subpopulations according to the geographic distance between sites and the mountain ranges they were located on (Fig. 1; Supplementary Fig. S1). We identified 18 haplotypes arising from 19 variable sites (Supplementary Table S2). We conducted a SAMOVA to examine best-fit grouping patterns by maximizing the proportion of genetic variance among regions, which revealed an optimal grouping of six regions, with the highest variance (68.78%) attributed to variation among regions ( $P < 0.01$ ) (Table 1). Of the nine a priori subpopulations, five (NSP, SSP, NTR, STR, NDD) were assigned by SAMOVA as being distinct regions, with one further region comprising the remaining SDD, NYS, SYS, and SGL subpopulations (Table 1). Hereafter, these six regions are denoted Southern Sheipa National Park (SSP), Northern Sheipa National Park (NSP), Southern Taroko National Park (STR), Northern Taroko National Park (NTR), Northern Danda Major Wildlife Habitat (NDD), and Central to Southern Range (CSR, formed by the grouping of the four remaining subpopulations).

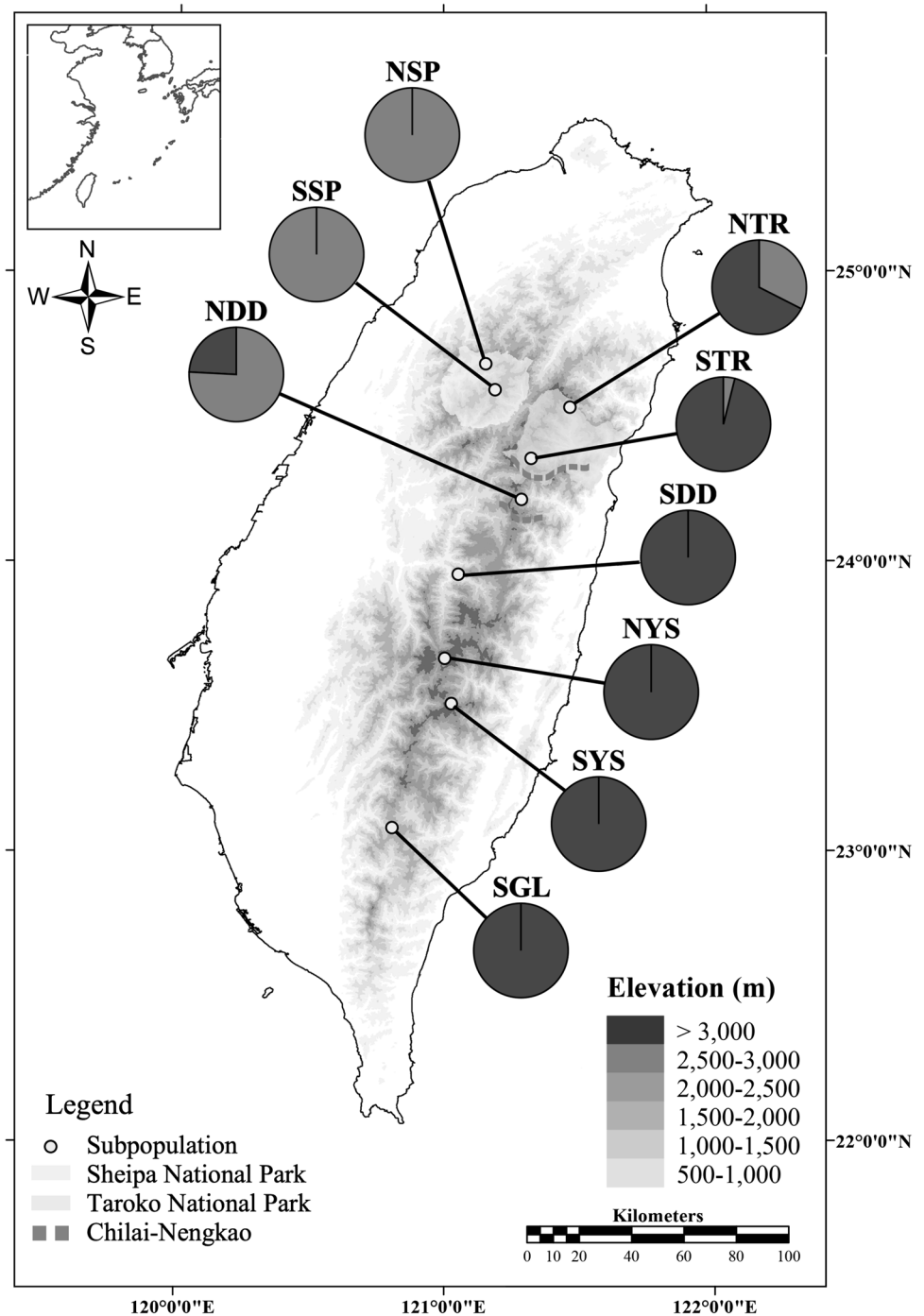
Haplotype diversity ( $h$ ) and nucleotide diversity ( $\pi$ ) for the entire Formosan sambar deer population were determined to be 0.814 and 0.00371, respectively. Only one haplotype was found in each of the NSP and SSP regions. Apart from those two regions, the value of  $h$  for the remaining four regions ranged from 0.255 (STR) to 0.693 (NTR), and values of  $\pi$  ranged from 0.00056 (CSR) to 0.00399 (NTR). All regions and subpopulations presented low values for  $\pi$  ( $< 0.005$ ). These results indicate that deer from NTR possess greater genetic diversity than those from other regions (Table 2).

We measured genetic distances ( $D_{xy}$ ) and fixation indices ( $F_{ST}$ ) to elucidate the genetic differentiation among subpopulations (Table 3). Regions NSP and SSP presented the closest genetic distance (0.0011). The largest genetic distance was identified between the SSP and STR subpopulations (0.0081), rather than between the NSP and SGL subpopulations (0.0069) that are the furthest apart. Similarly, we observed lower genetic distance but greater geographic distance between the SDD and SGL subpopulations (0.0002) than between adjacent STR and NDD subpopulations (0.0043). The lowest  $F_{ST}$  values were detected between the SDD and SGL subpopulations (0.0038), with the highest respective values being between NSP and SSP (1.0000), representing two adjacent subpopulations (Table 3). These results indicate that geographical distance may not be the only factor contributing to genetic barriers. Consequently, we explored historical demography to explain genetic differentiation across regions.

### Phylogenetic relationships and divergence

To understand the genetic characteristics of the Formosan sambar deer population, we performed a network analysis to depict the genealogy of the 18 identified haplotypes (Fig. 2). We uncovered two primary genetic clades comprising RUS01-04 and RUS05-18, with four diagnostic nucleotide substitutions between them at sites 151 (C/T), 180 (G/A), 352 (A/G), and 371 (C/T) (Supplementary Table S2). To further clarify the phylogenetic relationships among our control region sequences, we constructed a phylogenetic tree based on Bayesian inference. Sequences of *Rusa unicolor hainana*, *Muntiacus reevesi*, and *Muntiacus crinifrons* obtained from NCBI were used as outgroups and for fossil calibration of molecular clock dating. Our Bayesian phylogenetic tree clearly separated the 18

**Fig. 1** Geographic distribution of two clades identified from phylogenetic analysis of nine Formosan sambar deer subpopulations. The red and blue sections of the nine subpopulation pie charts represent the frequency of occurrence of Taroko-Sheipa (TS) and Central Mountain Range (CMR) clade haplotypes, respectively. The yellow dots denote mean geographical centers of the sampling sites. Details on sampling sites and subpopulations are presented in Supplementary Table S1



Formosan sambar deer haplotypes into two clades, with a significant Bayesian posterior probability ( $> 0.98$ ; Fig. 3). The two clades reflect spatial divergence between north and south Taiwan. Hereafter, we name the clade comprising haplotypes RUS01-04 as Taroko-Sheipa (TS), which encompasses deer distributed in regions SSP, NSP, STR, NTR, and NDD in northern Taiwan. Within the TS clade, haplotype RUS01 is restricted to the SSP region; haplotype RUS02 occurs in NSP, NTR, and STR; haplotype

RUS03 in NTR; and haplotype RUS04 in STR and NDD (Supplementary Table S3). The second clade (hereafter, the Central Mountain Range (CMR) clade) comprises haplotypes RUS05–18. Haplotypes RUS05 and RUS06 of the CMR clade are only found in regions STR, NTR, and NDD, apart from one individual that was detected in the SDD subpopulation of the CSR region. All other haplotypes (RUS07–18) were solely distributed in the CSR region. Notably, 147 individual deer possessed haplotype

**Table 2** Sample size ( $n$ ), number of haplotypes ( $nh$ ), haplotype diversity ( $h$ ), nucleotide diversity ( $\pi$ ), and neutrality test results for each region, subpopulation, and clade of Formosan sambar deer

Region/Clade	Subpopulations	$n$	$nh$	$h$	$\pi$	Tajima's $D$	Fu's $F_s$
Northern Sheipa National Park (NSP)	Northern Sheipa National Park (NSP)	49	1	0.000±0.000	0.00000±0.00000	–	–
Southern Sheipa National Park (SSP)	Southern Sheipa National Park (SSP)	26	1	0.000±0.000	0.00000±0.00000	–	–
Northern Taroko National Park (NTR)	Northern Taroko National Park (NTR)	107	4	0.693±0.014	0.00399±0.00024	3.1869	8.89707
Southern Taroko National Park (STR)	Southern Taroko National Park (STR)	50	4	0.255±0.077	0.00074±0.00036	–1.72724*	–0.19621
Northern Danda Major Wildlife Habitat (NDD)	Northern Danda Major Wildlife Habitat (NDD)	29	2	0.379±0.084	0.00213±0.00047	1.34754	5.81229
Central to Southern Range (CSR)		193	13	0.415±0.045	0.00056±0.00007	–1.84610**	–12.08819***
	Southern Danda Major Wildlife Habitat (SDD)	31	2	0.065±0.059	0.00007±0.00007	–1.14473	–1.23856**
	Northern Yushan National Park (NYS)	46	9	0.829±0.031	0.00138±0.00012	–0.90667	–3.34287*
	Southern Yushan National Park (SYS)	28	5	0.434±0.111	0.00054±0.00016	–1.39757	–2.70035**
	Shuangguei Lake Major Wildlife Habitat (SGL)	88	4	0.131±0.049	0.00025±0.00010	–1.45842*	–2.38072*
Taroko-Sheipa (TS) clade		134	4	0.571±0.037	0.00104±0.00009	1.17445	1.45254
Central Mountain Range (CMR) clade		320	14	0.701±0.018	0.00115±0.00005	–1.18201	–6.11141*
Total		454	18	0.814±0.010	0.00371±0.00012	0.40068	–0.36495

\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$

**Table 3** Pairwise genetic differentiation index ( $F_{ST}$ , below the diagonal) and genetic distance ( $D_{xy}$ , above the diagonal) based on mitochondrial control-region sequences for nine Formosan sambar deer subpopulations

Subpopulation	NSP	SSP	NTR	STR	NDD	SDD	NYS	SYS	SGL
NSP	–	0.0011	0.0058	0.0078	0.0036	0.0068	0.0076	0.0071	0.0069
SSP	1.0000	–	0.0069	0.0089	0.0048	0.0080	0.0087	0.0082	0.0080
NTR	0.5761	0.5992	–	0.0031	0.0044	0.0033	0.0041	0.0036	0.0035
STR	0.9519	0.9456	0.1878	–	0.0043	0.0014	0.0022	0.0017	0.0015
NDD	0.7645	0.7645	0.2693	0.6985	–	0.0037	0.0045	0.0040	0.0038
SDD	0.9959	0.9950	0.2829	0.6669	0.7088	–	0.0008	0.0003	0.0002
NYS	0.9116	0.9004	0.3049	0.5224	0.6200	0.0913	–	0.0010	0.0009
SYS	0.9723	0.9660	0.2824	0.6131	0.6606	0.0415	0.0570	–	0.0004
SGL	0.9768	0.9759	0.3654	0.7116	0.7884	0.0038	0.1180	0.0537	–

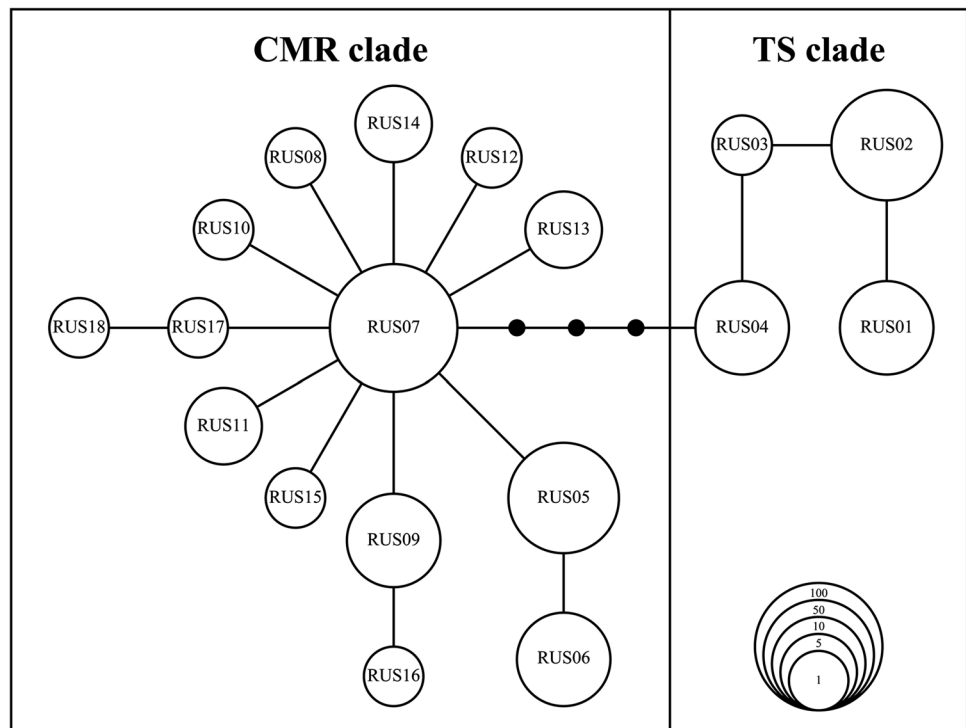
RUS07, representing the majority of our samples, and it is more widely distributed than any of the other haplotypes (Supplementary Table S3). Values of  $h$  and  $\pi$  for the TS clade were 0.571 and 0.00104, respectively, whereas for the CMR clade, they were 0.701 and 0.00115 (Table 2), indicating greater genetic diversity in the CMR clade relative to the TS clade.

The divergence times to the most recent common ancestor of each clade were estimated using a calibrated molecular clock analysis on our Bayesian phylogenetic tree (Fig. 3), which revealed coalescence of a common ancestor for *Rusa*

*unicolor hainana* and *Rusa unicolor swinhoii* (Formosan sambar deer) approximately 0.97 MYA (95% highest posterior density (HPD) = 0.55–1.41 MYA), i.e., during the late Pleistocene. Coalescence of our two Formosan sambar deer clades occurred 0.41 MYA (95% HPD = 0.21–0.63 MYA), also in the late Pleistocene. The most recent common ancestors for the TS and CMR clades arose 0.16 MYA (95% HPD = 0.04–0.30 MYA) and 0.24 MYA (95% HPD = 0.12–0.37 MYA), respectively.

We also generated a Bayesian skyline plot (BSP) to reconstruct the relationship between the timing of past events and

**Fig. 2** Haplotype network displaying 95% parsimonious connections of Formosan sambar deer haplotypes. The open circles show the 18 detected haplotypes, and the solid circles show hypothetical haplotypes. The lines connecting haplotypes indicate single-step mutations between them. The circle size is proportional to the sample size presented in the lower right corner. CMR, Central Mountain Range clade; TS, Taroko-Sheipa clade



the magnitude of change in the effective maternal population size of Formosan sambar deer since the appearance of its most recent common ancestor (Fig. 4). This analysis revealed a slow and steady increase in overall population size (Fig. 4a). Notably, the effective maternal population size of the CMR clade was relatively constant up to 10,000 years ago (YA) and then rapidly expanded thereafter (Fig. 4b). In contrast, the effective maternal population size of the TS clade has increased only gradually since 2000 YA, expanding to a much lesser extent than observed for the CMR clade (Fig. 4c).

### Historic demographic analyses

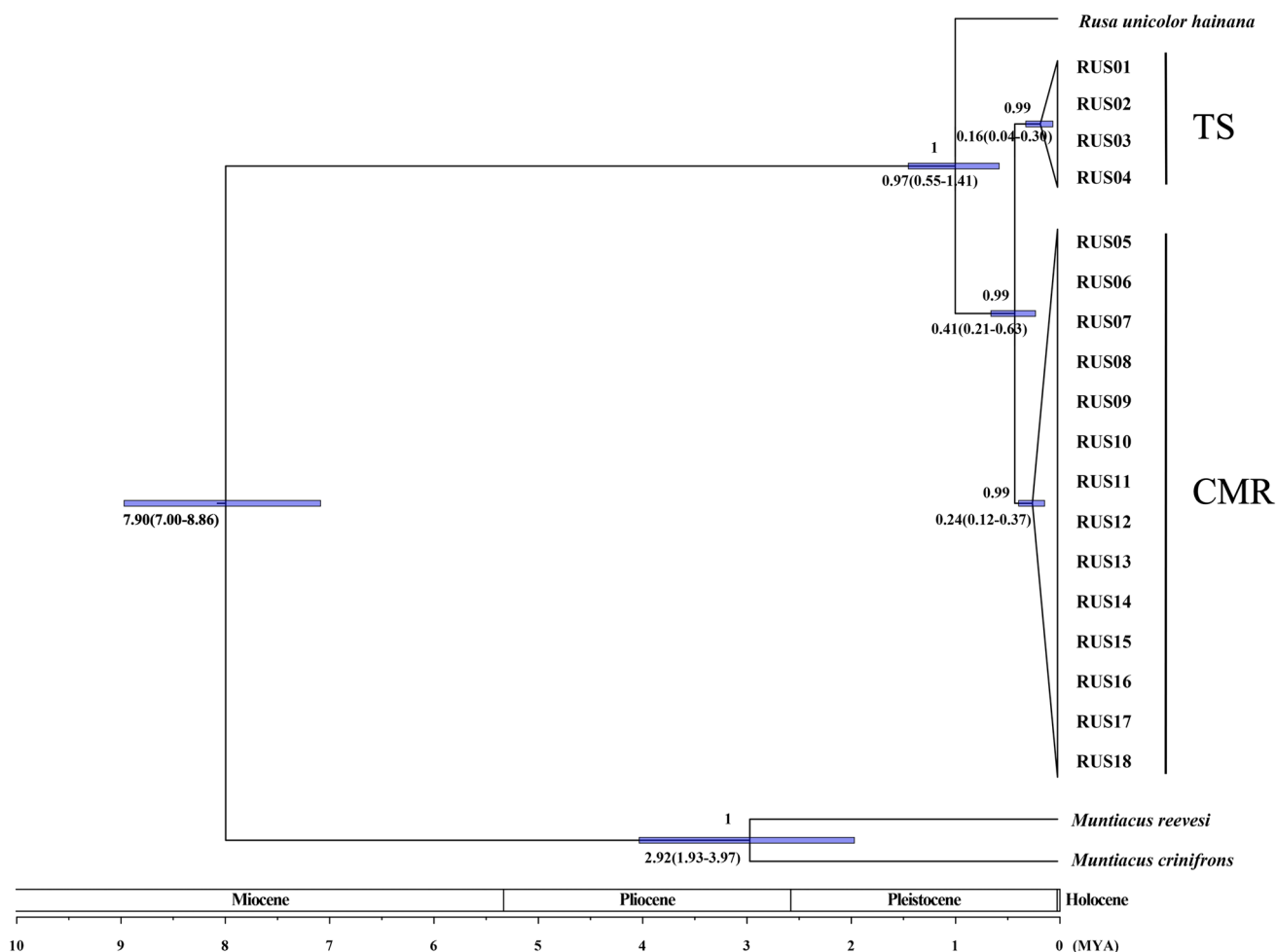
To better understand the demographic history of the Formosan sambar deer population, we applied neutrality tests for subpopulations, regions, and clades to detect population expansion (Table 2). Region CSR displayed significantly negative values for both Tajima's  $D$  ( $-1.84610$ ,  $P < 0.01$ ) and Fu's  $F_s$  ( $-12.08819$ ,  $P < 0.001$ ). Clade CMR showed a significantly negative value for Fu's  $F_s$  ( $-6.11141$ ,  $P < 0.05$ ) and a unimodal pattern in mismatch distribution (Fig. S3). Together, these results indicate that deer from regions CSR and STR, as well as deer in the CMR clade, have undergone historical population expansion.

We observed that the RUS05 and RUS06 haplotypes from the CMR clade occurred in regions NDD, NTR, and STR, i.e., where haplotypes from the TS clade are also found. Moreover, our neutrality tests and mismatch analysis revealed population

expansion for the CMR clade and the CSR region, indicating the possibility of secondary contact between CSR and more northerly regions. To explore the possible dispersal history of Formosan sambar deer in Taiwan, we used coalescent simulations to compare the probabilities of 10 potential scenarios (Fig. 5), which identified Scenario 4 as the most likely demographic history with a posterior probability of 0.5389 (a 95% credibility interval [CI] of 0.4544–0.6234). This scenario supports region CSR as being the likely origin of all others and experiencing secondary contact with STR to give rise to the NDD subpopulation. Other events encompassed by this scenario include splitting of CSR from other regions (t5), divergence of the STR and SSP+NSP regions (t4), secondary contact between the STR and CSR regions giving rise to the common ancestor of NDD (t3), divergence of the SSP and NSP regions (t2) and, finally, splitting of the STR and NTR regions (t1), all occurring in chronological order from most ancient to most recent.

### Discussion

Our SAMOVA revealed significant genetic differentiation among Formosan sambar deer from six defined regions. This result implies potential barrier effects between those six regions. However, minimal and maximal pairwise genetic distances and  $F_{ST}$  values did not correspond to geographic distances. This discordance between geographic and genetic distances could be due to non-mutually exclusive factors, such as (1) the dispersal capability of Formosan sambar deer



**Fig. 3** Fossil-calibrated Bayesian phylogenetic tree constructed from mitochondrial control region haplotypes. Bayesian posterior probabilities are shown above the nodes. Numbers below the nodes indicate

the divergence time (and 95% highest posterior density represented by a blue bar) to the most recent common ancestor of each clade. TS, Taroko-Sheipa clade; CMR, Central Mountain Range clade

could be higher than expected or (2) topological or climatic factors could restrict movement irrespective of dispersal capacity. To further explore potential contributory factors to the observed genetic patterning, we inferred the dispersal history of Formosan sambar deer.

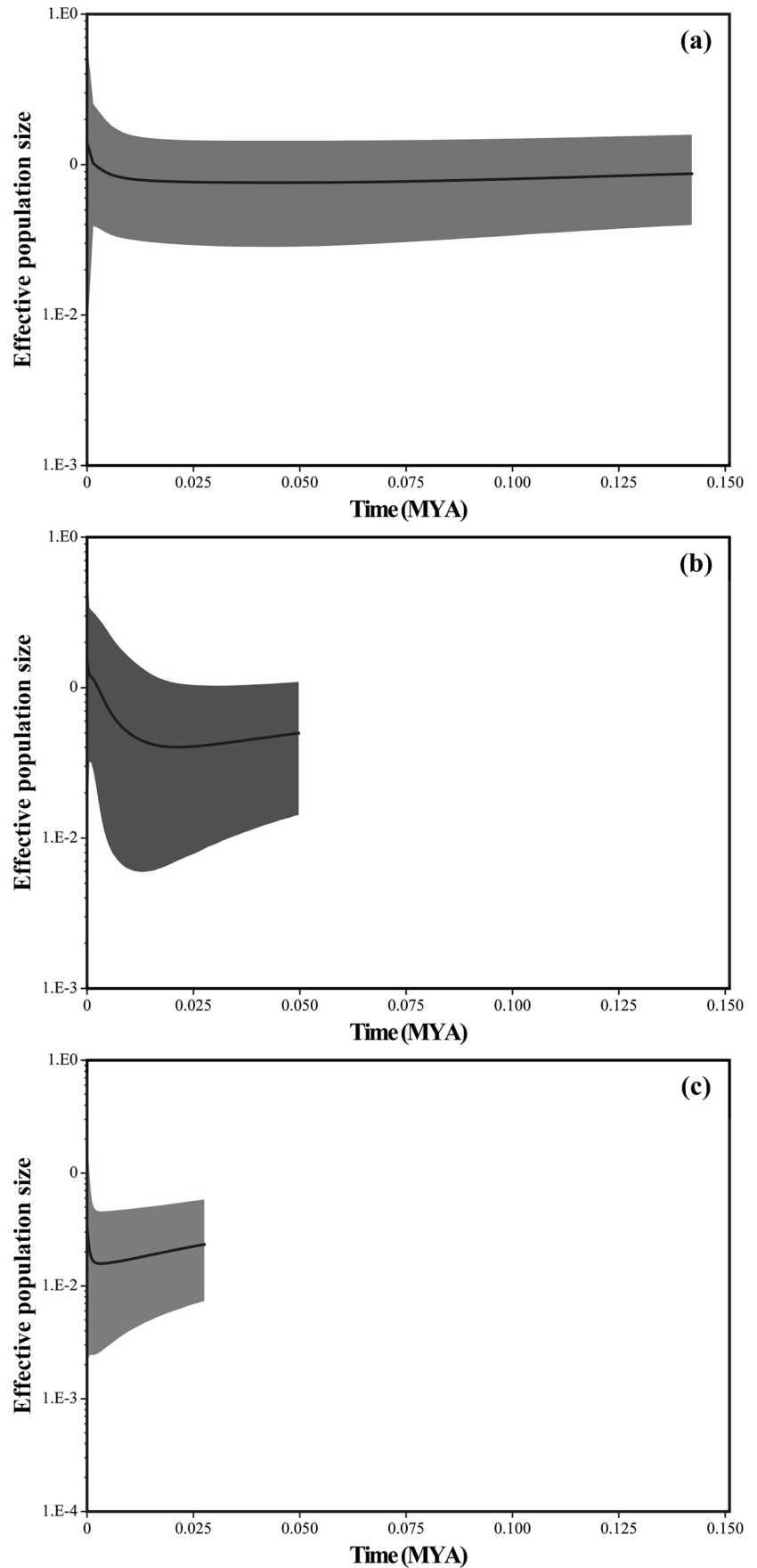
We noted that haplotype RUS07, which was assigned to the CMR clade, lay at the center of the star-like network and that it is distributed widely throughout the CSR region (Fig. 2). Moreover, we identified greater genetic diversity in the CMR clade relative to the TS clade based on  $h$  and  $\pi$  values (Table 2). Accordingly, we postulate that haplotype RUS07 represents the ancestral lineage for the two clades. Although scenarios linked to SP being the origin are also plausible, our approximate Bayesian computation (ABC) supports that CSR acted as the origin of the current subpopulations. Only two haplotypes, RUS01–02, occur in Sheipa National Park (the NSP and SSP regions), and haplotypes of the CMR clade are not found there, evidencing a

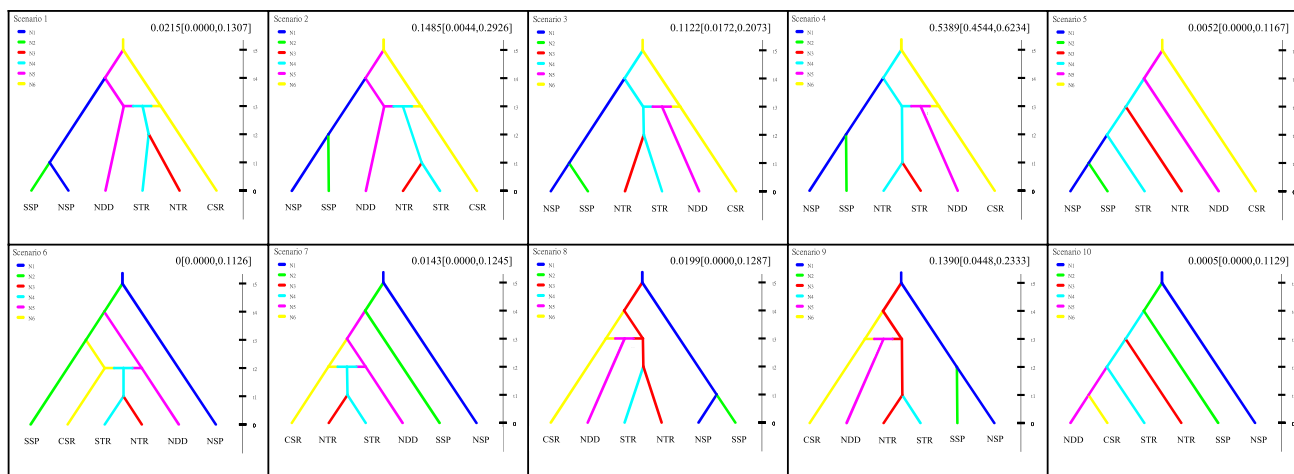
barrier to gene flow. Thus, from our haplotype phylogeny, we infer that the TS clade was derived from the CMR clade. The two clades are separated by three intermediate haplotypes (Fig. 2, indicated by three black dots), which were not detected among our samples. Neither RUS07 nor the three missing haplotypes were detected at NTR, STR, and NDD, indicating that deer hosting these haplotypes may be rare or have become extinct. This outcome implies that the trajectory of haplotype evolution RUS07 → RUS05 → RUS06 did not occur in TS, but arose from secondary contact due to sambar harboring RUS05 and RUS06 migrating into a region hosting the TS clade.

Haplotypes of both the TS and CMR clades (RUS02–RUS06) were detected in the NDD, NTR, and STR regions (Fig. 1; Supplementary Table S3), with secondary contacts potentially giving rise to this genetic admixture. Geographically, deer of the TS clade are primarily restricted to habitats north of the Chilai ridgeline (indicated



**Fig. 4** Bayesian skyline plots representing historical demographic trends for **a** the entire Formosan sambar deer population, **b** the Central Mountain Range clade, and **c** the Taroko-Sheipa clade. X-axis shows time before present in million years ago (MYA). Y-axis indicates effective population size on a logarithmic scale. The thick solid lines represent mean estimates for effective population size, and the differentially colored areas reflect the 95% highest posterior density intervals





**Fig. 5** Scenarios of lineage diversification and secondary contact for Formosan sambar deer based on mitochondrial control region sequences. Scenario 4 displayed the best fit to our dataset based on

approximate Bayesian computation (ABC). Posterior probabilities (PP) are shown for each scenario. Values in square brackets indicate the 95% highest posterior probability confidence interval

as Chilai-Nengkao in Fig. 1). In contrast, deer of the CMR clade occur south of it, except for deer with the RUS05 and RUS06 haplotypes that occur in the NTR, STR, NDD, and CSR regions. Our approximate Bayesian computation (ABC) analysis supports the hypothesis of the CSR being the origin of Formosan sambar subpopulations and that secondary contact between the STR and CSR regions gave rise to the NDD subpopulation (Fig. 5). Contact between subpopulations may have arisen from expansion of the CMR or TS clades, or both. The CMR clade presented a statistically significantly negative value for both Tajima's  $D$  and Fu's  $F_s$ , indicating that it had undergone population expansion (Table 2). In contrast, the same neutrality test results for the TS clade and the entire Formosan sambar deer population were not significant. These outcomes were supported by our BSP analysis, together revealing that the CMR and TS clades have experienced differing demographic trajectories. Sampling biases due to the rough terrain of the Taiwanese mountain ranges that limited sampling effort may have contributed to the fact that only one haplotype for each subpopulation was detected in SSP and NSP and, consequently, why we did not detect significant expansion for SP regions. If such biases exist, it is plausible that sambar in the STR, NTR, and NDD regions originally featured haplotypes of the CMR clade (RUS05 and RUS06) and came into contact with the TS clade (RUS03 and RUS04) during expansion of the TS. However, we assert that the frequency of missed haplotypes is not high, and if SP regions were the source of RUS03 and RUS04, the frequency of these two haplotypes should be high there. Yet, nowhere displays a sufficiently high frequency of haplotypes RUS03 and RUS04 that could have acted as the source of the TS clade, which could have expanded into NTR and STR. Instead, it is more likely that

haplotypes RUS03 and RUS04 detected in NTR and STR represent the origin of the TS clade where they evolved into RUS02 and RUS01, with these latter two haplotypes then dispersing into the SP region. This assumption is supported by the DIYABC result that depicts the CMR-origin scenario, with dispersal occurring from both south to north and from east to west. Hence, we suspect that dispersal of deer harboring the RUS05 and RUS06 haplotypes is responsible for the secondary contact between northern and southern Formosan sambar populations.

Based on all of our analyses, we present a likely scenario explaining the genetic patterning of Formosan sambar deer. We hypothesize that Formosan sambar deer first became established in the central or southern mountain ranges of Taiwan upon arrival from continental Asia and then dispersed northward. This northward dispersal resulted in the formation of two clades, i.e., CMR and TS. Based on molecular clock dating and fossil calibration, divergence between these two clades occurred 0.41 MYA, i.e., during the Nebraskan glacial stage (0.33–0.47 MYA) of the mid-Pleistocene, indicating that divergence of the TS clade coincided with a glacial period (Fig. 3). More recently, the ancestor of the TS clade migrated into the northern parts of the Central Mountain Range and into the region of Taroko National Park (from where we sampled the NTR and STR subpopulations) that are not separated by apparent geographic barriers. However, during the last glaciation, equilibrium-line altitudes (ELAs) of the Nanhu Mountain glacial cirque in Taroko National Park expanded, lowering snowlines (1500 m or less, Böse 2000, 2004; Hebenstreit et al. 2006). Thus, glaciers covering alpine hilltops formed closed-circle ELAs in the Taroko area, severely restricting southward migration of deer from the TS clade. Moreover, deer of the TS clade were forced to seek refuge in low-altitude canyons within Taroko where vegetation

could still be found. The walls of those canyons are surrounded by steep mountains and cliffs (Supplementary Fig. S3), which would have been capped by glaciers during glacial periods, creating topographic and climatic barriers to gene flow.

Radiotelemetry data and camera-trapping has revealed that Formosan sambar deer display altitudinal migration. In cold months, they migrate to lower elevations from the upland habitats they occupy during warmer months (Yen et al. 2013, 2019). This behavior implies that Formosan sambar may be sensitive to temperature, especially to cold weather. Temperatures during the Last Glacial Maximum (30,000–22,000 YA) were 8–11 °C lower than present (Tsukada 1967), so shifting of vegetation zones into lower elevations unsuitable for sambar deer likely induced genetic barriers during glacial times. Stress-induced selection and allotropic differentiation may have also resulted in the loss of intermediate haplotypes between the CMR and TS clades (represented as black dots in Fig. 2), facilitating their differentiation. However, during subsequent interglacial periods, the deer underwent post-glacial population expansion, allowing secondary contact between both clades. Hence, deer populations of the CMR and TS clades now occur sympatrically in the northern part of the Central Mountain Range (i.e., in the NTR, STR and NDD regions).

In conclusion, we have detected 18 mitochondrial control region haplotypes among Formosan sambar deer and assessed genetic differentiation across subpopulations. We have identified two primary haplotype lineages. Moreover, we have postulated a secondary contact event and predicted a likely evolutionary scenario for the demographic history of Formosan sambar. Finally, we have inferred dispersal histories in the context of Pleistocene climatic oscillations and speculated that the area of Taroko National Park acted as a refugium for Formosan sambar deer during glacial periods. Our study represents an example of how climate change plays an important role in patterns of genetic variation for large herbivorous mammals.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s13364-023-00682-6>.

**Author contribution** K. Y. L. and C. H. conceived and designed the experiments, performed the experiments, analyzed the data, and wrote the original draft. Y. Z. L., S. W. J., S. C. Y., M. H. H., Y. W., and S. W. C. contributed to sampling and manuscript revision. C. Y. H., P. J. Y., and K. L. C. contributed to molecular experiments and data analyses. G. J. W. and S. F. L. contributed to manuscript revision. Y. T. J. conceived and designed the experiments, revised the manuscript, and provided overall guidance. All authors have reviewed and approved the final manuscript.

**Funding** This work was supported by grants from the Ministry of Science and Technology of Taiwan (MOST 102–2313-B-002–024-MY3 and MOST 106–2313-B-002–048) and Taroko National Park of Taiwan (PG10101-0364, PG10202-0183, PG10303-0251, and PG10403-0179).

**Data Availability** The haplotypes of the Formosan sambar deer control region sequences, utilized in the present study, have been deposited into the National Center for Biotechnology Information (NCBI) GenBank. All data generated or analyzed during this study are included in this published article and its supplementary information files.

## Declarations

**Disclaimer** The funding organizations were not involved in experimental design.

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Angert AL, Sheth SN, Paul JR (2011) Incorporating population-level variation in thermal performance into predictions of geographic range shifts. *Integr Comp Biol* 51(5):733–750. <https://doi.org/10.1093/icb/acr048>
- Avice JC (2009) Phylogeography: retrospect and prospect. *J Biogeogr* 36(1):3–15. <https://doi.org/10.1111/j.1365-2699.2008.02032.x>
- Avice JC (2007) Twenty-five key evolutionary insights from the phylogeographic revolution in population genetics. In: W. S. and F. N (eds) *Phylogeography of Southern European Refugia*. Springer, AA Dordrecht, The Netherlands
- Bennett K, Provan J (2008) What do we mean by 'refugia'? *Quat Sci Rev* 27(27–28):2449–2455. <https://doi.org/10.1016/j.quascirev.2008.08.019>
- Böse M (2000) Glacial landforms in Taiwan and a reinterpretation of the last glacial snowline depression. In: Slaymaker O (ed) *Geomorphology, human activity and global environmental change*. Wiley, Chichester, pp 25–41
- Böse M (2004) Traces of glaciation in the high mountains of Taiwan. *Dev Quat Sci* 2:347–352. [https://doi.org/10.1016/S1571-0866\(04\)80141-6](https://doi.org/10.1016/S1571-0866(04)80141-6)
- Böse M, Hebenstreit R (2011) Late Pleistocene and Early Holocene glaciations in the Taiwanese High Mountain Ranges, Quaternary Glaciations — extent and chronology — a closer look. *Dev Quat Sci* 15:1003–1012
- Bouckaert R, Heled J, Kuhnert D et al (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10(4):e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Chan LM, Brown JL, Yoder AD (2011) Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Mol Phylogenet Evol* 59(2):523–537. <https://doi.org/10.1016/j.ympev.2011.01.020>

- Chavez AS, Maher SP, Arbogast BS, Kenagy GL (2014) Diversification and gene flow in nascent lineages of island and mainland North American tree squirrels (*Tamiasciurus*). *Evolution* 68(4):1094–1109. <https://doi.org/10.1111/evo.12336>
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9(10):1657–1659. <https://doi.org/10.1046/j.1365-294x.2000.01020.x>
- Cornuet JM, Pudlo P, Veyssier J et al (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30(8):1187–1189. <https://doi.org/10.1093/bioinformatics/btt763>
- DeMiguel D, Azanza B, Morales J (2010) Trophic flexibility within the oldest Cervidae lineage to persist through the Miocene Climatic Optimum. *Palaeogeogr Palaeoclimatol Palaeoecol* 289(1)–(4)
- Dong W, Pan Y, Liu J (2004) The earliest *Muntiacus* (Artiodactyla, Mammalia) from the Late Miocene of Yuanmou, southwestern China. *CR Palevol* 3(5):379–386. <https://doi.org/10.1016/j.crpv.2004.06.002>
- Drummond AJ, Suchard MA (2010) Bayesian random local clocks, or one rate to rule them all. *BMC Biol* 8(1):114. <https://doi.org/10.1186/1741-7007-8-114>
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Mol Ecol* 11(12):2571–2581. <https://doi.org/10.1046/j.1365-294X.2002.01650.x>
- Epps CW, Palsboll PJ, Wehausen JD, Roderick GK, McCullough DR (2006) Elevation and connectivity define genetic refugia for mountain sheep as climate warms. *Mol Ecol* 15(14):4295–4302. <https://doi.org/10.1111/j.1365-294X.2006.03103.x>
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Gilbert C, Ropiquet A, Hassanin A (2006) Mitochondrial and nuclear phylogenies of Cervidae (Mammalia, Ruminantia): Systematics, morphology, and biogeography. *Mol Phylogenet Evol* 40(1):101–117. <https://doi.org/10.1016/j.ympev.2006.02.017>
- Gorelick R (2008) Species richness and the analytic geometry of latitudinal and altitudinal gradients. *Acta Biotheor* 56(3):197–203. <https://doi.org/10.1007/s10441-008-9048-7>
- Hebenstreit R, Böse M, Murray A (2006) Late Pleistocene and early Holocene glaciations in Taiwanese mountains. *Quat Int* 147(1):76–88. <https://doi.org/10.1016/j.quaint.2005.09.009>
- Heled J, Drummond AJ (2011) Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Syst Biol* 61(1):138–149. <https://doi.org/10.1093/sysbio/syr087>
- Husemann M, Schmitt T, Zachos FE, Ulrich W, Habel JC, Riddle B (2014) Palaeoarctic biogeography revisited: evidence for the existence of a North African refugium for Western Palaeoarctic biota. *J Biogeogr* 41(1):81–94. <https://doi.org/10.1111/jbi.12180>
- Li CF, Chytrý M, Zelený D et al (2013) Classification of Taiwan forest vegetation. *Appl Veg Sci* 16(4):698–719. <https://doi.org/10.1111/avsc.12025>
- Liang Y, He D, Jia Y, Sun H, Chen Y (2017) Phylogeographic studies of schizothoracine fishes on the central Qinghai-Tibet Plateau reveal the highest known glacial microrefugia. *Sci Rep* 7(1):10983. <https://doi.org/10.1038/s41598-017-11198-w>
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>
- Lin HD, Hsu KC, Shao KT et al (2008) Population structure and phylogeography of *Aphyocypris kikuchii* (Oshima) based on mitochondrial DNA variation. *J Fish Biol* 72(8):2011–2025. <https://doi.org/10.1111/j.1095-8649.2008.01836.x>
- Lin HD, Chen YR, Lin SM (2012) Strict consistency between genetic and topographic landscapes of the brown tree frog (*Buergeria robusta*) in Taiwan. *Mol Phylogenet Evol* 62(1):251–262. <https://doi.org/10.1016/j.ympev.2011.09.022>
- Oshida T, Lin LK, Chang SW, Chen YJ, Lin JK (2011) Phylogeography of two sympatric giant flying squirrel subspecies, *Petaurista alborufus lena* and *P. philippensis grandis* (Rodentia: Scuridae), in Taiwan. *Biol J Linn Soc Lond* 102(2):404–419. <https://doi.org/10.1111/j.1095-8312.2010.01576.x>
- Potts R (2013) Hominin evolution in settings of strong environmental variability. *Quat Sci Rev* 73:1–13. <https://doi.org/10.1016/j.quascirev.2013.04.003>
- Rambaut A, Suchard MA, Xie W, Drummond AJ (2013) Tracer v.1.6.0: MCMC Trace Analysis Tool. <http://tree.bio.ed.ac.uk/software/tracer/>. Accessed 2022/12
- Rambaut A (2014) FigTree v1.4.2: Tree figure drawing tool. <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 2022/12
- Randi E, Pierpaoli M, Danilkin A (1998) Mitochondrial DNA polymorphism in populations of Siberian and European roe deer (*Capreolus pygargus* and *C. capreolus*). *Heredity (Edinb)*. 80(Pt 4):429–37. <https://doi.org/10.1046/j.1365-2540.1998.00318.x>
- Sommer RS, Zachos FE (2009) Fossil evidence and phylogeography of temperate species: ‘glacial refugia’ and post-glacial recolonization. *J Biogeogr* 36(11):2013–2020. <https://doi.org/10.1111/j.1365-2699.2009.02187.x>
- Stewart JR, Lister AM, Barnes I, Dalen L (2010) Refugia revisited: individualistic responses of species in space and time. *Proc Biol Sci* 277(1682):661–671. <https://doi.org/10.1098/rspb.2009.1272>
- Strani F, DeMiguel D, Bellucci L, Sardella R (2018) Dietary response of early Pleistocene ungulate communities to the climate oscillations of the Gelasian/Calabrian transition in Central Italy. *Palaeogeogr Palaeoclimatol Palaeoecol* 499:102–111
- Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol Biol Evol* 38:3022–3027
- Tsukada M (1967) Vegetation in subtropical Formosa during the pleistocene glaciations and the holocene. *Palaeogeogr Palaeoclimatol Palaeoecol* 3:49–64. [https://doi.org/10.1016/0031-0182\(67\)90005-3](https://doi.org/10.1016/0031-0182(67)90005-3)
- Valente LM, Etienne RS, Phillimore AB (2014) The effects of island ontogeny on species diversity and phylogeny. *Proc Biol Sci* 281(1784):20133227. <https://doi.org/10.1098/rspb.2013.3227>
- Yen SC, Wang Y, Ou HY (2013) Habitat of the vulnerable Formosan sambar deer *Rusa unicolor swinhoii* in Taiwan. *Oryx* 48(2):232–240. <https://doi.org/10.1017/s0030605312001378>
- Yen SC, Wang Y, Yu PH, Kuan YP, Liao YC, Chen KH, Weng GJ (2019) Seasonal space use and habitat selection of sambar in Taiwan. *J Wildl Manag* 83(1):22–31. <https://doi.org/10.1002/jwmg.21578>
- Yuan SL, Lin LK, Oshida T (2006) Phylogeography of the mole-shrew (*Anourosorex yamashinai*) in Taiwan: implications of interglacial refugia in a high-elevation small mammal. *Mol Ecol* 15(8):2119–2130. <https://doi.org/10.1111/j.1365-294X.2006.02875.x>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.