



ORIGINAL RESEARCH

Haplotype-resolved Genome of Sika Deer Reveals Allele-specific Gene Expression and Chromosome Evolution



Ruobing Han[#], Lei Han[#], Xunwu Zhao, Qianghui Wang, Yanling Xia, Heping Li^{*}

College of Wildlife and Protected Area, Northeast Forestry University, Harbin 150040, China

Received 23 May 2022; revised 24 October 2022; accepted 7 November 2022
 Available online 15 November 2022

Handled by Kai Ye

KEYWORDS

Allele-specific expression gene;
 Sika deer;
 Chromosome evolution;
 Structural variation;
 Rapid antler growth

Abstract Despite the scientific and medicinal importance of diploid **sika deer** (*Cervus nippon*), its genome resources are limited and haplotype-resolved chromosome-scale assembly is urgently needed. To explore mechanisms underlying the expression patterns of the allele-specific genes in antlers and the **chromosome evolution** in Cervidae, we report, for the first time, a high-quality haplotype-resolved chromosome-scale genome of sika deer by integrating multiple sequencing strategies, which was anchored to 32 homologous groups with a pair of sex chromosomes (XY). Several expanded genes (*RET*, *PPP2R1A*, *PPP2R1B*, *YWHAB*, *YWHAZ*, and *RPS6*) and positively selected genes (*eIF4E*, *Wnt8A*, *Wnt9B*, *BMP4*, and *TP53*) were identified, which could contribute to **rapid antler growth** without carcinogenesis. A comprehensive and systematic genome-wide analysis of allele expression patterns revealed that most alleles were functionally equivalent in regulating rapid antler growth and inhibiting oncogenesis. Comparative genomic analysis revealed that chromosome fission might occur during the divergence of sika deer and red deer (*Cervus elaphus*), and the olfactory sensation of sika deer might be more powerful than that of red deer. Obvious inversion regions containing olfactory receptor genes were also identified, which arose since the divergence. In conclusion, the high-quality allele-aware reference genome provides valuable resources for further illustration of the unique biological characteristics of antler, chromosome evolution, and multi-omics research of cervid animals.

Introduction

Cervidae is the second largest family of artiodactyl ruminants (second to Bovidae). As a unique appendage organ of male cervid species (except for the reindeer), the antler grows extremely fast that exceeds even certain cancer tissues [1,2]. Thus, antler provides an excellent model for studying rapid tissue

* Corresponding author.

E-mail: lihapinghrb2002@nefu.edu.cn (Li H).

[#] Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2022.11.001>

1672-0229 © 2023 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.
 This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

growth in biological sciences. Sika deer (*Cervus nippon*) is one of the famous cervid animals producing antlers. As of now, there are not many genome resources of sika deer for the study of biology and evolution of Cervidae. Recent studies have reported that expression variation of alleles occurs frequently in mammals [3–5]. The allelic variants might affect the expression levels of alleles, and the expression variation of alleles is crucial in determining phenotypic diversity [3]. However, as a typical diploid mammal with two homologous chromosome pairs, the allelic variation of cervid species has not been elucidated yet. Therefore, deciphering the genome sequence of each allele chromosome of sika deer is substantial for understanding the expression patterns of allele-specific genes and their phenotypic characteristics (rapid antler growth and inhibition of oncogenesis) [6,7].

It is generally believed that the variation of the chromosome number and structure is one of the sources of biodiversity. The chromosome number could be increased or decreased by chromosome fission or fusion, respectively [7,8]. However, the molecular basis and consequence of chromosome variation in mammals remain unsolved [9]. The chromosome number varies dramatically in cervid animals, ranging from $2n = 6$ (female Indian muntjac) [8] to $2n = 70$ (such as Siberian roe deer) [10]. Sika deer, as a typical representative animal of the genus *Cervus* in the cervid animals, also has different karyotypes with its closely related species, red deer (*Cervus elaphus*) [11]. However, the concrete mechanism of chromosome evolution between two species still remains to be elucidated. Assembling the phased genome of sika deer contributes to understanding chromosome evolution in cervid animals [12,13]. It is also crucial to study the biodiversity and social organization of cervid species [14].

In this work, the high-quality Illumina short reads, circular consensus sequencing (CCS) data, and high-throughput chromosome conformation capture (Hi-C) data were generated, which enabled us to assemble a high-quality haplotype-resolved chromosome-level genome of sika deer with the state-of-the-art assembly method. The allele-aware genome of sika deer contributed to exploring the cause of different karyotypes between sika deer and red deer, which would be of help for investigating the expression profiles of alleles in antler and researching the diversity and variation of alleles on homologous chromosomes. Our study also suggested the possible molecular basis for rapid antler growth. Overall, the high-quality genome and annotation information reported here not only investigated differences in expression between alleles on homologous chromosomes, but also provided valuable data and resources for studying structural variations (SVs), the mechanism of chromosome evolution of sika deer, and the molecular basis of rapid antler growth.

Results

Haplotype-resolved chromosome-scale genome assembly and annotation of sika deer

A total of 143 Gb ($53\times-55\times$) Illumina short reads and 96.4 Gb ($36\times-37\times$) PacBio CCS long reads were generated (Table S1). Hifiasm [15,16] and DipAsm pipeline [7] were used to assemble an accurate allele-aware genome of sika deer (Figure 1A), and the final diploid assembly of sika deer was phased into two

haplotypes named “haplotype 1” (Hap1) and “haplotype 2” (Hap2). Hap1 had the size of 2.71 Gb with a contig N50 length of 34.98 Mb, and Hap2 had the size of 2.58 Gb with a contig N50 length of 38.09 Mb (Table S2), indicating the well-resolved haplotype assembly (Figure 1B and C). For improving the haplotype-resolved genome to chromosome scale, approximately 269-Gb ($99\times-104\times$) Hi-C paired-end reads were obtained to anchor the contigs into chromosomes. The final assembled monoploid genome of sika deer contained 66 chromosomes, comprising the 32 homologous groups with a pair of sex chromosomes (XY) (Figure 1A and Figure 2). A total of 98.05%–100% of the phased scaffolds were anchored to 66 chromosomes, indicating that most chromosomes were phased correctly (Table S3). The Hi-C interaction matrix of Hap1 and Hap2 also indicated that the chromosome groups were clear cut (Figure 1D). Additionally, Hap1 and Hap2 had 94.5% and 95.0% Benchmarking Universal Single-Copy Orthologs (BUSCO) genes, respectively (Table S4). Compared with the chromosomal-level genome of female sika deer just released recently (Tables S5 and S6) [17], a haplotype-resolved chromosome-level genome of male sika deer assembled in this study had higher contig N50 and higher BUSCO scores, indicating that our assemblies are of high quality.

A total of 22,144 (Hap1) and 18,705 (Hap2) protein-coding genes were predicted in the monoploid genome by a combined strategy of *de novo* gene prediction, homology-based search, and RNA sequencing (RNA-seq). The results of Core Eukaryotic Genes Mapping Approach (CEGMA) showed that 99.57% (Hap1) and 98.28% (Hap2) of genes in the haplotype-resolved genome were predicted, respectively. The mapping ratios of Illumina reads were 99.83% and 98.92% in Hap1 and Hap2, respectively, and the mapping ratios of Expressed Sequence Tag (EST) sequences were both higher than 95% (Table S7). At least 93.80% and 91.60% of protein-coding genes were functionally annotated against databases in Hap1 and Hap2, respectively (Table S8). More than 93.74%–98.37% of the transcripts could be mapped to Hap1, whereas more than 92.22%–97.04% of the transcripts could be mapped to Hap2 (Tables S9 and S10). All the aforementioned results, together with the genome assembly quality standard established by the Vertebrate Genome Project consortium [18], indicate that the allele-aware chromosome-scale genome assembly and annotation of sika deer are of high quality. In addition, non-coding RNAs (ncRNAs) were predicted in the haplotype-resolved chromosome-level genome of sika deer (Table S11).

The allelic chromosome pairs were systematically compared for assessing the differences between the two haplotypes. The results showed that the homologous chromosomes were highly similar with respect to gene number, exon number, intron number, and repeat content (Figure 2; Tables S12 and S13), suggesting that two allelic chromosome pairs of sika deer were functionally equivalent. Nonsynonymous mutation (K_a)/synonymous mutation (K_s) of syntenic gene pairs also had no considerable difference between the two haplotypes (Figure 2). As the homologous chromosomes of sika deer had a highly similar gene content, Hap1 was used to represent the monoploid sika deer in the following analysis except for a special description. Collectively, the results of multiple approaches revealed the high-quality haplotype-resolved chromosome-scale genome of sika deer.

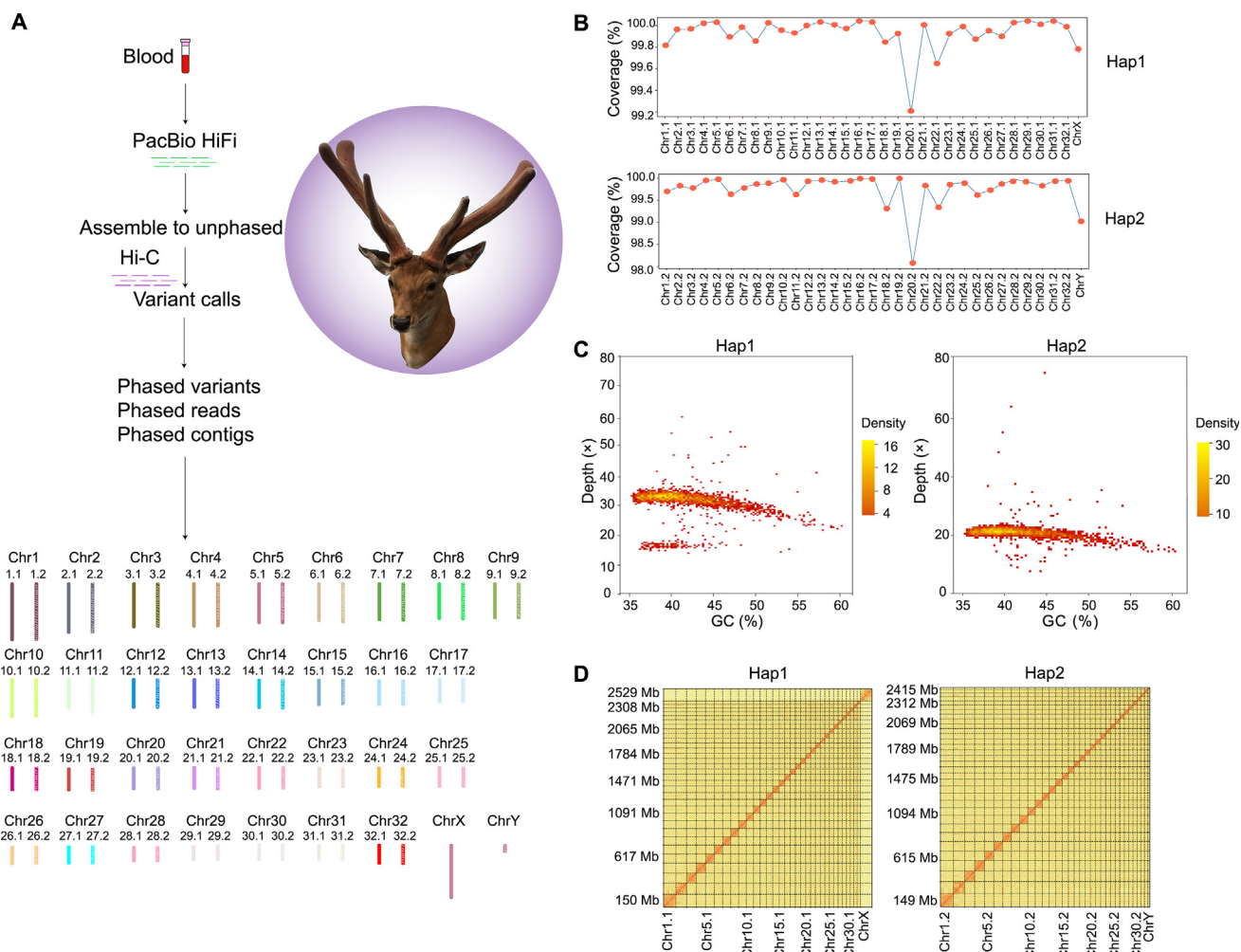


Figure 1 *De novo* assembly and assessment of genome quality

A. Overview of the *de novo* assembly of the haplotype-resolved chromosome-level genome of sika deer. **B.** Sequencing coverage of Hap1 and Hap2. **C.** GC depth distributions of Hap1 and Hap2. **D.** Hi-C interaction matrices of Hap1 and Hap2. The diagonal bar represents the frequency of contact between two loci on a chromosome, and the color from light to dark indicates the contact density from low to high. HiFi, high fidelity; Hi-C, high-throughput chromosome conformation capture; Chr, chromosome; Hap1, haplotype 1; Hap2, haplotype 2.

Phylogenetic relationship and demographic history of sika deer

The sika deer gene model was clustered with the genes from 13 mammals (see Materials and methods). A total of 2665 single-copy homologous genes were identified as shared by 14 mammalian genomes, which were used to construct a phylogenetic tree (Figure 3A). The results showed that the cervid species were closely related to *Bos taurus* and *Capra hircus*, and the common ancestor of cervid species diverged around 26.14 million years ago (MYA). In the cervid species, sika deer was the sister lineage of red deer and Tarim red deer, followed by muntjac, and then white-tailed deer and reindeer. The sika deer also demonstrated overall strong syntenic relationships with red deer and cattle, providing evidence for their phylogenetic relationships (Figure 3B).

To construct and investigate the demographic history of sika deer, the pairwise sequentially Markovian coalescent (PSMC) model was applied to infer the changes in the effective population size (N_e) of the ancestral populations of sika deer, reindeer, and cattle. The N_e of the ancestral population of sika

deer peaked twice at about 0.04 MYA and 0.25 MYA, respectively, whereas it increased sharply in 0.3–0.8 MYA. During the same period, the N_e of the ancestral population of cattle gradually decreased. Additionally, the N_e of the ancestral population of sika deer dropped two times starting from 0.01–0.04 MYA, and it underwent severe bottlenecks during the Last Glacial Maximum (Figure 3C), providing powerful evidence for the low genetic diversity of the modern population of sika deer [19]. It is probable that the uplift of the Tibetan Plateau led to a sharp decrease in the distribution of sika deer, and the habitat of sika deer has been destroyed to a certain extent by large-scale deforestation caused by humans [20,21]. Taken together, frequent human activities after the ice age were at least partially responsible for the low genetic diversity of the modern population of sika deer [20,21].

Expanded gene families involved in rapid antler growth

To elucidate the biological characteristics and adaptive evolution of sika deer, the gene families between sika deer

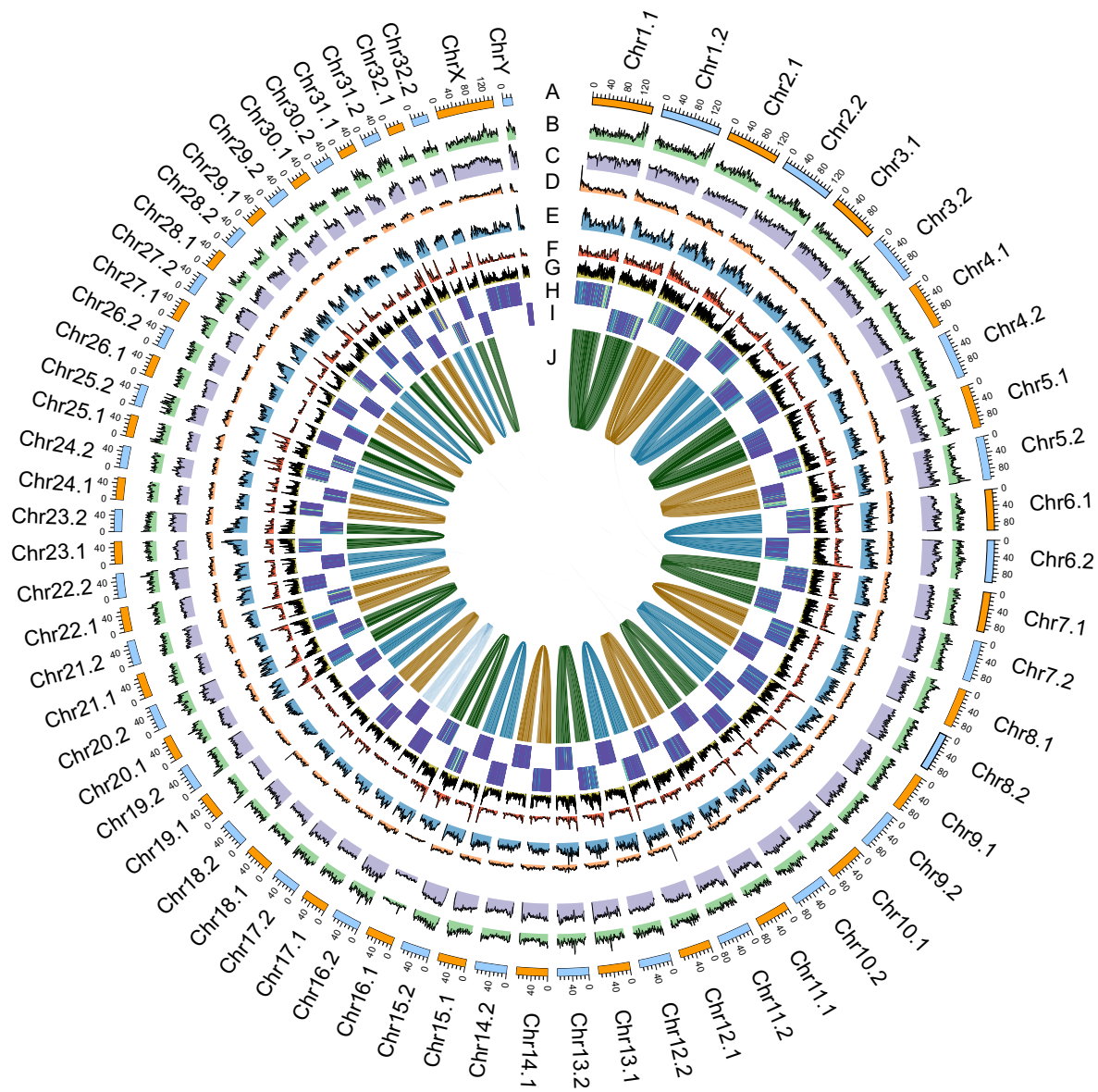


Figure 2 The overview of haplotype-resolved chromosome-level genome of sika deer

A. Length (Mb) of the chromosome. B. Density of LTR transposons. C. Density of LINE transposons. D. Density of DNA transposons. E. Density of SINE transposons. F. Gene number. G. GC content (the windows of 1 Mb). H. K_a/K_s of syntenic gene pairs in Hap1. I. K_a/K_s of syntenic gene pairs in Hap2. J. Links between the core connected alleles. LTR, long terminal repeat; LINE, long interspersed nuclear element; SINE, short interspersed nuclear element; K_a , nonsynonymous mutation; K_s , synonymous mutation.

and aforementioned 13 other mammals were analyzed, revealing that 378 significantly expanded gene families were functionally related to signal transduction (Hippo signaling pathway, PI3K-AKT signaling pathway, and calcium signaling pathway), cell growth and death (cell cycle and apoptosis), and pathways in cancer (Table S14). A total of 78 gene families were identified as significantly contracted, which were significantly enriched in ABC transporters, tight junction, focal adhesion, and ECM-receptor interaction (Table S15). In addition, a total of 42 genes were considered as positively selected genes (PSGs) (Table S16), which were functionally enriched in multiple signal transduction pathways, including MAPK signaling pathway, mTOR signaling pathway, Wnt signaling pathway, and Hippo signaling pathway (Table S17). The pre-

vious studies revealed that the PI3K-AKT signaling pathway was critical for the rapid antler growth with inhibition of oncogenesis [22,23], suggesting that these expanded gene families and PSGs are closely associated with rapid antler growth.

A total of 10 expanded genes (*COL4A1*, *COL4A2*, *COL4A5*, *COL4A6*, *RET*, *PPP2R1A*, *PPP2R1B*, *YWHAB*, *YWHAZ*, and *RPS6*) and 5 PSGs (*eIF4E*, *Wnt8A*, *Wnt9B*, *BMP4*, and *TP53*) were found to be functionally enriched in PI3K-AKT signaling pathway and related signaling pathway (Figure 4), of which three genes (*YWHAB*, *YWHAZ*, and *RPS6*) responding to DNA damage, cell proliferation, and cell apoptosis were significantly expanded in sika deer [24–27]. We also observed that six oncogenes were considered as expanded genes (*RET*, *PPP2R1A*, and *PPP2R1B*) or PSGs (*eIF4E*,

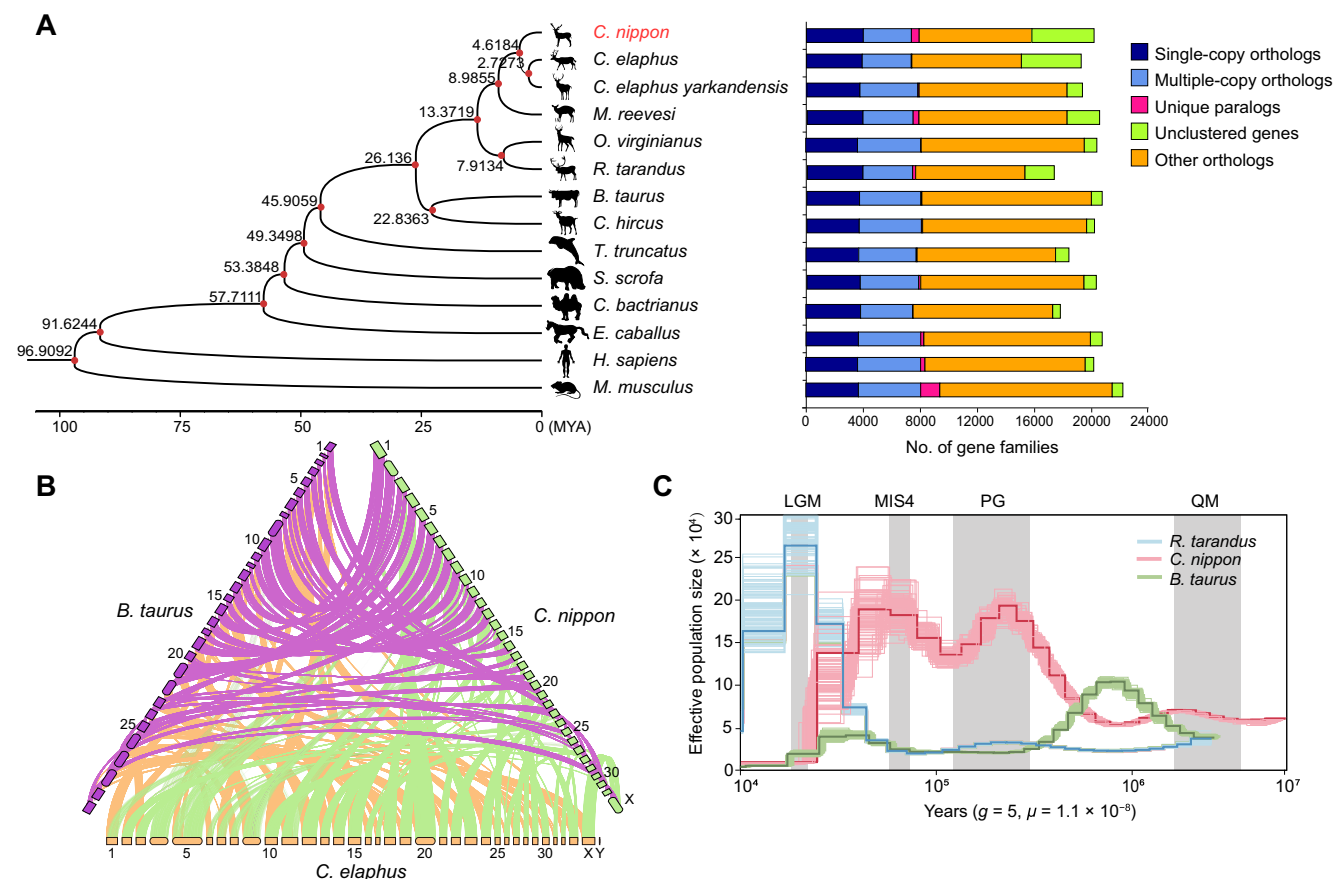


Figure 3 The phylogenetic relationship, synteny, and population size of sika deer

A. The phylogenetic relationship of sika deer and other 13 mammals. The divergence time is shown in the phylogenetic tree. The red nodes indicate that the support values of the branches are 100. The results of gene families are shown in the bar charts on the right side of the phylogenetic tree. **B.** Synteny analysis of sika deer, red deer, and cattle. Purple, green, and orange boxes represent the chromosomes of cattle, sika deer, and red deer, respectively. Different synteny blocks between one species and another are linked by lines of different colors. **C.** The effective population size of sika deer, cattle, and reindeer. “*g*” represents the generation length, and “*μ*” represents the mutation rate per generation. MYA, million years ago; LGM, Last Glacial Maximum; MIS4, Marine Isotope Stage 4; PG, Penultimate Glaciation; QM, Qingzang Movement.

BMP4, and *TP53*). Among them, *PP2A* regulates cell division, cell metabolism, and apoptosis by dephosphorylation of key proteins [28,29]. It is also important in the regulation of cell growth and proliferation, and the mice with knocked-out *PP2Ac* α subunit died in the embryonic stage [30,31]. More importantly, *RET* was significantly expanded in the sika deer compared with the 13 mammals. Previous studies reported that *RET* could regulate cell growth and differentiation by activating several downstream signaling pathways [32–34]. These results reveal that these genes enriched in the PI3K-AKT signaling pathway and related signaling pathways, together with genes involved in response to DNA damage, cell proliferation, and apoptosis, could coordinately regulate the rapid antler growth and possibly prevent the onset of cancer.

Genomic SVs and chromosome evolution between sika deer and red deer

Overall strong syntenic relationships between diploid sika deer and red deer were observed. It is worth noting that Chr1 of monopleid sika deer had strong syntenic relationships with

Chr4 and Chr23 of red deer (Figure 5A), which caused the different karyotypes between sika deer ($2n = 66$) and red deer ($2n = 68$). To further ascertain the mechanism of chromosome evolution between them, the haplotype-resolved genome of sika deer was used for variant calling. Compared with red deer, SVs were found to be distributed on the overwhelming majority of chromosomes in sika deer, among which the inversion was biased toward Chr1 and Chr28 of sika deer, accounting for 29% and 32% of the total length of Chr1 and Chr28, respectively. All these results indicated that there existed inversions during the divergence between Chr1 of sika deer and Chr4 and Chr23 of red deer. Additionally, inversions were also observed in Chr28 of sika deer and Chr2 of red deer (Figure 5B–D).

To test whether these inversion regions were associated with specific functions, the genes distributed in Chr1 inversion regions of sika deer and corresponding regions of red deer were identified, respectively. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed on the genes that were located in the Chr1 inversion region of sika deer, which were mainly enriched in some signaling

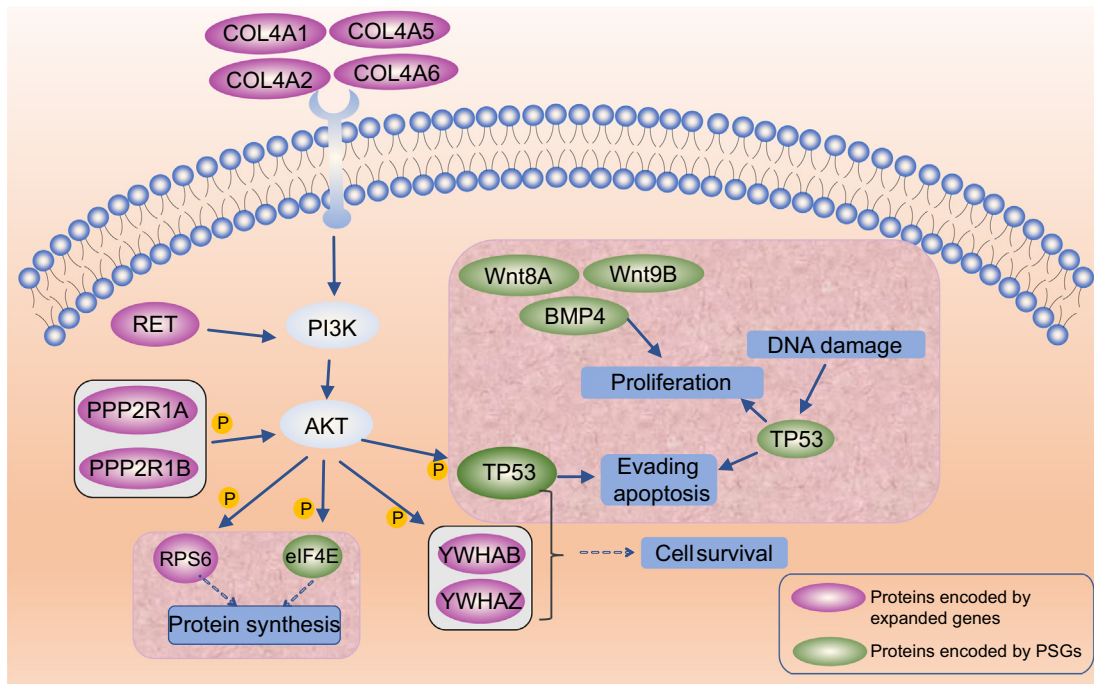


Figure 4 PI3K-AKT signaling pathway

Purple indicates proteins encoded by expanded genes in sika deer, and green indicates the proteins encoded by PSGs in sika deer. “P” represents phosphorylation. PSG, positively selected gene.

pathways related to biosynthesis (folate biosynthesis, ovarian steroidogenesis, and steroid hormone biosynthesis) and metabolism (metabolic pathways and arachidonic acid metabolism) (Table S18). Compared with the gene sets that were distributed in the Chr1 inversion region of sika deer, a total of 78 genes were lost in Chr4 and Chr23 of red deer. These 78 genes were also searched on National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/>) to reduce the false-positive results. It turned out that the Chr4 and Chr23 of red deer lost 33 genes compared with Chr1 of sika deer, inferring that the genome quality of red deer was not sufficient for investigating all genes. We further observed that 16 genes on Chr1 of sika deer distributed on multiple chromosomes of red deer, such as Chr10, Chr11, and Chr29, revealing that these 16 genes might be translocated during the divergence between sika deer and red deer. Finally, Chr4 and Chr23 in red deer uniquely lost 17 genes, one of which was enriched in olfactory transduction (*Calml3*) (Figure 5D).

The Chr28 in two haplotypes of sika deer showed high homology with Chr2 of red deer (Figure 5C). The genes distributed in the inversion regions of Chr28 in sika deer and corresponding regions of Chr2 in red deer were identified, respectively. KEGG enrichment analysis was performed on the gene sets that were distributed in the Chr28 inversion regions of sika deer, which were aligned in the p53 signaling pathway, HIF-1 signaling pathway, glycolysis/gluconeogenesis, propanoate metabolism, and pyruvate metabolism (Table S19). Compared with the gene sets that were distributed in the Chr28 inversion regions of sika deer by the aforementioned methods, seven genes were translocated, and six genes were lost in Chr2 of red deer. Among them, three genes were enriched in olfactory transduction (*Olfr149*, *OR10G7*, and *OR10S1*) (Figure 5C), suggesting that the olfactory sensation

of sika deer might be more powerful than that of red deer. Overall, our results cast a novel light on the mechanism of chromosome evolution between sika deer and red deer.

To further explore the impact of chromosome evolution on three-dimensional (3D) chromatin architectures in sika deer, we performed the 3D chromatin architecture analysis on the sika deer genome, including compartment A/B and topologically associated domains (TADs). The compartment A/B was identified at 500-kb resolution. As in previous studies [9], the results also showed that compartment A had higher GC content and gene density than compartment B (Figures S1 and S2). In addition, compartment A and compartment B were randomly distributed on all chromosomes without bias (Figure S3), speculating that compartment A/B may have little effect on chromosome evolution during the divergence between sika deer and red deer.

TADs were considered as fundamental units of 3D eukaryotic genome organization. In the present study, a total of 3427 self-interacting regions were identified at 40-kb resolution, with an average length of 713 kb. There is no correlation between the percentage of TADs on the chromosome and the length of the chromosome. However, the large chromosomes generally had more TADs than the short ones (Figure S4). In addition, the Hi-C contact heatmap of Chr1 was identified, including Hi-C contact heatmaps of two inversion sites (Figure S5). TADs were also identified at the sites of inversion regions in Chr1. The results showed that the TADs were scattered in the inversion regions of 30–70 Mb at 500-kb resolution (Figure S6). TADs were identified at both ends of the inversion regions to further understand the chromosome evolution in sika deer at 40-kb resolution. Multiple consecutive TADs were found in 28–33 Mb and 68–73 Mb (Figures S7 and S8). TADs were also identified in another

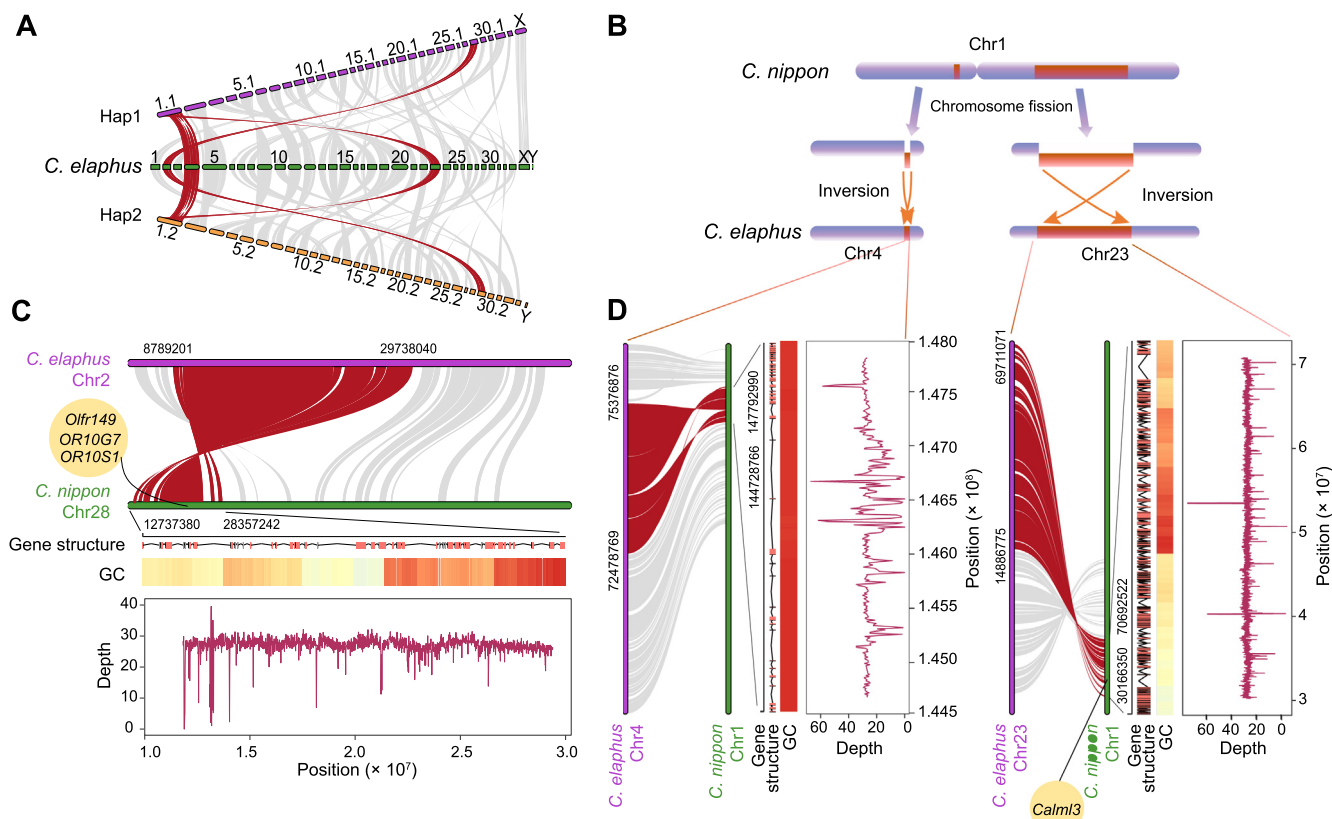


Figure 5 Chromosome evolution between sika deer and red deer

A. Synteny of the haplotype-resolved chromosome-level genome of sika deer and red deer. Synteny blocks between chromosomes of sika deer and red deer are illustrated by red and gray lines, with red lines indicating the inversion regions. **B.** Schematic diagram of divergence between sika deer and red deer. Orange indicates the inversion in the chromosomes. **C.** Synteny of Chr28 in sika deer and Chr2 in red deer. Red indicates the inversion regions. **D.** Synteny of Chr1 in sika deer and Chr4 in red deer, and synteny of Chr1 in sika deer and Chr23 in red deer. Red indicates the inversion regions. The schematic diagram of gene structure shows the distribution of genes in the inversion regions. The heatmap shows the GC content in the inversion regions. The line chart shows the distribution of sequencing depth in the inversion regions.

inversion regions (144–148 Mb), and several consecutive TADs were detected (Figure S9). However, this does not mean that the chromosome evolution between sika deer and red deer was greatly affected by TADs.

Expression pattern of alleles during the sika deer antler growth

Analysis of transcriptome data using one haploid genome as a reference genome could miss allele-specific expression (ASE) or novel expression patterns [35]. Therefore, the gene expression profiles of antlers were reanalyzed using our previous transcriptome data generated from three different periods that represented the whole antler development (BioProject: PRJNA552158) [36] (Figure 6A, Figure S10). The results of principal component analysis (PCA) of the monoploid genome of sika deer indicated that the antler transcriptomes were mainly shaped by genome-wide ASE, followed by the expression of genes at different developmental periods (Figure 6B). Additionally, the haploid genome sequences were aligned stringently for understanding the sequence divergence of two haplotypes, showing the 99.6% sequence identity (Figure 6C).

A total of 12,534 reliable homologous genes were identified in allelic chromosome pairs by combining the synteny and coordinate strategies (Table S20). The vast majority of alleles (97.12%) were found to be coordinately expressed in the two haplotypes during the rapid antler growth (Figure 6D), suggesting that the expression is generally not biased between the two haplotypes and most alleles have similar functions in regulating the rapid antler growth. The alleles in antler were also discovered to distribute on 32 homologous chromosomes without any bias (Figure 6E), revealing that the ASE genes were distributed randomly throughout the sika deer genome. Additionally, the phased diploid genome facilitated the detection of SVs between two haplotypes, including deletion and insertion. The SVs were distributed on 32 autosomes of sika deer (Figure 6F; Table S21) and spanned 18.9 Mb, representing almost 0.74% of the haploid genome (Figure 6G).

The ASE genes were investigated in the haplotype-resolved chromosome-level genome of sika deer without the parental information, which were in allele imbalance between two haplotypes. Only 2.9% (361) of homologous genes were regarded as ASE genes with the expression biased toward a single

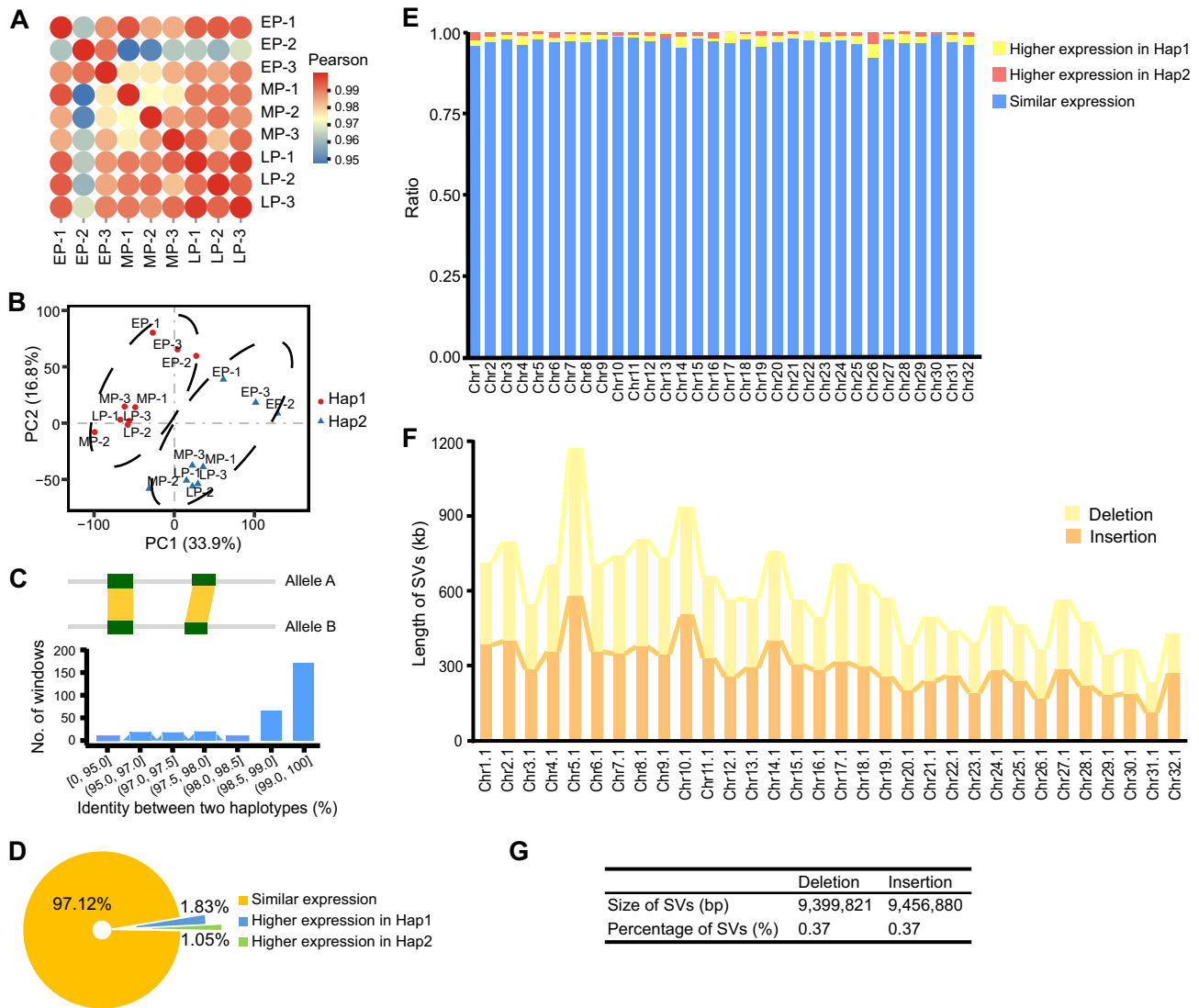


Figure 6 Haplotype comparison of diploid sika deer

A. Correlation heatmap of transcriptome data in Hap1. EP indicates the stage of antler growing to a saddle-like appearance; MP indicates the stage of antler growing with two branches; LP indicates the stage of antler growing with three branches. **B.** PCA of allele expression profiles of sika deer antler at three developmental periods. **C.** Comparison of two haplotypes using 10-Mb nonoverlapping windows. **D.** Statistics of alleles. **E.** Distribution histogram of alleles on the chromosomes. **F.** Distribution histogram of SVs between haplotypes on chromosomes. **G.** Statistics of SVs between two haplotypes. PCA, principal component analysis; PC, principal component; SV, structural variation.

haplotype, of which the expression of 229 ASE genes was biased toward the Hap1 (Figure S11) and the expression of 132 ASE genes was biased toward the Hap2 (Figure S12). These ASE genes showed functional enrichment in multiple biological processes, including ribosome, HIF-1 signaling pathway, axon guidance, mitochondrial biogenesis, metabolism of xenobiotics by cytochrome P450, and chemical carcinogenesis-reactive oxygen species (Figure S13). There were some oncogenes obviously biased toward Hap1 or Hap2, such as *RPLP1*, *RPL3*, *RPS10*, *RPL10*, *RPL23a*, *SLC7A3*, *COL2A1*, and *PEBP1*, indicating that the alleles might interplay to regulate rapid antler growth. In addition, based on the differential expression patterns observed in two haplotypes, we defined the smaller allele expression differences

with $2 < |\log_2 \text{fold change (FC)}| < 8$ ($P < 0.05$) and larger allele expression differences with $|\log_2 \text{FC}| > 8$ ($P < 0.05$). Results showed that the allele expression differences of diverse categories were relatively stable in two haplotypes (Figure S14).

Discussion

The genetic research of sika deer and related development efforts have not been commensurate with its importance due to the scarcity of genome resources. Currently, we decode the first high-quality haplotype-resolved chromosome-scale genome of diploid sika deer by combining the new sequencing

technology and Hi-C scaffolding, which is critical for studying the role of variation in genome function and phenotype [37]. We constructed the expression profile of alleles in the sika deer antler, and proposed the possible molecular basis for rapid antler growth. Our study also contributes to research on the chromosome evolution of cervid animals, especially the mechanism of chromosome evolution between sika deer and red deer. In summary, this high-quality monoploid genome is of sufficient quality for exploring the biological characteristics of cervid species and future multi-omics research of sika deer.

In the present study, the identification of the Y chromosome of sika deer undoubtedly fills the gap in reference genome resources of male sika deer. The genome of male sika deer is beneficial to genomic selection breeding, which is also helpful to retain the germplasm resources with higher antler yields. More importantly, two haplotype genomes of sika deer could more accurately and completely reflect its genetic information, which provides a more complete reference genome for the study of sika deer. The benefit of the haplotype-resolved chromosome-level genome of sika deer also includes that it can be used to explore the expression of alleles and the differences in some phenotypic characteristics of sika deer [38]. This has been confirmed by many haplotype-resolved genomes published recently, such as the haplotype-resolved genomes of plants have filled an important gap in exploring their unique traits using ASE analysis [15,39]. Additionally, as a secondary sex characteristic of male cervid animals, male cervid species annually grow deciduous antlers. Thus, comparing the diversity of alleles between the two haplotypes of sika deer contributes to understanding the potential molecular basis of male antler growth, and it is important to study the molecular mechanism of the unique biological characteristics of antlers.

In the genus *Cervus* of cervid animals, the variation of chromosome number is dominated by the variation of autosome number, of which chromosome fission is considered to be the main molecular mechanism [11]. In our study, the results revealed that Chr1 of monoploid sika deer showed strong syntenic relationships with Chr4 and Chr23 of red deer, inferring that the sika deer ($2n = 66$) may diverge into red deer ($2n = 68$) through chromosome fission. The results of SVs between sika deer and red deer revealed that inversion played a dominant role in this process, providing a novel perspective to understand the mechanism of chromosome evolution in cervid animals. However, we cannot further determine the chromosome evolution between sika deer and red deer by comparing the 3D chromatin architectures due to the unavailability of Hi-C data of red deer. Additionally, red deer was found to uniquely lose three genes associated with olfactory by comparing the gene sets in the inversion regions of sika deer and red deer. A previous study also analyzed the adaptive evolution of olfactory-related genes in cervid animals using comparative genomics [40]. Our study provides a further solid foundation and detailed reference for the studies of the adaptive evolution of olfactory-related genes in cervid species.

The role of alleles on homologous chromosome pairs in cervid animals has long been overlooked. The phased haplotype-resolved genome of sika deer facilitated the interpretation of the expression patterns and functions of alleles. In the present study, the alleles generally had no notable differential expression between two haplotypes, whereas

differential gene expression was still observed in a few ASE genes, which was potentially associated with multiple biological processes. We found that some genes showing ASE were involved in oncogenesis. However, whether this means that one of the haplotypes plays a more important role during the rapid antler growth without carcinogenesis requires further experiments.

A recent study reported that the expression of proto-oncogenes was vital for antler fast growth [23], and the expression of several tumor suppressor genes was necessary for antler growth and inhibition of oncogenesis [23,41]. We found that several proto-oncogenes and tumor suppressor genes were identified as the expanded genes or PSGs in the sika deer lineage, which were assigned to the PI3K-AKT signaling pathway and related pathways. It has been postulated that the PI3K-AKT signaling pathway was involved in regulating cell proliferation, differentiation, migration, and cell-cycle progression [42], and it was one of the crucial pathways regulating the initiation, development, and regeneration of antler [42–45]. Thus, we inferred that the multiple copies of *RET* in sika deer might enhance the rapid antler growth through interaction with the PI3K-AKT signaling pathway. *TP53* has one copy and it was identified as a PSG in the sika deer lineage, which strengthens the insight that cervid species may have evolved an enhanced p53 signaling pathway for an efficient cancer-defense mechanism [46]. The rapid antler growth is a complex process regulated by multiple factors, and these expanded genes and PSGs (*RET*, *PPP2R1A*, *PPP2R1B*, *eIF4E*, *BMP4*, and *TP53*) might play a pivotal role in this process. These cancer-related genes might coordinately regulate the rapid antler growth, and form the unique cancer defense mechanism of cervid species. To thoroughly elucidate the molecular mechanism of antler rapid growth and defend against cancer in cervid species, other haplotype-resolved chromosome-scale genomes of cervid species and further functional experiments are required in the future. In summary, our haplotype-resolved chromosome-scale genome of sika deer offers a holistic view of its expression profiles, functions of alleles, and chromosome evolution of cervid species. It will also be helpful for studying the therapy of cancer in humans.

Conclusion

In the present study, we provided a haplotype-resolved chromosome-scale genome, which is, to our knowledge, the first high-quality available diploid genome of sika deer. This allowed us to explore the ASE patterns between the two homologous chromosomes. Most alleles were found to be co-expressed in the rapidly growing antlers, while at the same time preventing the onset of cancer. Several expanded genes or PSGs (*RET*, *PPP2R1A*, *PPP2R1B*, *YWHAB*, *YWHAZ*, and *RPS6*) were also considered to play a key role in this process. In addition, our results revealed that chromosome fission might occur during the divergence of sika deer and red deer, which resulted in an increase in the chromosome numbers of red deer. Overall, our study will promote the research of the unique characteristics of antler (rapid growth and low cancer rate) and the chromosome evolution in cervid species. It also provides valuable resources and references for multi-omics studies of sika deer.

Materials and methods

Sample collection

A fresh blood sample was collected from a 7-year-old healthy male sika deer in Jindi Deer Industry Co. Ltd. (Ear tags: 1220; Harbin, Heilongjiang Province, China), whose parents were both purebred sika deer. DNA was extracted from the fresh blood sample for constructing the libraries.

Library construction and sequencing

An Illumina paired-end library was constructed using the TruSeq Nano DNA HT Sample Preparation Kit (Catalog No. TG-202-1003, Illumina, San Diego, CA) according to the manufacturer's instructions, and sequenced by Illumina NovaSeq 6000 (Illumina, San Diego, CA). Three paired-end CCS libraries with the small insert size of 15 kb were constructed, and sequenced using the PacBio Sequel II platform (PacBio, San Francisco, CA). High-fidelity (HiFi) reads were generated by CCS software (<https://github.com/PacificBiosciences/ccs>). To improve the allele-aware genome to chromosome scale, two Hi-C libraries were constructed and sequenced by Illumina NovaSeq 6000 using DNA extracted from the same individual.

Genome assembly

Haplotype-resolved genome of sika deer was assembled using Hifiasm [15,16] with the default parameters. DipAsm pipeline [7] was applied to combine HiFi reads and Hi-C raw data. The haplotype-resolved chromosome-level genome was assembled as follows: (1) Hifiasm was first used to obtain an unphased assembly. Juicer (v1.5) [47] and a 3D *de novo* assembly (3D-DNA; v180922) [48] were used to scaffold the contigs. (2) DeepVariant (v0.8.0) [49] was used to perform small variants, and Hi-C reads were mapped to the scaffolds. (3) HapCUT2 (v1.1) [50] was used to generate sparse phasing at the chromosome scale. (4) WhatsHap (v0.18) [51] was used to generate fine-scale chromosome-long phasing by combining the haplotypes with PacBio HiFi data. (5) To reference-assisted scaffolding, minimap2 (v2.17) [52] was used to align contigs. (6) Purge_dups pipeline [53] was used to remove haplotig sequences from the initial assembly genome. (7) HiC-Pro was used to align Hi-C reads to contigs, and 3D-DNA was applied for correcting misassemblies, anchor, order, and orient fragments of DNA. Juicebox Assembly Tools (v1.9.9) was used to manual correction connections. (8) Juicebox [47] and plotHicGenome (v0.1.0) were used to analyze and visualize Hi-C-assembled scaffolds. (9) Y chromosome of the sika deer was detected using the Diamond (v0.9.10) based on the previous study of the red deer genome.

Genome annotation

Gene prediction

The completeness and accuracy of the genome assembly of sika deer were evaluated based on BUSCO, and the transcriptome was mapped to the genome to assist in verifying the genome quality. Illumina short reads were mapped into the haplotype-resolved genome of sika deer by Burrows–Wheeler

Aligner (BWA) for estimating the accuracy of genome assembly, and CEGMA software was also used to estimate the quality of the genome assembly. In addition, 2743 EST sequences of sika deer were downloaded from NCBI and aligned against the two haplotype-resolved genomes to further verify the quality of the genome assembly.

Protein-coding genes were annotated by integrated evidence from the homology-based search, *de novo* prediction, and transcriptome data. Briefly, the homologous gene sets of cattle, red deer, sheep, mouse, and reindeer were downloaded from NCBI for homology-based prediction. *De novo* prediction was performed by AUGUSTUS (v3.0.3) [54] and SNAP [55]. We also downloaded the transcriptome data of multiple organs of sika deer from NCBI, including heart (SRA: SRS3900676), liver (SRA: SRS3900672), spleen (SRA: SRS3900692), lung (SRA: SRS3900690), and kidney (SRA: SRS3900689).

Gene function annotation was then performed by aligning against multiple public databases, including Non-Redundant Protein Sequence Database (NR), Gene Ontology (GO), InterProScan, Swiss-Prot, Translation of EMBL (TrEMBL), KEGG, and Clusters of orthologous groups for eukaryotic complete genomes (KOG). In addition, the haplotype-resolved genome was aligned against the Rfam database and vertebrate ribosomal RNA (rRNA) database to predict ncRNAs, including small nuclear RNA (snRNA), microRNA (miRNA), and rRNA.

Repeat annotation

The homology-based search and *ab initio* method were both used to annotate repetitive sequences of the sika deer genome. To identify the types of repeat elements, a transposable element library was constructed by Tandem Repeat Finder (TRF) [56], LTR_FINDER [57], and RepeatModeler [58]. The known repeat element was searched against the Repbase database using RepeatMasker [57] and RepeatProteinMask [59].

Construction of phylogenetic tree and estimation of divergence time

TreeFam (v4.0) [60] was used to construct gene families in the six cervid animals (sika deer, red deer, white-tailed deer, reindeer, Tarim red deer, and muntjak) and eight other mammals (cattle, sheep, human, horse, mouse, pig, camel, and dolphin). Based on the aforementioned results, CAFÉ (v4.2) [61] was used to determine the gene family expansion and contraction. KEGG enrichment analysis was implemented for expanded and contracted gene families in sika deer. To construct the phylogenetic tree, the single-copy genes shared within 14 mammalian genomes were identified.

The divergence time among different species was calculated based on fossil evidence. MCMCtree from the PAML (v4.8) [62] package was applied to estimate the divergence time with default parameters. The PSMC [63] model was used to infer the demographic history of sika deer with the parameter: *psmc* -N25 -t15 -r5 -p “4+25*2+4+6”.

Identification of PSGs

To identify the PSGs of sika deer, the high-confidence single-copy orthologous genes were identified based on the

14 mammals mentioned above. MUSCLE (v3.8.31) was used to conduct the multiple sequence alignment. The branch-site model was first selected to determine the PSGs, and the likelihood ratio test (LRT) in the CodeML of PAML was then used to detect the PSGs with the sika deer lineage as the foreground branch. *P* value was calculated by chi-square statistic and corrected by the Benjamini–Hochberg method. The genes with *P* < 0.05 were considered as PSGs. The PSGs obtained were subjected to KEGG enrichment analysis using the clusterProfiler package [64].

SVs between sika deer and red deer

Genomic SVs, detected by PacBio CCS reads, provided more convenient conditions for studying polymorphic variations, chromosome evolution [65], cancer research [66], and phenotypes in organisms [67–70]. The haplotype-resolved genome was beneficial for better interpreting the SVs. To understand the SVs between the sika deer and its sister lineage (red deer), MCScanX [71] was used to perform synteny analysis. Sniffles [72] was then used to identify SVs between sika deer and red deer, including inversion, deletion, duplication, and insertion.

3D chromatin architecture was performed on sika deer at two different hierarchical levels using HiCExplorer (<https://training.galaxyproject.org/training-material/topics/epigenetics/tutorials/hicexplorer/tutorial.html>), including compartment A/B and TADs. HicPCA was used to calculate compartment A/B, and hicFindTADs program was used to call TADs at 40-kb resolution in the sika deer genome.

Allele identification and haplotype comparison

The alleles were identified by combining synteny, coordinate, and the mapping ratio. Blastn was used to perform the multiple sequence alignment between the two haplotype genomes. MCScanX and MUMmer were used to define the synteny blocks between Hap1 and Hap2. Accordingly, the paired genes (identity > 97%) with high similarity in each synteny block were considered as reliable alleles A and B.

The phased monoploid genome was helpful for understanding the expression profile of alleles in the antlers of sika deer. In the present study, HISAT2 was used to map the RNA-seq reads that were sequenced from the mesenchymal tissue of the sika deer antler [36] to Hap1 and Hap2, respectively. The expression level of each transcript was estimated with HTSeq. The differences in expression between alleles were estimated by DESeq2 [73]. To ensure the accuracy and reliability of the results of ASE genes, the genes with the count of 0 in all samples from two haplotypes were filtered out. The homologous genes that meet $|\log_2 FC| > 2$ and *P* < 0.05 were defined as ASE genes. KEGG enrichment analyses of ASE genes were implemented by the clusterProfiler package in R. Additionally, Assemblytics [74] was used to identify SVs between haplotypes.

Ethical statement

All experimental designs and animal handling were approved by the Institutional Animal Care and Use Committee of Northeast Forestry University, China (Approval No. 2022049).

Data availability

The raw sequencing data generated in this study have been deposited in the Genome Sequence Archive [75] at the National Genomics Data Center (NGDC), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (CAS) / China National Center for Bioinformation (CNCB) (GSA: CRA007487), and are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The whole-genome sequence data reported in this study have been deposited in the Genome Warehouse [76] at the NGDC, BIG, CAS / CNCB (GWH: GWHBJVV000000000 and GWHBJVU000000000), and are publicly accessible at <https://ngdc.cncb.ac.cn/gwh>.

Competing interests

The authors have declared no competing interests.

CRedit authorship contribution statement

Ruobing Han: Resources, Writing – original draft, Data curation, Investigation, Writing – review & editing, Visualization, Formal analysis. **Lei Han:** Data curation, Visualization, Investigation, Formal analysis, Writing – review & editing. **Xunwu Zhao:** Writing – review & editing. **Qianghui Wang:** Writing – review & editing, Formal analysis. **Yanling Xia:** Resources. **Heping Li:** Funding acquisition, Conceptualization, Supervision, Project administration, Resources. All authors have read and approved the final manuscript.

Acknowledgments

We would like to thank the National Key R&D Program of China (Grant No. 2018YFC1706601) and the Natural Science Foundation of Heilongjiang Province of China (Grant No. C2017012). We thank Xianlan Cui for the language editing of this manuscript.

Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gpb.2022.11.001>.

ORCID

ORCID 0000-0003-4115-7533 (Ruobing Han)
ORCID 0000-0001-6487-7520 (Lei Han)
ORCID 0000-0002-7499-8817 (Xunwu Zhao)
ORCID 0000-0002-7959-0482 (Qianghui Wang)
ORCID 0000-0002-5407-9047 (Yanling Xia)
ORCID 0000-0001-8284-0449 (Heping Li)

References

- [1] Bubenik GA, Bubenik AB. Horns, pronghorns, and antlers. New York: Springer; 1990.
- [2] Li C, Zhao H, Liu Z, McMahon C. Deer antler - a novel model for studying organ regeneration in mammals. *Int J Biochem Cell Biol* 2014;56:111–22.

- [3] Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA, Smith OS. Allelic variation of gene expression in maize hybrids. *Plant Cell* 2004;16:1707–16.
- [4] Cowles CR, Hirschhorn JN, Altshuler D, Lander ES. Detection of regulatory variation in mouse genes. *Nat Genet* 2002;32:432–7.
- [5] Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. Allelic variation in human gene expression. *Science* 2002;297:1143.
- [6] Yang C, Zhou Y, Marcus S, Formenti G, Bergeron LA, Song Z, et al. Evolutionary and biomedical insights from a marmoset diploid genome assembly. *Nature* 2021;594:227–33.
- [7] Garg S, Fungtammasan A, Carroll A, Chou M, Schmitt A, Zhou X, et al. Chromosome-scale, haplotype-resolved assembly of human genomes. *Nat Biotechnol* 2021;39:309–12.
- [8] Yang F, O'Brien PC, Wienberg J, Ferguson-Smith MA. A reappraisal of the tandem fusion theory of karyotype evolution in the Indian muntjac using chromosome painting. *Chromosome Res* 1997;5:109–17.
- [9] Yin Y, Fan H, Zhou B, Hu Y, Fan G, Wang J, et al. Molecular mechanisms and topological consequences of drastic chromosomal rearrangements of muntjac deer. *Nat Commun* 2021;12:6858.
- [10] Yu M, Yang F, Xing X. Advances on the chromosome of cervid in China. *China Animal Husbandry Veterinary Med* 2012;39:65–8.
- [11] Tang L, Zhang R, Dong S, Xing X. Research progress in Cervidae's chromosome. *Special Wild Econ Animal Plant Res* 2020;42:61–64+89.
- [12] Li C, Yang F, Sheppard A. Adult stem cells and mammalian epimorphic regeneration - insights from studying annual renewal of deer antlers. *Curr Stem Cell Res Ther* 2009;4:237–51.
- [13] Ba H, Cai Z, Gao H, Qin T, Liu W, Xie L, et al. Chromosome-level genome assembly of Tarim red deer, *Cervus elaphus yarkandensis*. *Sci Data* 2020;7:187.
- [14] Bartoš L, Bubenik GA. Relationships between rank-related behaviour, antler cycle timing and antler growth in deer: behavioural aspects. *Anim Prod Sci* 2011;51:303–10.
- [15] Sun X, Jiao C, Schwaninger H, Chao CT, Ma Y, Duan N, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet* 2020;52:1423–32.
- [16] Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat Methods* 2021;18:170–5.
- [17] 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 for high-tannin adaptation. *Genomics Proteomics Bioinformatics* 2023;21:203–15.
- [18] A reference standard for genome biology. *Nat Biotechnol* 2018;36:1121.
- [19] Hu P, Shao Y, Xu J, Wang T, Li Y, Liu H, et al. Genome-wide study on genetic diversity and phylogeny of five species in the genus *Cervus*. *BMC Genomics* 2019;20:384.
- [20] Guo Y, Zheng H. On the geological distribution, taxonomic status of species and evolutionary history of sika deer in China. *Acta Theriol Sinica* 2000;20:168–79.
- [21] Wu F, Li H, Jin L, Li X, Ma Y, You J, et al. Deer antler base as a traditional Chinese medicine: a review of its traditional uses, chemistry and pharmacology. *J Ethnopharmacol* 2013;145:403–15.
- [22] Dong Z, Coates D. Bioactive molecular discovery using deer antlers as a model of mammalian regeneration. *J Proteome Res* 2021;20:2167–81.
- [23] 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:eaav6335.
- [24] Li Y, Mitsuhashi S, Ikejo M, Miura N, Kawamura T, Hamakubo T, et al. Relationship between ATM and ribosomal protein S6 revealed by the chemical inhibition of Ser/Thr protein phosphatase type I. *Biosci Biotechnol Biochem* 2012;76:486–94.
- [25] Chen B, Zhang W, Gao J, Chen H, Jiang L, Liu D, et al. Downregulation of ribosomal protein S6 inhibits the growth of non-small cell lung cancer by inducing cell cycle arrest, rather than apoptosis. *Cancer Lett* 2014;354:378–89.
- [26] Stevers L, Sijbesma E, Botta M, Mackintosh C, Obsil T, Landrieu I, et al. Modulators of 14-3-3 protein–protein interactions. *J Med Chem* 2018;61:3755–78.
- [27] Aitken A. 14-3-3 proteins: a historic overview. *Semin Cancer Biol* 2006;16:162–72.
- [28] Mumby M. PP2A: unveiling a reluctant tumor suppressor. *Cell* 2007;130:21–4.
- [29] Lin YC, Chen HF, Lin LN, Jie L, Wen L, Zhang SJ, et al. Identification and functional analyses of polymorphism haplotypes of protein phosphatase 2A- α gene promoter. *Mutat Res* 2011;716:66–75.
- [30] Rahman M, Nakayama K, Rahman MT, Nakayama N, Katagiri H, Katagiri A, et al. *PPP2R1A* mutation is a rare event in ovarian carcinoma across histological subtypes. *Anticancer Res* 2013;33:113–8.
- [31] Li W. Mutational analysis of the *ppp2r1a* and *ppp2r1b* genes in Chinese patients with ovarian cancer. A Master thesis. Nanchang University 2014.
- [32] De Groot JWB, Links TP, Plukker JTM, Lips CJM, Hofstra RMW. *RET* as a diagnostic and therapeutic target in sporadic and hereditary endocrine tumors. *Endocr Rev* 2006;27:535–60.
- [33] Frank-Raue K, Rondot S, Raue F. Molecular genetics and phenomics of *RET* mutations: impact on prognosis of MTC. *Mol Cell Endocrinol* 2010;322:2–7.
- [34] Manié S, Santoro M, Fusco A, Billaud M. The RET receptor: function in development and dysfunction in congenital malformation. *Trends Genet* 2001;17:580–9.
- [35] Qi W, Lim YW, Patrignani A, Schlapfer P, Bratus-Neuenschwander A, Gruter S, et al. The haplotype-resolved chromosome pairs of a heterozygous diploid African cassava cultivar reveal novel pan-genome and allele-specific transcriptome features. *Gigascience* 2022;11:giac028.
- [36] Han R, Han L, Wang S, Li H. Whole transcriptome analysis of mesenchyme tissue in sika deer antler revealed the ceRNAs regulatory network associated with antler development. *Front Genet* 2020;10:1403.
- [37] Fernando A. The necessity of diploid genome sequencing to unravel the genetic component of complex phenotypes. *Front Genet* 2017;8:148.
- [38] Low W, Tearle R, Liu R, Koren S, Rhie A, Bickhart D, et al. Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. *Nat Commun* 2020;11:2071.
- [39] Cheng SP, Jia KH, Liu H, Zhang RG, Li ZC, Zhou SS, et al. Haplotype-resolved genome assembly and allele-specific gene expression in cultivated ginger. *Hortic Res* 2021;8:188.
- [40] Ba H, Qin T, Cai Z, Liu W, Li C. Molecular evidence for adaptive evolution of olfactory-related genes in cervids. *Genes Genomics* 2020;42:355–60.
- [41] Chonco L, Landete-Castillejos T, Serrano-Heras G, Serrano MP, Perez-Barberia FJ, Gonzalez-Armeito C, et al. Anti-tumour activity of deer growing antlers and its potential applications in the treatment of malignant gliomas. *Sci Rep* 2021;11:42.
- [42] Sun H, Sui Z, Wang D, Ba H, Zhao H, Zhang L, et al. Identification of interactive molecules between antler stem cells and dermal papilla cells using an *in vitro* co-culture system. *J Mol Histol* 2020;51:15–31.
- [43] Ruan HN, Luo JC, Wang LL, Wang J, Wang ZY, Zhang J. Sika deer antler protein against acetaminophen-induced nephrotoxicity by activating Nrf2 and inhibition FoxO1 via PI3K/Akt signaling. *Int J Biol Macromol* 2019;141:961–87.

- [44] Dong Z, Ba H, Wei Z, Coates D, Li C. iTRAQ-based quantitative proteomic analysis of the potentiated and dormant antler stem cells. *Int J Mol Sci* 2016;17:1778.
- [45] Su H, Tang X, Zhang X, Liu L, Jing L, Pan D, et al. Comparative proteomics analysis reveals the difference during antler regeneration stage between red deer and sika deer. *PeerJ* 2019;7:e7299.
- [46] Lin Z, Chen L, Chen X, Zhong Y, Yang Y, Xia W, et al. Biological adaptations in the Arctic cervid, the reindeer (*Rangifer tarandus*). *Science* 2019;364:eaav6312.
- [47] Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* 2016;3:95–8.
- [48] Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 2017;356:92–5.
- [49] Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* 2018;36:983–7.
- [50] Edge P, Bafna V, Bansal V. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Res* 2017;27:801–12.
- [51] Patterson M, Marshall T, Pisanti N, Van Iersel L, Stougie L, Klau GW, et al. WhatsHap: weighted haplotype assembly for future-generation sequencing reads. *J Comput Biol* 2015;22:498–509.
- [52] Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–100.
- [53] Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 2020;36:2896–8.
- [54] Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 2004;32:W309–12.
- [55] Korf I. Gene finding in novel genomes. *BMC Bioinformatics* 2004;5:59.
- [56] Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999;27:573–80.
- [57] Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 2007;35:W265–8.
- [58] Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A* 2020;117:9451–7.
- [59] Tempel S. Using and understanding RepeatMasker. *Methods Mol Biol* 2012;859:29–51.
- [60] Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, Osmotherly L, et al. TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* 2006;34:D572–80.
- [61] Bie TD, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 2006;22:1269–71.
- [62] Xu B, Yang Z. PAMLX: a graphical user interface for PAML. *Mol Biol Evol* 2013;30:2723–4.
- [63] Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature* 2011;475:493–6.
- [64] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284–7.
- [65] Lupski JR. Structural variation mutagenesis of the human genome: impact on disease and evolution. *Environ Mol Mutagen* 2015;56:419–36.
- [66] Macintyre G, Ylstra B, Brenton JD. Sequencing structural variants in cancer for precision therapeutics. *Trends Genet* 2016;32:530–42.
- [67] Zichner T, Garfield DA, Rausch T, Stutz AM, Cannavo E, Braun M, et al. Impact of genomic structural variation in *Drosophila melanogaster* based on population-scale sequencing. *Genome Res* 2013;23:568–79.
- [68] Dennenmoser S, Sedlazeck FJ, Iwaszkiewicz E, Li XY, Altmüller J, Nolte AW. Copy number increases of transposable elements and protein-coding genes in an invasive fish of hybrid origin. *Mol Ecol* 2017;26:4712–24.
- [69] Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2504 human genomes. *Nature* 2015;526:75–81.
- [70] Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* 2017;8:14061.
- [71] Yupeng W, Haibao T, Debarry JD, Xu T, Jingping L, Xiyin W, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 2012;40:e49.
- [72] Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 2018;15:461–8.
- [73] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
- [74] Nattestad M, Schatz MC. Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics* 2016;32:3021–3.
- [75] Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, et al. The Genome Sequence Archive Family: toward explosive data growth and diverse data types. *Genomics Proteomics Bioinformatics* 2021;19:578–83.
- [76] Chen M, Ma Y, Wu S, Zheng X, Kang H, Sang J, et al. Genome Warehouse: a public repository housing genome-scale data. *Genomics Proteomics Bioinformatics* 2021;19:584–9.