RESEARCH



Population genomics of sika deer reveals recent speciation and genetic selective signatures during evolution and domestication

Huamiao Liu^{1†}, Bo Zhu², Tianjiao Wang^{1†}, Yimeng Dong^{1†}, Yan Ju¹, Yang Li¹, Weilin Su¹, Ranran Zhang¹, Shiwu Dong¹, Hongliang Wang¹, Yongna Zhou¹, Yanmin Zhu¹, Lei Wang¹, Zhengyi Zhang¹, Pei Zhao¹, Shuyan Zhang³, Rui Guo³, E. A⁴, Yuwen Zhang⁵, Xin Liu⁶, Hidetoshi B. Tamate^{7*}, Qiqi Liang^{8*}, De Ma^{2*} and Xiumei Xing^{1*}

Abstract

Background Population genomic analysis can reconstruct the phylogenetic relationship and demographic history, and identify genomic selective signatures of a species. To date, fundamental aspects of population genomic analyses, such as intraspecies taxonomy, evolutionary history, and adaptive evolution, of sika deer have not been systematically investigated. Furthermore, accumulating lines of evidences have illustrated that incorrect species delimitation will mislead conservation decisions, and even lead to irreversible mistakes in threatened species.

Results In this study, we resequenced 81 wild and 71 domesticated sika deer representing 10 main geographic populations and two farms to clarify the species delimitation, demographic and divergence histories, and adaptive evolution of this species. First, our analyses of whole genomes, Y chromosomes and mitochondrial genomes revealed substantial genetic differentiation between the continental and Japanese lineages of sika deer, representing two phylogenetically distinct species. Second, sika deer in Japan were inferred to have experienced a "divergence-mixing-isolation" evolutionary scenario. Third, we identified four candidate genes (*XKR4*, *NPAS3*, *CTNNA3*, and *CNTNAP5*) possibly involved in body size regulation of sika deer by selective sweep analysis. Furthermore, we also detected two candidate genes (*NRP2* and *EDIL3*) that may be associated with an important economic trait (antler weight) were under selection during the process of domestication.

Conclusion Population genomic analyses revealed that the continental and Japanese lineages represent distinct phylogenetic species. Moreover, our results provide insights into the genetic selection signatures related to body size differences and a valuable genomic resource for future genetic studies and genomics-informed breeding of sika deer.

[†] Huamiao Liu, Bo Zhu, Tianjiao Wang and Yimeng Dong contributed equally to this work.

*Correspondence: Hidetoshi B. Tamate tamate@sci.kj.yamagata-u.ac.jp Qiqi Liang qiqi.liang@glbizzia.com De Ma made08@mails.ucas.edu.cn Xiumei Xing xingxiumei@caas.cn Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

Keywords Sika deer, *Cervus nippon*, Population genomics, Species delimitation, Genetic selective signature, Domestication

Introduction

Population genomic analysis is crucial for reconstructing the phylogenetic relationship and demographic history, and identifying genomic signatures underlying selection of a species [1, 2]. Native to East Asia, sika deer (*Cervus nippon*) were once widely distributed in the eastern Eurasian mainland (Far East of Russia, China, the Korean Peninsula, and Vietnam), Taiwan island and the Japanese archipelago [3, 4]. However, wild sika deer populations have gradually declined due to human activities such as overexploitation via hunting and habitat destruction caused by urbanization [5, 6], and they are only distributed in nature reserves in some areas. To date, population genomics research on this specie is scarce [7], as it is distributed in different countries and has limited wild populations, making sampling difficult.

Based on morphological characteristics (such as antlers, skulls, body size, and coat color pattern), this species has been taxonomically classified into 13 subspecies: six subspecies in eastern Eurasian continent and Taiwan island, six in Japan, and one in Vietnam [8-10]. Of the six subspecies that have inhabited eastern Eurasian continent, two (C. n. grassianus and C. n. mandarinus) are considered extinct today. The remaining four subspecies are (1) C. n. hortulorum, which is distributed in northeastern China, the Far East of Russia, and the Korean Peninsula; (2) C. n. sichuanicus, which lives in the Tiebu Nature Reserve in Sichuan Province; (3) C. n. kopschi, which lives in the Qingliangfeng Nature Reserve in Zhejiang and the Taohongling Nature Reserve in Jiangxi; and (4) C. n. taiouanus, which is extinct in the wild but currently lives in Kenting National Park after restoration from captive-bred wild individuals. In Japan, sika deer are classified into six subspecies [8]: (1) C. n. yesoensis (Hokkaido Island), (2) C. n. centralis (Honshu mainland and Tsushima Island), (3) C. n. nippon (Kyushu, Shikoku, and Goto Islands), (4) C. n. mageshimae (Mageshima and Tanegashima Islands), (5) C. n. yakushimae (Yakushima and Kuchinoerabu Islands), and (6) C. n. keramae (Ryukyu Islands). The last subspecies, C. n. pseudaxis is distributed in Vietnam but is extinct in the wild; it is only raised on deer farms and relocated to parks, reserves, and zoos (pp. 541–548 in [10]).

Fundamental aspects of evolutionary and conservation biology, such as intraspecies taxonomy, evolutionary history, and adaptive evolution, of this species have not been investigated by genome-wide single nucleotide polymorphism (SNP) markers but only by mitochondrial gene fragments [5, 11–15] and Y-chromosome genes [16], which have some limitations [17, 18]. Accumulating lines of evidences have illustrated that incorrect species delimitation will mislead conservation decisions, such as protection and management practices, and even lead to irreversible mistakes in threatened species [19–21]. SNPs identified by whole-genome sequencing show great potential not only to assist in species delimitation and phylogeographic evolutionary analyses at an appropriate temporal scale [19] but also to help elucidate micro-evolutionary mechanisms on a population scale [22], for instance, dissecting genomic signatures of selection and local adaptation [21, 23].

In addition, ecological pressures and environmental conditions could influence the body size evolution of vertebrates living on islands. The classic insular pattern characterized by gigantism in small animals and dwarfism in large animals has been summarized as a macroevolutionary or biogeographic concept, the so-called "island rule" [24-26]. However, the microevolutionary mechanism driving patterns of insular dwarfism in large animals remains largely elusive. Since body size is intimately linked to many physiological and ecological characteristics of vertebrates, it is predictable that shifts in body size would leave signals of selection and local adaptation in the genome during the evolution of a species. For instance, a recent study revealed that selective signatures in the fatty acid desaturase (FADS) gene cluster, likely related to diet, are associated with stature size differences between an extant pygmy population on Flores Island and modern humans, which provided a paradigm to bridge the gap between selective signatures in the genome and insular dwarfism in humans [27]. Overall, research on screening genes related to body size differences between island and mainland mammals is still scarce. Furthermore, it is also unknown whether the genes regulating body size related to the "island rule" in humans and other mammals are conserved. Paired island-mainland populations of mammals, birds, reptiles and amphibians are frequently used as models for exploring the patterns of body size evolution in terrestrial vertebrates [26]. In this study, we have been able to sample populations from the eastern Eurasian continent and the Japanese archipelago, which provide an excellent model for exploring the microevolutionary mechanism of this phenomenon.

Furthermore, the sika deer has a history of domestication and breeding for more than 300 years in China [28], and antler has been used as a traditional Chinese medicine for more than 2,000 years [29, 30]. Due to their economic and medicinal value, sika deer are commonly raised in captivity in eastern Asia. During the domestication process, antler weight is an important target of artificial selection, resulting in a 30–60% increase in antler weight in artificially bred populations compared to uncultivated ones [28]. However, the molecular mechanism regulating the weight of velvet antler has not been well explained.

Recently, we assembled the first high-quality chromosome-level reference genome of sika deer, which represents a valuable genetic resource for research on and the conservation of this species [4]. In the present study, we resequenced the genomes of 81 wild and 71 domesticated sika deer representing 10 main geographic populations and two farms with an average depth of $26.7 \times$. We used these data to clarify the species delimitation, demographic and divergence histories, and adaptive evolution of wild sika deer. Furthermore, we identified potential candidate genes that may underlie the observed body size variations among population pairs from the Japanese archipelago. Finally, we characterized the signatures of selection related to antler size during the domestication of sika deer.

Results

Whole-genome sequencing and genetic variation

We resequenced the genomes of 81 wild sika deer from 10 sampling sites, including a nature reserve (RU1, n = 16) and a wild region (RU2, n = 6) in the Far East of Russia, the Tiebu Nature Reserve in Sichuan Province (SC, n = 8), the Qingliangfeng National Nature Reserve in Zhejiang Province (ZJ, n = 8), the Liugongdao National Forest Park in Shandong Province (TW, n = 2), Hokkaido (HK, n = 9), Nikko in Honshu (NK, n = 8), Shimane in Honshu (SM, n = 2), Tsushima Island (TS, n = 14), and Yakushima Island (YAK, n = 8). The last five sampling sites are located in Japan. In addition, a total of 71 domesticated sika deer from farms in Jilin and Heilongjiang Provinces (Northeast China, NEC) were also collected for genome resequencing (Fig. 1A and Supplementary table S1).

A total of 13.22 Tb of high-quality Illumina clean reads were generated (Supplementary table S2). Genome mapping showed that approximately 98.46% of the reads covered 98.68% of the reference genome, with an average genome depth of $26.7 \times$. After SNP calling and quality filtering, we identified a total of 31,320,151 SNPs. Among these SNPs, 242,811 were located within exonic regions, 9,605,504 in intronic regions, 796 in splicing sites, 430,223 in upstream or downstream regions, and the remaining 21,040,817 in intergenic regions. In the coding regions, we annotated 102,530 nonsynonymous, 218 stop-loss, and 1,799 stop-gain SNPs that cause amino acid changes, elongated transcripts or premature termination (Supplementary table S3). In addition, we also detected 2,034,186 insertions and deletions (InDels) with a minor allele frequency (MAF)>0.05 (Supplementary table S4). Genetic variants of sika deer along the 33 chromosomes indicated they were evenly distributed across the genome (Supplementary Figure S1 and S2).

Since our genome assembly was from a female individual [4], the Y chromosome of another sika deer genome (GCA_040085125.1) in National Center for Biotechnology Information (NCBI) RefSeq database (https://www. ncbi.nlm.nih.gov/datasets/genome/GCA_040085125.1/) was used to call Y-chromosome (Y-chr) SNPs. We identified a total of 543,807 Y-chr SNPs in 129 male individuals. In addition, we assembled the mitochondrial genome of each individual from the genome resequencing reads of 81 wild and 71 domesticated sika deer and detected 914 mitochondrial SNPs.

The average SNP densities were 12.53/kb, 16.67/kb, and 19.72/kb in autosomes, X-chromosomes, and Y-chromosomes, respectively. This result indicated that the nucleotide diversities of nonrecombining sex chromosomes were higher than those of autosomes, which contrasted with the observed pattern in some livestock species where sex chromosomes exhibit reduced nucleotide diversity compared to autosomes due to selective breeding [31, 32]. This distinction was possibly due to that most of our samples were wild (n=81), and although 71 domesticated samples were also included, the domesticating and breeding history of sika deer was much shorter than other livestock.

In addition, we detected a total of 132,644 SVs, including insertions (INS), deletions (DEL), intrachromosomal translocations (ITX), interchromosomal translocations (CTX), and inversions (INV) (Supplementary tables S5, S6). In total, 48,570 copy number variant (CNV) were also identified (Supplementary table S7).

Genomic evidence of two phylogenetic species of sika deer Based on the autosomal SNPs, we performed phylogenetic tree construction, population structure analysis, and principal component analysis (PCA) (Fig. 1B-D). In the phylogenetic tree, we found two divergent genetic lineages, the continental lineage and the Japanese lineage, in sika deer. The continental lineage included four clades, northern (NEC, RU1 and RU2), SC, ZJ, and TW, which was consistent with their geographic origin, corresponding well to the four subspecies *C. n. hortulorum, C. n. sichuanicus, C. n. kopschi*, and *C. n. taiouanus* as classified in traditional taxonomy [9]. On the other hand, populations from HK, NK, SM, TS, and YAK were classified into five clades, with individuals in each population



Fig.1 Geographic distributions and population genetic structure of the sika deer. **A** Sampling locations in this study. A total of 81 wild sika deer representing populations from China, Russia, and Japan and 71 domesticated sika deer from farms in Jilin and Heilongjiang Provinces (Northeast China, NEC) were sampled. Triangles and circles were used to distinguish between mainland and island populations, and two domestic populations were labelled with black edges. The black squares denoted non-native populations: the Kazakhstan population was introduced from a nature reserve in the Far East of Russia (RU1) in the 1930s, whereas the population in the Liugongdao National Forest Park, Shandong province, was introduced from Taiwan on April 16, 2011. **B** NJ tree constructed using p-distances of all 152 individuals. **C** ADMIXTURE results obtained using all 152 sika deer with K=2, 3 and 8. **D** Principal component analysis plot of all 152 samples used in this study. **E** Phylogenetic tree based on mitochondrial genome haplotypes. **F** Median-joining network map based on Y-chromosome SNP haplotypes

forming a single clade (Fig. 1B). In Japan, six subspecies have been reported for sika deer, but unfortunately, only three of these subspecies were sampled during our

efforts. The five clades together formed the Japanese lineage, which displayed a smaller cluster size with a shorter branch length than the continental lineage, indicating that the genetic differentiation within the Japanese lineage was relatively shallow. Moreover, the species tree reconstructed with SNAPP [33, 34] also demonstrated that the continental lineage and Japanese lineage were reciprocally monophyletic and the divergence events within the Japanese lineage occurred within the last 0.1 million years (Ma) (Supplementary Figure S3 and S4).

The major lineage classification was further validated through ADMIXTURE analysis (Fig. 1C and Supplement Figure S5). Although cross-validation (CV) error minimization indicated K=3 as the optimal value (Supplement Figure S6), examination of alternative K-values provided additional insights into the admixture history of sika deer populations. At K = 2, a clear genetic divergence emerged between the continental and Japanese lineages. At K=3, this primary division persisted while additionally resolving a secondary divergence between northern (RU1, RU2, NEC) and southern (SC, ZJ, TW) subclades within the continental lineage. At K=8, finer-scale population structure was resolved: within the continental lineage, the SC and ZJ populations exhibited relatively homogeneous genetic composition, whereas RU1, RU2, NEC and TW displayed admixture. Similarly, in the Japanese lineage, the HK, YAK and TS populations showed limited admixture, while NK and SM were mixed populations (Fig. 1C and Supplementary Figure S5).

The PCA revealed clear genetic structure, with samples from each geographic region clustering together (Fig. 1D). The first component separated the continental lineage from the Japanese lineage. Within the continental lineage, separation was detected between the northern (including NEC, RU1 and RU2) and southern (SC, TW and ZJ) populations in eastern Eurasian continent and Taiwan island along the axis of the second component. Our PCA results is consistent with the pattern revealed by the phylogenetic tree and ADMIXTURE.

Complementing autosomal SNP evidence, both uniparental inheritance systems - maternally transmitted mitochondrial DNA and paternally inherited Y chromosome – consistently resolved the deep evolutionary split between continental and Japanese sika deer lineages (Fig. 1E-F and Supplementary Figure S7). Phylogenetic reconstruction of mitochondrial haplotypes (39 haplotypes defined by 914 SNPs) revealed reciprocally monophyletic clades (Fig. 1E and Supplementary Figure S7A), while Y chromosome analyses integrating phylogenetic topology (52 haplotypes derived from 211,326 SNPs filtered with TrimAl) and median-joining network patterns (35 haplotypes resolved by 1,248 high-confidence SNPs after removing heterozygous sites and sites with missing genotypes) further confirmed complete lineage partitioning, with no haplotypes shared between the two groups (Fig. 1F and Supplementary Figure S7B).

To further evaluate whether the two lineages represent distinct species, we conducted a coalescent model-based species delimitation analysis using BPP v. 4.3.0 [35]. The results unequivocally supported their delimitation as distinct species, with posterior probabilities (PP) of 1.0 for the two-species model and 0 for the one-species model. A PP of 1.0 indicated that the genomic data decisively rejected lineage coalescence under a shared ancestral population, strongly favoring independent evolutionary histories. Under the phylogenetic species concept (PSC) [36], species are defined as irreducible monophyletic lineages diagnosable by fixed characters or coalescent thresholds. The BPP results was consistent with PSC since the two lineages form reciprocally monophyletic clades in the species tree inference.

Population genetic diversity

To assess the genetic diversity of different populations of sika deer, we calculated the genetic differentiation statistic (F_{ST}), nucleotide diversity ($\theta \pi$), observed heterozygosity (Het), linkage disequilibrium (LD), and inbreeding coefficient of each population (Fig. 2). The TW and SM populations were excluded from the genetic diversity analysis since only two samples were collected at both sites; thus, we were unable to provide an accurate assessment for these two populations.

 $F_{\rm ST}$ measures based on whole-genome SNPs showed that $F_{\rm ST}$ between the continental and Japanese lineages ranged from 0.24 to 0.55, which was much higher than that between populations within the same lineage (0.19–0.43 within the Japanese lineage, 0.02–0.26 within the continental lineage) (Fig. 2A).

The average $\theta\pi$ value of the populations from the continental lineage was $1.62-2.89 \times 10^{-3}$ (Fig. 2B), higher than that of the Japanese lineage $(0.77-1.32 \times 10^{-3})$. The $\theta\pi$ value of populations in the Japanese lineage was comparable to that of the threatened artiodactyl *Budorcas taxicolor* distributed in East Asia $(0.45-0.75 \times 10^{-3})$ [21], suggesting that the Japanese lineage experienced a severe bottleneck effect and still harbors low genetic diversity. The value of observed heterozygosity (Het) showed large variation among populations in the continental lineage, possibly due to the situation that these populations live in nature reserves or farms with different area sizes and engage in non-random mating (Fig. 2C and Supplementary Fig. 8).

In addition, populations in the Japanese lineage showed an overall slow decay rate and a high level of LD (as measured by r^2), whereas the populations in the continental lineage exhibited a rapid decay rate and a low level of LD (Fig. 2D). The difference in the genome-wide LD patterns between populations is consistent with the larger effective population size (*Ne*)



Fig. 2 Population differentiation (F_{ST}), genetic diversity, linkage disequilibrium, and inbreeding coefficient of different sika deer populations. **A** Genetic differentiation statistics (F_{ST}) of different sika deer populations. **B**, **C**. nucleotide diversity ($\theta \pi$) (**B**) and Observed heterozygosity (**C**) of different sika deer populations. **D** Genome-wide average linkage disequilibrium decay estimated from different sika deer populations. **E** Inbreeding coefficient of the 9 sika deer populations. The inbreeding coefficient is calculated as the fraction of runs of homozygosity (ROH). The TW and SM populations were excluded in these analyses since only two samples were collected at both sites; thus, we were unable to provide an accurate assessment for these two populations

for the continental lineage compared with the Japanese lineage. Moreover, the assessment of inbreeding level showed that populations in the Japanese lineage had higher inbreeding coefficient (F values: $1.83-4.56 \times 10^{-6}$) than those in the continental lineage ($0.64-2.93 \times 10^{-6}$) (Fig. 2E).

Demographic history

To explore the demographic histories of sika deer, we inferred historical population size changes using the pairwise sequential Markovian coalescent (PSMC) model based on the distribution of heterozygous sites across the genome [37]. PSMC analysis showed that the

demographic histories of sika deer can be traced back from the early Holocene (~10,000 years ago) to the middle Miocene (~10 Ma). The historical trends of *Ne* for different populations (Eurasian mainland vs. the Japanese archipelago) of sika deer showed a distinct pattern, especially during the period from the late Pliocene to the late Pleistocene (Fig. 3A).

We then used the multiple sequentially Markovian coalescent (MSMC) approach [38] to estimate the divergence time between the continental and Japanese



Fig. 3 Demographic history of different sika deer populations. **A** The ancestral population size of each population was inferred using a PSMC model. A generation time (g) = 5 years and neutral mutation rate (μ) = 2.2 × 10⁻⁹ per nucleotide site per generation were used. **B** Relative cross-coalescence rates (RCCR) over time between populations. Each line depicts the RCCR profile estimated using pairs of high-coverage genomes. The pairwise comparisons were between populations from Japan and RU1. **C** FastSIMCOAL2 v.2.6 simulation used to reconstruct the divergence, admixture, and demographic history of different sika deer populations. **D** TreeMix analysis of populations in Japan using RU1 as the outgroup when m = 2. **E** *D*-statistics with the form *D* (A, B; X, Y) for populations in Japan using RU1 as the outgroup. Positive *D*-statistic values indicate gene flow from X to A, whereas negative values indicate gene flow from X to B

lineages. We detected a decrease in the cross-coalescence rate between populations from the Japanese lineage and RU1 to 0.5 approximately 0.2-0.7 Ma (Fig. 3B). Furthermore, we also performed FastSIMCOAL2 v.2.6 analysis to infer demographic histories and estimate divergence times. The divergence between the continental and Japanese lineages occurred at approximately 0.60 Ma (Fig. 3C). Intriguingly, this divergence time coincided with the occurrence of the Naynayxungla Glaciation (0.72-0.50 Ma, corresponding to Marine Isotopic Stages (MIS) 14-18), which is speculated to be the most extensive glaciation among the four major glaciations during Quaternary climatic history [39]. Moreover, gene flow was detected from the population in southern China (ZJ) to the Japanese population in the late Pleistocene (Fig. <u>3</u>C).

Next, we investigated potential gene flow among populations in the Japanese archipelago using two different approaches. In the TreeMix analysis, which included populations from Japan and population RU1, we detected ancestral gene flow under the optimal migration parameter (m=2, as determined by OptM) (Supplementary Figure S9). Specifically, gene flow was inferred from the common ancestor of NK and HK to SM, as well as from TS to the common ancestor of NK and HK (Fig. 3D). Furthermore, *D*-statistics showed significant values (|Z-score|>3) indicating gene flow from NK to HK, from NK to SM, from TS to YAK, from TS to SM, and from YAK to SM (Fig. 3E).

Genomic selection signatures associated with body size differences between different populations of wild sika deer

To explore the genetic mechanism of selection and local adaptation, we used two population pairs, YAK vs NK+SM and TS vs NK, to detect selective sweep signatures associated with body size (Supplementary Figure S10). We used these two population pairs for the following reasons. First, comparisons between population pairs that are genetically very similar would help to minimize the effects of population genetic differences that are not related to the selection signatures of interest. By reconstructing the phylogenetic relationship, we found that populations from HK, NK, SM, TS, and YAK were clustered together and each clade displayed a shorter branch length than the continental lineage, indicating that the genetic differentiation within the Japanese lineage was relatively shallow (Fig. 1B, D). Second, different populations in the Japanese lineage showed similar demographic histories (Fig. 3A). Therefore, using population pairs within the Japanese lineage could also minimize the effects of demographic changes in population size, such as bottlenecks or population expansions, on selective sweep analysis. Third, since both the TS and YAK populations are characterized by their smaller body size compared to the other populations in their respective pairs, the candidate genes under selection they share are likely to be linked to the small body size trait. Last, we used a population pair (TS vs NK) with relatively close latitudes, so as to reduce the influence caused by Bergman's rule [40]. Due to similar body size [41], we did not conduct selective sweep analysis between TS and SM populations.

Furthermore, it is widely accepted that all methods detecting signatures of positive selection rely on different explicit or implicit assumptions and thus have limitations [42]. Therefore, using a range of methods in recombination is a powerful approach to resolving different forms of selection in practice [43]. In this study, we used five selective sweep methods, including the genetic differentiation statistic F_{ST} , the genetic diversity $\theta \pi$, Tajima's D, XP-CLR, and XP-EHH, to identify genomic regions under selection. The selection signature plots displaying all the outlier results across the chromosomes detected by five selective sweep methods were shown in Fig. 4A and B, and identified genes were listed in Supplementary table S8-S17. Those supported by all five methods (Supplementary Figure S11A and S11B) were considered as candidate genes associated with the body size. We identified 131 and 117 candidate genes in YAK vs NK+SM and TS vs NK, respectively, and 12 genes were shared between these two population pairs (Fig. 4C, Supplementary table S18-S22). Among the 12 candidate genes, four (XKR4, NPAS3, CTNNA3, and CNTNAP5) were supported in published papers suggesting a possible correlation with body size regulation (see details in discussion) [44–53]. Selective signals detected by the five selective

(See figure on next page.)

Fig. 4 Genomic regions with strong selection signals in the Yakushima (YAK) and Tsushima Island (TS) populations. **A**, **B** Selection signature plots displaying all the outlier results across the chromosomes detected by five selective sweep methods (FST, XP-CLR, XP-EHH, θπ, and Tajima's D) in population pairs NK+SM vs YAK (A) and NK vs TS (B). **C** Candidate genes shared between population pairs NK+SM vs YAK and NK vs TS. **D**, **E** Selective signals detected by five selective sweep methods around the XKR4 gene region in YAK vs NK+SM populations (D) and TS vs NK populations (E). **F**, **G** The patterns of genotypes of the XKR4 gene region in YAK and NK+SM populations (F), and in TS and NK populations (G) based on 833 SNPs



Fig. 4 (See legend on previous page.)





Fig. 5 Genomic regions with strong selection signals in the domesticated sika deer population. **A** Selection signature plots displaying all the outlier results across the chromosomes detected by five selective sweep methods (F_{ST} , XP-CLR, XP-EHH, $\theta\pi$, and Tajima's D) in population pair wild vs domesticated. **B** Genes supported by more than 3 methods were considered as candidate genes in population pair wild vs domesticated. **C** Overlapping genes from three datasets, namely, genes detected by selective sweep analysis in this study, genes detected with transcriptomic data from different antlers [56], and antler-specific genes from a previous study [57], were used as candidate genes related to antler size to decrease the false-positive rate. **D** Selective signals detected by five selective sweep methods around the *NRP2* (left panel) and *EDIL3* (right panel) gene regions in domesticated populations

sweep methods around these gene regions in YAK and TS populations were shown in Fig. 4D and E. Our data revealed a striking pattern at the *XKR4* locus among populations. An enrichment of homozygous variant genotypes (green) was observed across SNPs around *XKR4* region in population YAK, contrasting with the predominant heterozygous variants (yellow) in the NK+HK populations, suggesting this gene may be under selection in population YAK (Fig. 4F). Similarly, this gene was also likely to be under selection in population TS as suggested by Fig. 4G.

Genomic selection signatures in domesticated sika deer

Sika deer have experienced strong artificial selection during the process of domestication, as their antler size and reproductive traits have changed substantially compared with those of their wild ancestors. Therefore, in this study, we resequenced 81 wild and 71 domesticated sika deer representing 10 main geographic populations and two farms to explore the genetic mechanism of selection during domestication. Aforementioned five selective sweep methods were used to identify genomic regions under selection (Fig. 5A, Supplementary table S23-S27). Genes supported by more than three methods were considered as candidate genes associated with the domesticated process (Fig. 5B and Supplementary table S28, S29) [54, 55].

One of the most important traits under selection during the domestication of sika deer is antler weight, which increased by $30 \sim 60\%$ during the process of domestication [28]. To screen genes related to antler growth, we selected genes that overlapped among three different datasets: genes detected by selective sweep analysis in this study, genes identified by transcriptomic data from different antlers [56], and antler-specific genes from a previous study [57]. We identified 17 overlapping genes between the selective sweep and antler-specific genes, five between the selective sweep and transcriptomic datasets, and two (*NRP2* and *EDIL3*) among all three datasets (Fig. 5C and Supplementary table S30). The candidate genes *NRP2* and *EDIL3* were shown to be under selection in the domesticated population by the five selective sweep methods (Fig. 5D).

Discussion

The sika deer has historically been subdivided into 13 subspecies based on morphological characteristics [8-10]. As phylogenomic approaches can reveal discordances between molecular systematics and conventional morphological taxonomy [58, 59], we conducted phylogenomic investigations on sika deer. Based on the phylogenetic tree, PCA, and ADMIXTURE analysis of autosomal SNPs, together with mitochondrial and Y-chromosome evidence, we identified two lineages, the continental lineage and the Japanese lineage, with deep genetic divergence. Furthermore, Bayesian species delimitation decisively supported the species-level divergence between these lineages, with complete separation of posterior distributions (two-species model: PP=1.0 vs single-species model: PP = 0). The divergence time between the two lineages was estimated to be approximately 0.60 Ma, which coincided with the occurrence of the Naynayxungla Glaciation (0.72-0.50 Ma; MIS 14-18), the most extensive of the four major glaciations during Quaternary climatic history [39]. This glacial interval has been proposed as a key driver of evolutionary dynamics in species distributed across the eastern Eurasian mainland [21, 60, 61].

The ancestral origins of sika deer in the Japanese archipelago were considered to derive from Eurasian continental populations that migrated via the land bridges repeatedly formed between Japan and Eurasian continent during the middle and late Pleistocene [12, 62]. Although previous paleontological investigation had posited a potential northern China origin based on fossil record [63], the precise continental progenitor of the Japanese lineage remains elusive due to their deep divergence time (~0.60 Ma). Our phylogenetic analysis demonstrated that populations from southern China and Taiwan (TW, ZJ and SC) exhibited a closer genetic relationship to the Japanese lineage compared to populations from northern China and Russia (RU1, RU2 and NEC) (Fig. 1B). In addition, ADMIXTURE results at K=8 revealed that the genetic composition of the TW and ZJ populations were partially shared with populations from Japan (HK, NK, SM and YAK) (Fig. 1C). This admixed genetic structure likely reflected secondary contact events in late Pleistocene, as revealed by our demographic modeling demonstrating post-divergence gene flow from population ZJ to the Japanese lineages (Fig. 3C). These findings suggested complex evolutionary dynamics involving both early divergence and subsequent limited genetic exchange between the continental and Japanese lineages, complicating the ancestral origins inference of the Japanese lineage through contemporary genomic data alone.

Prior phylogeographic studies suggest that following their initial migration to Japan, sika deer ancestral populations likely inhabited in Honshu-Shikoku-Kyushu region [62, 63]. Nevertheless, some uncertainties remain regarding their demographic history, particularly the marked temporal discrepancy between mitochondrial and nuclear genomic data. Our whole mitochondrial genome analysis was consistent with previous D-loop and cytochrome b findings [12, 13], revealing two deeply divergent clades (0.53 Ma) (Supplementary Figure S7A). By contrast, the divergence times among populations in Japan estimated by genomic data were less than 0.10 Ma (Supplementary Figure S3 and S4). This inconsistency can be explained by a three-stage scenario of "divergence-mixing-isolation".

Stage I, Divergence. Our mitochondrial data supported the divergence of the two clades in Japan, yet it remains unclear whether this divergence resulted from dual colonization events from the continent [12] or occurred post-colonization in Japan [64]. Notably, while a previous study detected no differentiation in Y-chromosomal markers among populations in Japan [16], our analysis of SNPs across the entire Y-chromosome clearly showed that populations in Japan were highly structured, with individuals from the same population forming distinct clusters (Supplementary Figure S7B). This discrepancy can be attributed to our more comprehensive dataset, which encompasses SNPs from the entire Y-chromosome, thus providing higher resolution. The differences in topological structures observed among populations based on mitochondrial and Y-chromosomal data were likely due to female philopatry and male-biased dispersal patterns in sika deer [65-67]. The substantial genetic divergence between the northern and southern sika deer clades in Japan renders the hypothesis of simultaneous continental divergence followed by concurrent migration biologically unlikely. While this dual-origin scenario could account for the existence of two distinct lineages, it fails to provide a parsimonious explanation for their pronounced latitudinal segregation. The invocation of female philopatry to reconcile this geographic pattern with the dual migration hypothesis is further weakened by evidence of Y chromosome-based genetic differentiation among populations (Supplementary Figures S7B). This genetic structure suggests that even male-mediated gene flow has been insufficient to homogenize these

lineages, thereby challenging any simplistic model predicated solely on continental divergence and synchronous colonization.

Stage II, Mixing. Secondary contact and subsequent genetic admixture between the two clades occurred in Honshu-Shikoku-Kyushu [12, 62, 63]. The relatively recent divergence times among populations estimated by genomic data was likely attributed to this historical admixture event. In contrast, mitochondrial DNA-based estimates between the two clades remained unaffected by post-contact gene flow, as maternal inheritance of mitochondria allows the persistence of deeply divergent matrilineal lineages even when nuclear genomic differentiation becomes homogenized through introgression [68]. Moreover, TreeMix and *D*-statistics analyses further reveled asymmetry gene flow patterns, with significant migration detected from TS to NK, TS to HK, and NK to SM (Fig. 3D and 3E).

Stage III, Isolation. The sika deer in Japan were isolated to different islands leading to the formation of the current genetic structure (Supplementary Figure S3 and S4).

In addition, the NEC population is a domesticated group living in farms in northeast China. Our analysis of mitochondrial haplotypes revealed that this population had a relatively distant genetic relationship with other populations in the continental lineage (Fig. 1E). By contrast, biparental genomic data suggested that this population was more closely related to RU1 and RU2 than to other populations within the continental lineage (Fig. 1B). This discrepancy may be attributed to admixture of this population with RU1 and RU2, thereby reducing their genetic differences in nuclear markers.

Previous studies on patterns in ungulate body size variations emphasized the effect of a particular environmental factor, such as island area or precipitation, which reflects the availability of food resources [24]. This macroevolutionary phenomenon, termed the "island rule", is pervasive across terrestrial vertebrates [26]. There is geographical variation in the body size of sika deer [8]. Substantial body weight variations among sika deer populations from the Japanese archipelago were reported: 50 kg in Kyushu (KY), 55 kg in Tanegashima (TN), 35 kg in Yakushima Island (YAK), and 30 kg in Kerama (KR) for adult males [8]. Moreover, the body size differences among different populations in the Japanese archipelago were verified in both adult male and female, respectively [41]. Hence, the conclusion that the body size of sika deer populations in Japan adheres to the island rule was neither influenced by disparate life histories nor attributable to sexual dimorphism [40]. By selective sweep analyses, we identified 12 genes shared between two small body size populations (YAK and TS) were under selection. Through a further literature search, we found four

genes (XKR4, NPAS3, CTNNA3, and CNTNAP5) might be related to body size regulation as suggested by published papers. Among these genes, phospholipid scramblase XKR4 encoded a member of XK-related (XKR) membrane protein family, catalyzing the rapid, nonspecific, and bi-directional translocation of phospholipids between the two leaflets of the plasma membrane [44]. It was reported to be associated with feed intake and growth phenotypes in cattle [45], carcass weight (CWT) and eye muscle area (EMA) in Korean Hanwoo cattle [46], growth and fatness traits in Sujiang pigs [47], and body size traits in Pekin ducks [48], indicating that this gene played a crucial role in regulating body size in various animals. NPAS3 was reported to play a critical role for lung development in mice [49]. In addition, a genome-wide association study suggested that it was candidate gene associated with growth and fatness traits in large white pigs [50]. CTNNA3 was reported to be involved in growth traits of red deer [51] and Qinchuan cattle [52]. CNTNAP5 was speculated to be a candidate gene associated with growth and carcass traits including body weight and body fat deposition in cattle [53].

These genes may contribute to the difference in body size between sika deer among the Japanese archipelago. To our knowledge, only one study has explored genes under selection that are related to body size differences between island and continental populations [27]. In that work, selective signatures in the fatty acid desaturase (FADs) gene cluster were inferred to be associated with stature size differences between an extant pygmy population on Flores Island and modern humans. In addition, other genes (PLAG1, HMGA2, LCORL, CCND2, TNN2, and TCP11) were found to regulate body size in cattle [69], suggesting that the genetic architecture of body size in large mammals is affected by many polymorphisms of small effect, as observed in cattle and humans [69, 70]. Therefore, more studies are needed to investigate whether the genes regulating body size related to the "island rule" in humans and other mammals are conserved.

Understanding the genetic changes underlying phenotypic variation in sika deer during domestic processes could facilitate our efforts toward further improvement. In this study, we detected genomic regions harboring genes associated with antler size. In particular, a strong selective signature located around two gene regions (*NRP2* and *EDIL3*) implied intense artificial selection during domestication, suggesting that they may be involved in antler growth. Neuropilin-2 (*NRP2*) is a member of the neuropilin protein family. It was identified as a potential biomarker for chondrocyte programmed cell death in a recent study [71]. Epidermal growth factor (EGF)-like repeats and discoidin I-like domains-containing protein 3 (*EDIL3*) was reported to maintain cartilage extracellular matrix and protect chondrocyte apoptosis [72, 73]. Our data provide a valuable genomic resource for facilitating future genomics-informed breeding and genetic improvement of domestic sika deer.

Conclusions

Our analyses of whole genomes, Y chromosomes, and mitochondrial genomes revealed substantial genetic differentiation between the continental and Japanese lineages of sika deer, indicating that they belong to two distinct phylogenetic species. The divergence time between the two species was estimated to be approximately 0.60 Ma, which coincided with the Navnavxungla Glaciation (0.72–0.50 Ma, MIS 14–18). Sika deer in Japan were inferred to have experienced a "divergence-mixingisolation" evolutionary scenario. In addition, we also identified four candidate genes (XKR4, NPAS3, CTNNA3, and CNTNAP5) that may be involved in body size regulation. Furthermore, we also detected two candidate genes (NRP2 and EDIL3) that is putatively associated with an important economic trait (antler weight) under selection during the process of domestication. The genomic data generated in this study will serve as a valuable resource for genomics-assisted breeding to improve economic traits in domesticated sika deer.

Limitation of the study

Our work has several limitations. First, sika deer are listed as an endangered animal in China's Red Book of Endangered Animals (National Forestry and Grassland Administration, National Park Administration, 2021). Therefore, the number of sika deer that could be collected and used for this study was limited. Second, some population genetic parameters may be affected by sample size, for example, genetic diversity, linkage disequilibrium and even selective sweep statistics. Third, body size has been considered one of the most important traits of an organism or a species, as it covaries with a wide range of morphological, physiological, and ecological features. Our work is only a preliminary analysis of this complex trait.

Methods

Sample collection, DNA extraction, and genome sequencing

We collected blood, ear, antler or skin samples from 81 wild and 71 domesticated sika deer representing 10 main geographic populations and two farms, respectively. During the process of sampling, the deer were anesthetized by intramuscular injection of 5% xylazine with a dose of 0.5 mg/kg body weight, which was performed in accordance with the ARRIVE guidelines (https://arriv eguidelines.org). The deer were not treated harshly during sampling and were released after sampling. Of the 81 wild individuals, 16 were from a nature reserve in the Far East of Russia (RU1, including 10 individuals collected in Kazakhstan introduced from this nature reserve in the 1930s), 6 were from the wild region in the Far East of Russia (RU2), 8 were from the Tiebu Nature Reserve in Sichuan Province (SC), 8 were from the Qingliangfeng National Nature Reserve in Zhejiang Province (ZJ), 2 were from the Liugongdao National Forest Park in Shandong Province (introduced from Taiwan on April 16, 2011, TW), 9 were from Hokkaido (HK), 8 were from Nikko in Honshu (NK), 2 were from Shimane in Honshu (SM), 14 were from Tsushima Island (TS), and 8 were from Yakushima Island (YAK). The last five sampling sites are located in Japan. In addition, a total of 71 domesticated sika deer from farms in Jilin and Heilongjiang Provinces were also included in this study (Fig. 1a and Supplementary table S1).

Russian and Kazakhstani samples were provided by Professor Nurlan Toktarov. Chinese samples were collected by Huamiao Liu, Yimeng Dong, Yan Ju, Yang Li, Weilin Su, Ranran Zhang, Shiwu Dong, Hongliang Wang, Yongna Zhou, Yanmin Zhu, Lei Wang, Zhengyi Zhang, Pei Zhao, Shuyan Zhang, Rui Guo, E A, Yuwen Zhang, and Xin Liu. DNA of the Japanese samples were provided by Professor Hidetoshi B. Tamate.

Genomic DNA of all samples sequenced in this study was extracted using the DNeasy Blood and Tissue Kit (QiaGen, Shanghai, China), including RNase A treatment. DNA quality and quantity were evaluated using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and electrophoresis on a 0.8% agarose gel, respectively. Individual DNA libraries with an insert size of 350 bp were constructed with at least 1 μ g of genomic DNA according to the manufacturer's protocol and then sequenced using the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA).

Sequencing read filtering and SNP variant calling

In total, we obtained ~13.52 terabases (Tb) of raw sequencing data (~88.95 Gb per sample). To avoid reads with artificial bias (i.e., low-quality paired reads, which primarily result from base-calling duplicates and adapter contamination), we used fastp v0.23.4 [74] Wheeler aligner (BWA-mem) software version 0.7.84] to remove the following types of reads: (1) reads with \geq 10% unidentified nucleotides (N); (2) reads with > 10 nt aligned to the adapter, with \leq 10% mismatches allowed; and (3) reads with > 50% bases having a Phred quality <5.

Consequently, we retained 13.22 Tb (~ 86.94 Gb per sample) of high-quality genomic data (Supplementary table S2).

The remaining high-quality paired-end reads were aligned to the chromosome-level reference genome of sika deer assembled from a domesticated individual [4] and the Y chromosome of GCA_040085125.1 in NCBI RefSeq database (https://www.ncbi.nlm.nih.gov/datasets/ genome/GCA_040085125.1/) using Burrows-Wheeler aligner (BWA-mem) software version 0.7.8 [75] with the command "mem -t 4 -M -R". We then converted the mapped reads into BAM files and sorted the files using SAMtools v1.9 [76] with the parameter setting "-bS -t". PCR duplicates were subsequently removed using "MarkDuplicates" in GATK version 4.1.2.0 [77]. We calculated the whole-genome sequencing coverage and depth of each sample using SAMtools. Only samples with miss < 10% and depth > 5 × were kept to call SNP variations.

SNPs and InDels were called from the BAM files using the *HaplotypeCaller* module in GATK version 4.1.2.0 with the "-ERC GVCF" option. Raw GVCFs with the samples called individually were merged using CombineGVCFs and called for SNPs using GenotypeGVCFs. We then selected the candidate SNPs and created the selected SNP dataset using the *SelectVariants* module in GATK. To obtain reliable candidate SNPs and InDels, we implemented "VariantFiltering" in GATK according to the following parameter settings: "QUAL < 30.0 | QD < 2.0 | MQ < 40.0 | FS > 60.0 | SOR > 3.0 | MQRankSum < -12.5 | ReadPosRankSum < -8.0" for SNPs and "QD < 2.0 | QUAL < 30.0 | FS > 200.0 | ReadPosRankSum < -20.0" for InDels.

After filtering, all the identified SNPs and InDels were further annotated using the ANNOVAR software tool v2022-12–28 [78] and subsequently divided into the following groups on the basis of the sika deer genome annotation information: variants occurring in intergenic regions, within 1 kb upstream of the translation start and downstream of the translation stop sites, in coding sequences, and in introns.

Structural variant and copy number variant calling

Only samples with depth>15×were kept to identify structural variants (SVs). SVs, including INS, DEL, ITX, CTX, and INV, were detected by LUMPY v.0.2.13 [79] and BreakDancer v1.4.5 [80] with the following criteria: 50 bp \leq SV \leq 1 Mb. CNV calling was performed using CNVnator 0.3.3 [81].

Mitochondrial genome assembly and haplotype analysis

Clean reads were mapped to a complete mitochondrial genome (GenBank: GU457433.1) to generate a SAM file for each individual. After filtering potential PCR duplications by SAMtools software, the resulting BAM file was used as input for MITObim version 1.9.1 [82] to assemble the mitochondrial genome of each individual with the parameters "-start 0 -end 30".

The mitochondrial genome alignment was refined using TrimAl v1.4.1 [83] with automated parameter optimization (-automated1 mode), which removed ambiguously aligned regions while preserving phylogenetically informative positions, resulting in a conserved alignment length of 15,980 bp. From this curated alignment, we identified 914 variable sites. Thirty-nine unique haplotypes were resolved in wild sika deer populations using DnaSP version 5.10.01 [84]. Bayesian phylogenetic reconstruction and divergence time estimation were performed in BEAST version 2.6.6 [85] using an uncorrelated lognormal relaxed clock, with a calibration node prior set at 1.4 Ma [95% Highest Posterior Density (HPD): 0.73-2.20 Ma] as the divergence time between Cervus elaphus and Cervus nippon [86]. Additionally, a maximum likelihood (ML)-based unrooted phylogeny was reconstructed in IQ-TREE v2.2.0 [87].

Y chromosome SNPs and haplotype analysis

Following aforementioned GATK variant calling procedures, we identified 543,807 Y-chromosomal SNPs across 129 male sika deer. Subsequent quality filtering with TrimAl (-automated1 mode) retained 211,326 SNPs, from which 52 distinct haplotypes were resolved in wild populations. A neighbor-joining tree was reconstructed in MEGA11 [88] under the Kimura two-parameter (K2P) model [89], with nodal support assessed through 100 bootstrap replicates.

After removing heterozygous sites and sites with > 5% missing genotypes, we obtained 1,248 high-confidence SNPs that delineated 35 Y-chromosomal haplotypes. A median-joining network was constructed using NET-WORK v10.1 (https://www.fluxus-engineering.com/) [90].

Population genetic analysis

To reduce the required computing resources, we implemented LD-based pruning for the genotype data using PLINK software version 1.9 [91] with the option "-indeppairwise 50 10 0.2" for the autosomal SNP dataset. To clarify phylogenetic relationships from a genome-wide perspective, we constructed an NJ tree according to the p distance using the software TreeBeST v1.9.2 [92] with 1000 bootstrap replications. Furthermore, population genetic structure was examined via an expectation

maximization algorithm, as implemented in the program ADMIXTURE v1.23 [93]. The number of assumed genetic clusters K ranged from 2 to 9, with 10,000 iterations for each run. CV error minimization was used to determine the optimal K value [94]. In addition, principal component analysis (PCA) was performed using PLINK software [91].

To infer the species tree and estimate the divergence times among the major clades, we used the SNAPP package version 1.5.2 [33] implemented in BEAST version 2.6.6 [85] to perform phylogenetic analysis. Due to the high computational demand of SNAPP, we performed this analysis using only four individuals for each population and a reduced set of 5,000 SNPs randomly selected from all biallelic SNPs with no missing genotypes. The input files for SNAPP were prepared with the script snapp_prep.rb [95], implementing a strict-clock model and a pure-birth tree model. We used the divergence time of Cervus elaphus and Cervus nippon of 1.4 Ma (95% HPD: 0.73-2.20 Ma) [86] as the calibration point. We carried out three independent SNAPP analyses, each with a Markov chain Monte Carlo (MCMC) chain length of 1,500,000 and parameter values recorded every 250 generations, resulting in ESS values that were all greater than 400. The posterior tree distribution was again summarized as a maximum clade credibility (MCC) consensus tree. To verify the repeatability of SNAPP analysis, we repeated this analysis by re-selecting the reduced set of 5,000 SNPs randomly for another 9 times, and the topological structures of species trees were integrated into one figure (Supplementary Figure S4).

Species delimitation

We used BPP v. 4.3.0 [35] for joint species delimitation and species tree inference with unguided species delimitation [96]. Due to the relatively high computational burden of Bayesian methods such as BPP, we performed inference on reduced datasets consisting of a subset of 1000 of the most variable loci, which were selected by Geneious v9.1.7 [97] based on the percentage of variable sites. Three independent MCMC runs were implemented with a burn-in of 10,000, a sample frequency of 2, and 100,000 total samples.

Genetic diversity

Genetic differentiation statistics (F_{ST}), observed heterozygosity (Het), and nucleotide diversity ($\theta\pi$) were calculated using VCFtools v0.1.14 [98], with sliding windows of 40 kb and 20-kb overlap between adjacent windows. To detect the genomic signatures of inbreeding in different populations of sika deer, genome-wide runs of homozygosity (ROH) were identified for each individual using the tool in PLINK software version 1.9 [91] with

the parameter setting "-threshold 0.05 -snp 65 -kb 100 -missing 3 -gap 5,000 -density 5,000".

We estimated the genome-wide LD decay of each population using PopLDdecay version 3.40 [99], setting the maximum distance between two SNPs to 500 kb, MAF \geq 0.05, and missing ratio \leq 10%. In addition, PLINK version 1.9 [91] was used to calculate the LD coefficient (r²) between pairs of high-quality SNPs, with the parameters set as "-ld-window-r2 0 -ld-window 99,999 -ld-window-kb 1000". The average r² value was calculated for pairs of markers in a 10-kb window, and values were averaged across the whole genome.

Demographic history and divergence time analyses

To trace the demographic history of each sika deer population, we employed the PSMC model [37] to reconstruct changes in historical effective population size (N_e) over time using heterozygous sites. Analyses for each population were based on the two to three individuals with the highest mean sequencing depth. The PSMC analysis was run with the parameters -N25 -t15 -r5 -p "4+25 *2+4+6", together with 100 bootstrap replicates. Referring to Tarim red deer (*C. e. yarkandensis*) and other species in Cervidae, we adopted a mutation rate of 2.2×10^{-9} per nucleotide site per generation [100, 101] and a generation time of 5 years to scale the results.

The MSMC method [38] was used to estimate the relative cross-coalescence rate (RCCR) between pairs of high-coverage individuals. The time at which the RCCR decreased below the 0.5 threshold was taken as a point estimate of the divergence time. For the range of divergence times, we used the first time point at which the cross-coalescence rate was at or below 0.25 and 0.75. All parameter estimates were scaled using a mutation rate of 2.2×10^{-9} per nucleotide site per generation [100, 101] and a generation time of 5 years to scale the results.

Demographic histories and migration rates among RU1, ZJ and Japan (JPN) were estimated by using FastSIMCOAL2 v.2.6 [102], using a mutation rate of 2.2×10^{-9} per nucleotide site per generation [100, 101] and a generation time of 5 years. We first inferred the best parameters and the likelihoods for each of the demographic models: no gene flow, early gene flow and recent gene flow, since the divergence between the continental populations and JPN by running 100 independent iterations with 300,000 coalescent simulations and 60 optimization cycles. The demographic model with the highest likelihood was selected to run parameter estimation with block-bootstrap of 100 replicates.

To detect potential gene flow among populations in Japan, we implemented two distinct statistical analyses: TreeMix v1.13 [103] and *D*-statistics [104]. For the TreeMix analysis, we conducted maximum likelihood trees

using blocks of 1,000 SNPs and RU1 as the root, setting the number of tested migration events from 1 to 5. Subsequently, the optimal number of migration edges was determined by OptM [105]. Additionally, we calculated *D*-statistics using the program AdmixTools [104]. The four populations were set to be ((A, B), (X, Y)), where Y represented the outgroup RU1, and A, B, X were selected population combinations according to the phylogenetic tree constructed by autosomal SNPs (Supplementary Figure S3 and S4). The statistical significance of the *D* value was assessed using a two-tailed *Z* test, with |Z-score|>3 to be significant [106].

Genomic signatures of selection for body size and domestication

To identify genome-wide selective sweeps associated with body size, we used two population pairs: YAK (n=8)vs NK (n=8) + SM (n=2), and TS (n=14) vs NK (n=8). In addition, the wild population (n=81) and domesticated population (n=71) were compared to detect selective signatures related to the domestication process. We used five selective sweep methods to identify genomic regions under selection, including the genetic differentiation statistic F_{ST} , the genetic diversity $\theta \pi$, Tajima's D, XP-CLR, and XP-EHH. F_{ST} , $\theta\pi$ and Tajima's D were calculated using VCFtools v0.1.14 [98], with sliding windows of 40 kb and 20-kb overlap between adjacent windows. We also scanned for signals of selection across the genome using XP-EHH scores [107] as implemented in the Python package "ALLEL" with a sliding window of 40 kb in 20 kb step size. The XP-CLR statistic was calculated using the XP-CLR v1.1.2 [108]. According to the simple majority vote strategy [54, 55], genes supported by more than three methods can be considered as candidate genes related to the phenotype. To identify candidate genes related to body size, we adopted a stricter criterion (supported by all five methods) to decrease the false-positive rates. To identify candidate genes related to antler weight, we adopted a common criterion (supported by more than three methods), since we already used two different datasets (genes identified by transcriptomic data from different antlers [56], and antler-specific genes from a previous study [57]).

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s12864-025-11541-w.

Below is the link to the electronic supplementary material. Supplementary Material 1: Supplementary Table S1 Sample information. Supplementary Table S2 Summary of whole-genome sequencing data of 152 individuals. Supplementary Table S3 Summary of categorized SNPs. Supplementary Table S4 Summary of InDels. Supplementary Table S5 Summary of SVs detected by LUMPY v.0.2.13. Supplementary Table S6 Summary of SVs detected in each individual by BreakDancer v1.4.5. Supplementary Table S7 Summary of CNVs.Supplementary Table S8 Candidate genes related to body size in YAK population identified by FST. Supplementary Table S9 Candidate genes related to body size in YAK population identified by $\theta \pi$. Supplementary Table S10 Candidate genes related to body size in YAK population identified by Tajima's D. Supplementary Table S11 Candidate genes related to body size in YAK population identified by XP-CLR. Supplementary Table S12 Candidate genes related to body size in YAK population identified by XP-EHH. Supplementary Table S13 Candidate genes related to body size in TS population identified by FST. Supplementary Table S14 Candidate genes related to body size in TS population identified by $\theta\pi$. Supplementary Table S15 Candidate genes related to body size in TS population identified by Tajima's D. Supplementary Table S16 Candidate genes related to body size in TS population identified by XP-CLR. Supplementary Table S17 Candidate genes related to body size in TS population identified by XP-EHH. Supplementary Table S18 131 candidate genes related to body size in YAK population Supplementary Table S19 Variant types of SNPs around 131 candidate genes related to body size in YAK population. Supplementary Table S20 117 candidate genes related to body size in TS population. Supplementary Table S21 Variant types of SNPs around 117 candidate genes related to body size in TS population. Supplementary Table S22 12 candidate genes related to body size shared between YAK and TS population. Supplementary Table S23 Candidate genes related to the domesticated process identified by FST.Supplementary Table S24 Candidate genes related to the domesticated process identified by $\theta \pi$. Supplementary Table S25 Candidate genes related to the domesticated process identified by Tajima's D. Supplementary Table S26 Candidate genes related to the domesticated process identified by XP-CLR. Sup plementary Table S27 Candidate genes related to the domesticated process identified by XP-EHH.Supplementary Table S28 1298 candidate genes under selection during the domesticated process. Supplementary Table S29 Variant types of SNPs around 1298 candidate genes under selection during the domesticated process. Supplementary Table S30 Two candidate genes under selection may be associated with antler size

Supplementary Material 2: Supplementary Figure S1 Visualization of genomic SNP density distributed across 33 chromosomes with a window size of 20 kb. Supplementary Figure S2 Circos plot illustrating genomic variants distributed across 33 chromosomes with a window size of 100 kb. a, b. Number of SNPs in continental (a) and Japanese (b) populations; c, d. Number of SVs in continental (c) and Japanese (d) populations; e, f. Number of CNVs in continental (e) and Japanese (f) populations. Supplementary Figure S3 Species tree reconstructed with SNAPP. This analysis was performed using four individuals for each population and a reduced set of 5.000 SNPs randomly selected from all biallelic SNPs without missing genotypes. We used the divergence time of Cervus elaphus and Cervus nippon of 1.4 Ma [95% Highest Posterior Density (HPD): 0.73-2.20 Ma] as the calibration point. Node annotations indicate mean divergence estimates with corresponding 95% HPD intervals for the major clades. Note that the TW and SM populations were excluded since only two samples were collected at each site. Supplementary Figure S4 Topological structures of species trees reconstructed with SNAPP. This analysis was performed by ten times independently and integrated into one figure. Supplementary Figure S5 Population structure of all 152 sika deer revealed by ADMIX-TURE analysis across K=2-9. Supplementary Figure S6 Determination of optimal ancestral clusters through ADMIXTURE cross-validation. We inferred genetic structure using ADMIXTURE v1.23 across a range of K=2-9, identifying K=3 as the optimal cluster number based on the lowest cross-validation error (CV error=3.13). Supplementary Figure S7 Phylogenetic trees constructed by mitochondrial and Y-chromosomal data, respectively. A. Phylogenetic tree of mitochondrial DNA haplotypes of wild sika deer and divergence times estimated by BEAST version 2.6.6. We used the divergence time of Cervus elaphus and Cervus nippon of 1.4 Ma [95% Highest Posterior Density (HPD): 0.73-2.20 Ma] as the calibration point. Node annotations indicate mean divergence estimates with corresponding 95% HPD intervals for the major clades. B. NJ phylogenetic tree of Y-chromosomal haplotypes of

wild sika deer. Bootstrap analysis was carried out with 100 replicates, and values higher than 50 were shown above the branches. Cervus elaphus was used as outgroup. The same populations in the Japanese lineage between the two trees were denoted with dashed lines. Supplementary Figure S8 Relationship between observed heterozygosity and area sizes of different nature reserves and farms. Supplementary Figure S9 Output generated by OptM analysis. A. Composite likelihood L(m) with mean \pm standard deviation (SD) (left axis; black circles), and proportion of variance explained (right axis; red dots) across 10 iterations for migration edges m=1-5. B. Second-order rate of change (Δm) across m values, with the maximum Δ m indicating optimal migration edges. Supplementary Figure S10 Picture displays the body size difference of different subspecies of sika deer in Japan: C. n. yesoensis distributed on Hokkaido Island, C. n. centralis on Honshu mainland and Tsushima Island, and C. n. yakushimae on Yakushima and Kuchinoerabu Islands (photographic credit: Professor Hidetoshi B. Tamate). Supplementary Figure S11 Candidate gene numbers identified by different methods of selective sweep analysis. Genes identified by all five methods were considered as candidate genes in population pairs NK+SM vs YAK (A) and NK vs TS (B).

Acknowledgements

We would like to thank Professor Nurlan Toktarov for providing samples.

Authors' contributions

X. X. conceived and supervised the project. H. L., Y. D., Y. J., Y. L., W. S., R. Z., S. D., H. W., Y. Zhou, Y. Zhu, L. W., Z. Z., P. Z., S. Z., R. G., E A, Y. Z. and X. L. collected the samples. H.L., T. W. and Y. D. carried out the experiments. B. Z., T. W. and Q. L. performed the bioinformatics analyses. H.B. T., D. M., Q. L. and X.X. discussed the analysis results. H. L., T. W., D. M. and X. X. wrote the manuscript. All authors have read and approved the final manuscript.

Funding

This study was supported by funds from the Agricultural Science and Technology Innovation Program of China (CAAS-ASTIP-201X-ISAPS) and the Special Economic Animals Sharing Platform in China (TZDW2020).

Data availability

The raw whole-genome sequencing data reported in this study have been deposited to the National Genomics Data Center (NGDC, https://ngdc.cncb. ac.cn/), under accession number CRA018955 (https://ngdc.cncb.ac.cn/gsa/browse/CRA018955).

Declarations

Ethics approval and consent to participate

Informed consent was obtained from all participating institutions and individual collaborators. Handling and sampling of the 152 sika deer were performed in accordance with the ARRIVE guidelines (https://arriveguidelines.org) and the guidelines for the care and use of experimental animals established by the Ministry of Agriculture and Rural Affairs of China, and all experiments were approved by the Institutional Animal Care and Use Committee of the Institute of Special Economic Animal and Plant Sciences, Chinese Academy of Agricultural Sciences (Approval No. ISAPSAEC-2014–016), Changchun, China.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹ State Key Laboratory for Molecular Biology of Special Economic Animals, Institute of Special Animal and Plant Sciences, Chinese Academy of Agricultural Sciences, Changchun 130112, China. ²Novogene Bioinformatics Institute, Beijing 100083, China. ³Administration of Zhejiang Qingliangfeng National Nature Reserve, Hangzhou 310000, China. ⁴Sichuan Tiebu Sika Deer Nature Reserve, Aba 624000, China. ⁵Administrative Office of Liugong Island National Forest Park, Weihai 264200, China. ⁶Northeast Forestry University, Harbin 150006, China. ⁷Faculty of Science, Yamagata University, Yamagata 9908560, Japan. ⁸Glbizzia Bioinformatics Institute, Beijing 102208, China.

Received: 27 August 2024 Accepted: 28 March 2025 Published online: 11 April 2025

References

- Schlebusch CM, Sjödin P, Breton G, Günther T, Naidoo T, Hollfelder N, et al. Khoe-San Genomes Reveal Unique Variation and Confirm the Deepest Population Divergence in Homo sapiens. Mol Biol Evol. 2020;37(10):2944–54.
- 2. Bergström A, Stanton DWG, Taron UH, Frantz L, Sinding MS, Ersmark E, et al. Grey wolf genomic history reveals a dual ancestry of dogs. Nature. 2022;607(7918):313–20.
- Harris RB. 2015. Cervus nippon. The IUCN Red List of Threatened Species 2015:e.T41788A22155877. https://doi.org/10.2305/IUCN.UK.2015-2. RLTS.T41788A22155877.en
- Xing X, Ai C, Wang T, Li Y, Liu H, Hu P, et al. The first high-quality reference genome of sika deer provides insights into high-tannin adaptation. Genomics Proteomics Bioinformatics. 2023;21(1):203–15.
- Nagata J, Masuda R, Kaji K, Kaneko M, Yoshida MC. Genetic variation and population structure of the Japanese sika deer (Cervus nippon) in Hokkaido Island, based on mitochondrial D-loop sequences. Mol Ecol. 1998;7(7):871–7.
- van Doormaal N, Ohashi H, Koike S, Kaji K. Influence of human activities on the activity patterns of Japanese sika deer (Cervus nippon) and wild boar (Sus scrofa) in Central Japan. Eur J Wildl Res. 2015;61:517–27.
- Liu H, Ju Y, Tamate H, Wang T, Xing X. Phylogeography of sika deer (Cervus nippon) inferred from mitochondrial cytochrome-b gene and microsatellite DNA. Gene. 2021;772:145375.
- Ohtaishi N. Preliminary memorandum of classification, distribution and geographic variation on Sika deer. Mammal Science. 1986;53:13–7 (In Japanese.).
- Ohtaishi N, Gao Y. A review of the distribution of all species of deer (Tragulidae, Moschidae and Cervidae) in China. Mammal Rev. 1990;20(2–3):125–44.
- 10. McCullough DR, Takatsuki S, Kaji K, editors. Sika Deer: Biology and Management of Native and Introduced Populations. Tokyo: Springer; 2009.
- 11. Tamate HB, Tsuchiya T. Mitochondrial DNA polymorphism in subspecies of the Japanese Sika deer. Cervus nippon J Hered. 1995;86(3):211–5.
- Nagata J, Masuda R, Tamate HB, Hamasaki Si, Ochiai K, Asada M, et al. Two genetically distinct lineages of the sika deer Cervus nippon, in Japanese islands: comparison of mitochondrial D-loop region sequences. Mol Phylogenet Evol. 1999;13(3):511–9.
- Tamate HB, Tatsuzawa S, Suda K, Izawa M, Doi T, Sunagawa K, Miyahira F, Tado H. Mitochondrial DNA Variations in Local Populations of the Japanese Sika Deer, Cervus nippon. J Mammalogy. 1998;79:1396–403.
- Yamada M, Hosoi E, Nagata J, Tamate HB, Tado H. Phylogenetic relationship of the southern Japan lineages of the sika deer (Cervus nippon) in Shikoku and Kyushu Islands. Japan Mamm Study. 2007;32:121–7.
- Takiguchi H, Tanaka K, Ono K, Hoshi A, Minami M, Yamauchi K, Takatsuki S. Genetic variation and population structure of the Japanese sika deer (Cervus nippon) in the Tohoku District based on mitochondrial D-loop sequences. Zoolog Sci. 2012;29(7):433–6.
- Tanaka K, Hoshi A, Nojima R, Suzuki K, Takiguchi H, Takatsuki S, et al. Genetic Variation in Y-Chromosome Genes of Sika Deer (Cervus nippon) in Japan. Zoolog Sci. 2020;37(5):411–6.
- 17. Zhang DX, Hewitt GM. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Mol Ecol. 2003;12(3):563–84.
- van Riemsdijk I, Arntzen JW, Babik W, Bogaerts S, Franzen M, Kalaentzis K, et al. Next-generation phylogeography of the banded newts (Ommatotriton): A phylogenetic hypothesis for three ancient species with geographically restricted interspecific gene flow and deep intraspecific genetic structure. Mol Phylogenet Evol. 2022;167:107361.
- Georges A, Gruber B, Pauly GB, White D, Adams M, Young MJ, et al. Genomewide SNP markers breathe new life into phylogeography and species delimitation for the problematic short-necked turtles (Chelidae: Emydura) of eastern Australia. Mol Ecol. 2018;27(24):5195–213.

- Pedraza-Marrón CDR, Silva R, Deeds J, Van Belleghem SM, Mastretta-Yanes A, Domínguez-Domínguez O, et al. Genomics overrules mitochondrial DNA, siding with morphology on a controversial case of species delimitation. Proc Biol Sci. 1900;2019(286):20182924.
- Yang L, Wei F, Zhan X, Fan H, Zhao P, Huang G, et al. Evolutionary Conservation Genomics Reveals Recent Speciation and Local Adaptation in Threatened Takins. Mol Biol Evol. 2022;39(6):111.
- 22. Reznick DN, Ricklefs RE. Darwin's bridge between microevolution and macroevolution. Nature. 2009;457(7231):837–42.
- Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, et al. Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. Mol Biol Evol. 2016;33(10):2576–92.
- 24. Lomolino MV. Body size evolution in insular vertebrates: generality of the island rule. J Biogeogr. 2005;32:1683–99.
- 25. Lokatis S, Jeschke JM. The island rule: an assessment of biases and research trends. J Biogeogr. 2018;45:289–303.
- Benítez-López A, Santini L, Gallego-Zamorano J, Milá B, Walkden P, Huijbregts MAJ, Tobias JA. The island rule explains consistent patterns of body size evolution in terrestrial vertebrates. Nat Ecol Evol. 2021;5(6):768–86.
- Tucci S, Vohr SH, McCoy RC, Vernot B, Robinson MR, Barbieri C, et al. Evolutionary history and adaptation of a human pygmy population of Flores Island. Indonesia Science. 2018;361(6401):511–6.
- Li H. Study on the performance of the velvet antler of Sika Deer. Zhongguo Xumu Zazhi. 2003;39:31–2.
- Zha E, Gao S, Pi Y, Li X, Wang Y, Yue X. Wound healing by a 3.2 kDa recombinant polypeptide from velvet antler of Cervus nippon Temminck. Biotechnol Lett. 2012;34(4):789–93.
- Wu F, Li H, Jin L, Li X, Ma Y, You J, Li S, Xu Y. Deer antler base as a traditional Chinese medicine: a review of its traditional uses, chemistry and pharmacology. J Ethnopharmacol. 2013;145(2):403–15.
- Raudsepp T, Finno CJ, Bellone RR, Petersen JL. Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era. Anim Genet. 2019;50(6):569–97.
- Lv FH, Cao YH, Liu GJ, Luo LY, Lu R, Liu MJ, et al. Whole-Genome Resequencing of Worldwide Wild and Domestic Sheep Elucidates Genetic Diversity, Introgression, and Agronomically Important Loci. Mol Biol Evol. 2022;39(2):msab353.
- Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol Biol Evol. 2012;29(8):1917–32.
- Matschiner M. Species Tree Inference with SNP Data. Methods Mol Biol. 2022;2512:23–44.
- Flouri T, Jiao X, Rannala B, Yang Z. Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent. Mol Biol Evol. 2018;35(10):2585–93.
- 36. De Queiroz K. Species concepts and species delimitation. Syst Biol. 2007;56(6):879–86.
- 37. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475(7357):493–6.
- Schiffels S, Wang K. MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. Methods Mol Biol. 2020;2090:147–66.
- Zheng B, Xu Q, Shen Y. The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: review and speculation. Quat Int. 2002;98:93–101.
- Takatsuki S, Hosoi E, Suzuki K. Age-Related Changes in the Body Size and Weight of Sika Deer (Cervus nippon): A Comparison between Northern and Southern Populations of Honshu, The Main Island of Japan. Mammal Study. 2024;49(4):299–307.
- Kubo MO, Takatsuki S. Geographical Body Size Clines in Sika Deer: Path Analysis to Discern Amongst Environmental Influences. Evol Biol. 2015;42:115–27.
- Weigand H, Leese F. Detecting signatures of positive selection in nonmodel species using genomic data. Zool J Linn Soc. 2018;184:528–83.
- Horscroft C, Ennis S, Pengelly RJ, Sluckin TJ, Collins A. Sequencing era methods for identifying signatures of selection in the genome. Brief Bioinform. 2019;20:1997–2008.
- 44. Maruoka M, Zhang P, Mori H, Imanishi E, Packwood DM, Harada H, Kosako H, Suzuki J. Caspase cleavage releases a nuclear protein

fragment that stimulates phospholipid scrambling at the plasma membrane. Mol Cell. 2021;81(7):1397-1410.e9.

- 45. Lindholm-Perry AK, Kuehn LA, Smith TP, Ferrell CL, Jenkins TG, Freetly HC, Snelling WM. A region on BTA14 that includes the positional candidate genes LYPLA1, XKR4 and TMEM68 is associated with feed intake and growth phenotypes in cattle. Anim Genet. 2012;43(2):216–9.
- 46. Bhuiyan MSA, Lim D, Park M, Lee S, Kim Y, Gondro C, Park B, Lee S. Functional Partitioning of Genomic Variance and Genome-Wide Association Study for Carcass Traits in Korean Hanwoo Cattle Using Imputed Sequence Level SNP Data. Front Genet. 2018;22(9):217.
- Xu P, Li D, Wu Z, Ni L, Liu J, Tang Y, Yu T, Ren J, Zhao X, Huang M. An imputation-based genome-wide association study for growth and fatness traits in Sujiang pigs. Animal. 2022;16(8):100591.
- Yu JZ, Zhou J, Yang FX, Hao JP, Hou ZC, Zhu F. Genome-Wide Association Analysis Identifies Important Haplotypes and Candidate Gene XKR4 for Body Size Traits in Pekin Ducks. Animals (Basel). 2024;14(16):2349.
- Zhou S, Degan S, Potts EN, Foster WM, Sunday ME. NPAS3 is a trachealess homolog critical for lung development and homeostasis. Proc Natl Acad Sci U S A. 2009;106(28):11691–6.
- Yang W, Wu J, Yu J, Zheng X, Kang H, Wang Z, Zhang S, Zhou L, Liu J. A genome-wide association study reveals additive and dominance effects on growth and fatness traits in large white pigs. Anim Genet. 2021;52(5):749–53.
- Pi T, Yi W, Hu M, Quan X, Tian L, Sun H, Yan S. Assessing genomic diversity and signatures of selection in Qingyuan Wapiti. Anim Genet. 2025;56(1):e13505.
- 52. Yu H, Yu S, Guo J, Cheng G, Mei C, Zan L. Genome-Wide Association Study Reveals Novel Loci Associated with Body Conformation Traits in Qinchuan Cattle. Animals (Basel). 2023;13(23):3628.
- 53. Vanvanhossou SFU, Scheper C, Dossa LH, Yin T, Brügemann K, König S. A multi-breed GWAS for morphometric traits in four Beninese indigenous cattle breeds reveals loci associated with conformation, carcass and adaptive traits. BMC Genomics. 2020;21(1):783.
- 54. Wall ME, Raghavan S, Cohn JD, Dunbar J. Genome majority vote improves gene predictions. PLoS Comput Biol. 2011;7(11):e1002284.
- Zhou G, Shen D, Zhang J, Su J, Tan S. Recognition of protein/gene names from text using an ensemble of classifiers. BMC Bioinformatics. 2005;6Suppl 1(Suppl 1):S7.
- Hu P, Wang T, Liu H, Xu J, Wang L, Zhao P, Xing X. Full-length transcriptome and microRNA sequencing reveal the specific gene-regulation network of velvet antler in sika deer with extremely different velvet antler weight. Mol Genet Genomics. 2019;294(2):431–43.
- Wang Y, Zhang C, Wang N, Li Z, Heller R, Liu R, et al. Genetic basis of ruminant headgear and rapid antler regeneration. Science. 2019;364(6446):eaav6335.
- Nitnaware KM, Raskar KB, Agarwal G, Chávez Montes RA, Chopra R, López-Arredondo DL, et al. Whole-genome characterization and comparative genomics of a novel freshwater cyanobacteria species: Pseudanabaena punensis. Mol Phylogenet Evol. 2021;164:107272.
- Rasplus JY, Rodriguez LJ, Sauné L, Peng YQ, Bain A, Kjellberg F, et al. Exploring systematic biases, rooting methods and morphological evidence to unravel the evolutionary history of the genus Ficus (Moraceae). Cladistics. 2021;37(4):402–22.
- Yang FS, Liu M, Guo X, Xu C, Jiang J, Mu W, et al. Signatures of Adaptation and Purifying Selection in Highland Populations of Dasiphora fruticosa. Mol Biol Evol. 2024;41(6):msae099.
- Zhang YZ, Zhu RW, Zhong DL, Zhang JQ. Nunataks or massif de refuge? A phylogeographic study of Rhodiola crenulata (Crassulaceae) on the world's highest sky islands. BMC Evol Biol. 2018;18(1):154.
- Sato JJ. A review of the processes of mammalian faunal assembly in Japan: insights from molecular Phylogenetics. In: Motokawa M, Kajihara H, editors. Species diversity of animals in Japan. Tokyo: Springer; 2017; p. 49–116.
- 63. Kawamura Y. Fossil Record of Sika Deer in Japan. In: McCullough, D.R., Takatsuki, S., Kaji, K. (eds) Sika Deer. Tokyo: Springer; 2009; p.11–25.
- Ohnishi N, Uno R, Ishibashi Y, Tamate HB, Oi T. The influence of climatic oscillations during the Quaternary Era on the genetic structure of Asian black bears in Japan. Heredity. 2009;102(6):579–89.
- 65. Agetsuma N, Agetsuma YY, Hino T. Autumn long-distance movements of male Japanese sika deer Cervus nippon yesoensis in western Hokkaido. Japan Eurasian J For Res. 2011;14:13–9.

- Takii A, Izumiyama S, Mochizuki T, Okumura T, Sato S. Seasonal migration of Sika deer in the Oku-Chichibu mountains, central Japan. Mamm Study. 2012;37:127–37.
- 67. Takii A, Izumiyama S, Taguchi M. Partial migration and effects of climate on migratory movements of sika deer in Kirigamine Highland, central Japan. Mamm Study. 2012;37:331–40.
- Lagcher E, Lensing K, Bosse M, Fischer K, Camacho G, McManus J, Tensen L. Red, gold, and green: comparative genomics of polymorphic leopards from South Africa. Evolution. 2025;79(3):442–56.
- Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat Genet. 2018;50(3):362–7.
- Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46(11):1173–86.
- Ren J, Li R, Meng C, Xu Y, Li C. Identification of BCL3 as a biomarker for chondrocyte programmed cell death in osteoarthritis. Int J Exp Pathol. 2025;106(1):e12522.
- Chen MF, Hu CC, Hsu YH, Lin YC, Chen KL, Ueng SWN, Chang Y. The role of EDIL3 in maintaining cartilage extracellular matrix and inhibiting osteoarthritis development. Bone Joint Res. 2023;12(12):734–46.
- 73. Wang Z, Boyko T, Tran MC, et al. DEL1 protects against chondrocyte apoptosis through integrin binding. J Surg Res. 2018;231:1–9.
- 74. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
- 75. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
- 79. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15(6):R84.
- Fan X, Abbott TE, Larson D, Chen K. BreakDancer: Identification of Genomic Structural Variation from Paired-End Read Mapping. Curr Protoc Bioinformatics. 2014;45:15.6.1–11.
- Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011;21(6):974–84.
- Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads–a baiting and iterative mapping approach. Nucleic Acids Res. 2013;41(13):e129.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.
- Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2.
- Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 25: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4):e1006650.
- Meiri M, Kosintsev P, Conroy K, Meiri S, Barnes I, Lister A. Subspecies dynamics in space and time: A study of the red deer complex using ancient and modern DNA and morphology. J Biogeogr. 2018;45(2):367–80.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1530–4.
- Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38(7):3022–7.

- 90. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16(1):37–48.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
- Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 2009;19(2):327–35.
- 93. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–64.
- 94. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011;12:246.
- Stange M, Sánchez-Villagra MR, Salzburger W, Matschiner M. Bayesian Divergence-Time Estimation with Genome-Wide Single-Nucleotide Polymorphism Data of Sea Catfishes (Ariidae) Supports Miocene Closure of the Panamanian Isthmus. Syst Biol. 2018;67(4):681–99.
- 96. Yang Z, Rannala B. Unguided species delimitation using DNA sequence data from multiple Loci. Mol Biol Evol. 2014;31(12):3125–35.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.
- 99. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019;35(10):1786–8.
- Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science. 2019;364(6446):eaav6202.
- 101. Ababaikeri B, Abduriyim S, Tohetahong Y, Mamat T, Ahmat A, Halik M. Whole-genome sequencing of Tarim red deer (Cervus elaphus yarkandensis) reveals demographic history and adaptations to an arid-desert environment. Front Zool. 2020;17:31.
- Excofffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics. 2021;37(24):4882–5.
- Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 2012;8(11):e1002967.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. Genetics. 2012;192:1065–93.
- Fitak RR. OptM: estimating the optimal number of migration edges on population trees using Treemix. Biol Methods Protoc. 2021;6(1):bpab017.
- Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. Mol Biol Evol. 2011;28:2239–52.
- Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007;449:913–8.
- 108. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. Genome Res. 2010;20:393–402.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.