# A consensus genome of sika deer (*Cervus nippon*) and transcriptome analysis provided novel insights on the regulation mechanism of transcript factor in antler development

Qianghui Wang[1†], Ruobing Han[1†], Haihua Xing[1] and Heping Li[1*]

## Abstract

**Background** Sika deer (*Cervus nippon*) holds significance among cervids, with three genomes recently published. However, these genomes still contain hundreds of gaps and display significant discrepancies in continuity and accuracy. This poses challenges to functional genomics research and the selection of an appropriate reference genome. Thus, obtaining a high-quality reference genome is imperative to delve into functional genomics effectively.

**Findings** Here we report a high-quality consensus genome of male sika deer. All 34 chromosomes are assembled into single-contig pseudomolecules without any gaps, which is the most complete assembly. The genome size is 2.7G with 23,284 protein-coding genes. Comparative genomics analysis found that the genomes of sika deer and red deer are highly conserved, an approximately 2.4G collinear regions with up to 99% sequence similarity. Meanwhile, we observed the fusion of red deer's Chr23 and Chr4 during evolution, forming sika deer's Chr1. Additionally, we identified 607 transcription factors (TFs) that are involved in the regulation of antler development, including RUNX2, SOX6, SOX8, SOX9, PAX8, SIX2, SIX4, SIX6, SPI1, NFAC1, KLHL8, ZN710, JDP2, and TWST2, based on this consensus reference genome.

**Conclusions** Our results indicated that we acquired a high-quality consensus reference genome. That provided valuable resources for understanding functional genomics. In addition, discovered the genetic basis of sika-red hybrid fertility and identified 607 significant TFs that impact antler development.

**Keywords** Sika deer, Consensus genome, Chromosome evolution, TFs, Antler development

## Introduction

Cervidae is an important ruminant, which consists of two subfamilies, including nineteen genera and fifty-five species [1–3]. The sika deer (*Cervus nippon*), as a significant cervid naturally distributed throughout East Asia, possesses a lot of unique features, such as a special evolutionary status, economic and ecological value. Particularly, sika deer are renowned in East Asian countries such as China, Japan, and South Korea for the production of antlers [4]. Antler as a unique appendage organ of male cervid species (except for the reindeer), it grows extremely fast that exceeds even certain cancer tissues [5–7] and can completely regenerate periodically [8]. Thus, antlers can be to a crucial biological model for mammalian organ regeneration and cancer treatment, it is also a valuable

†Qianghui Wang and Ruobing Han contributed equally to this work.

*Correspondence:
Heping Li
lihepinghrb2002@nefu.edu.cn
[1] College of Wildlife and Protected Area, Northeast Forestry University, Harbin 150040, China

Wang *et al. BMC Genomics*      (2024) 25:617

Page 2 of 13

traditional Chinese medicinal ingredient with significant economic value [4].

A high-quality reference genome can provide significant convenience for in-depth research on the functional genomic. so far, three sika deer genomes have been reported [3, 5, 9], including one female and two male genomes. However, we found that three genomes exhibit substantial differences in continuity and accuracy (Table 1), especially with the recently published genomic comparisons between GCA_034195675.1 (Cni-M ONT) [5] and GWHANOY0000000 (Cni-F 1.0) [3] (Supplementary Fig.S1A, B). Importantly, they might misguide researchers of sika deer. On the other hand, the growth and development of antlers is an exceedingly intricate biological process regulated by numerous factors. In addition to the conventional growth factors such as light, hormones, and so on [8], one of the most critical factors is the genetic, especially the programmed gene expression, which is a pivotal process in the antler development. Recently, lots of studies have revealed the significant impact of transcription factors (TFs) on various life activities, as well as gene expression during cell growth and differentiation [10]. Thus, it is crucial to study the regulatory mechanisms of transcription factors on gene expression in male sika deer based on a high-quality reference genome.

In this study, using a variety of assembly techniques, we present here a high-quality consensus genome (JAYKZF000000000, Cni-M) of male sika deer, offering a reliable source for the genome of sika deer. Based on the high-quality consensus reference genome of sika deer, we conducted a comprehensive investigation into the impact of TFs on the growth and development of sika deer antler. In parallel, we investigated the regulatory mechanisms, providing a novel genomic viewpoint for a detailed study of sika deer antler proliferation and development.

## Results

### A high-quality consensus genome of sika deer assembly and annotation

We assembled a high-quality consensus genome of male sika deer using 143.03G (53–55×) Illumina short reads (from Han et al.2022), 96.39G (36–37×) PacBio hifi longs reads (from Han et al.2022), 161G (57–64×) Oxford Nanopore longs reads (from Qin et al.2023), and 263.03G (99–104×) Hi-C paired-end reads (from Han et al.2022) (Supplementary Table S1) [5, 9]. The final genome size of male sika deer was 2.7G, which was similar to the genome size of the haplotype-resolved sika deer genome and larger than the other two sika deer genomes recently published (Table 1) [3, 5, 9]. Compared to the previous sika deer genome, our consensus genome significantly improved assembly quality. All 34 chromosomes are assembled into single-contig pseudomolecules without any gaps (Fig. 1A), including 34 centromeric regions and 45 telomeres were identified on consensus genome (Supplementary Table S7, Fig.S4). Compared with the haplotype-resolved genome published recently, contig N50 was improved from 49.4 Mb to 89.6 Mb [9] and 13 chromosomes reach to telomere-to-telomere levels (Supplementary Table S7, Fig. 1B). In addition, the high accuracy and completeness was represented by 99.58% mapping rate of Illumina short reads, 4.73e-8 of homozygous SNPs (Supplementary Table S4 and S5), and 98.46% and 97.8% genome completeness as assessed by CEGMA and BUSCO (Fig. 1D), respectively. The assembly quality value (QV) was estimated as 43.85 by k-mer-based approach, exceeding the Vertebrate Genome Project

**Table 1** Summary of sika deer and red deer genome assemblies

| Item | *Cervus nippon* | | | | | *Cervus elaphus* |
|---|---|---|---|---|---|---|
| | Male (consensus genome, This study) | Female (Cni-F 1.0: GWHANOY0000000, Xing et al.,2022) | Male (Cni-M ONT: GCA_034195675.1, Qin et al.,2023) | Male (Hap1, Han et al., 2022) | Male (Hap2, Han et al., 2022) | Female (mCerEla1.1, Pemberton J et al., 2021) |
| Assembly size (Gb) | 2.78 | 2.5 | 2.54 | 2.71 | 2.58 | 2.89 |
| Contig number | 51 | 2,041 | 3,501 | 3,283 | 3,113 | 184 |
| Contig N50 (Mb) | 89.56 | 23.56 | 54.19 | 34.98 | 49.44 | 68.74 |
| Gap-free chromosome number | 34 | 0 | 0 | 0 | 0 | 0 |
| Gaps number | 0 | 1,453 | 1,612 | 913 | 726 | 38 |
| Gene models | 23,284 | 21,449 | 20,281 | 22,144 | 18,705 | 22,941 |
| GC Content (%) | 42.43 | 41.62 | 41.5 | 42.06 | 42.17 | 43 |
| TE proportion (%) | 44.49 | 45.38 | - | 42.39 | 43.06 | - |
| Completeness (% BUSCO) | 97.8 | 93.9 | 89.56 | 94.2 | 93.7 | 97.8 |

Wang *et al. BMC Genomics*    (2024) 25:617

Page 3 of 13

standard of QV40 [11, 12]. Together, these results indicated that our assembly a consensus genome of male sika deer has higher accuracy, completeness, and contiguity.

Annotation using EVidenceModeler combining ab initio prediction, homology to other species, and RNA-seq data identified 23,284 protein-coding genes, which 99% have functional (Supplementary Fig.S5). The number of gene in male sika deer genome is slightly larger than most published ruminants but similar to that of *Rangifer tarandus* (23,233) (Supplementary Fig.S6) [7, 13]. Repetitive sequences account for ~ 44.94% of the genome, while transposable elements (TEs) account for 44.49% with long terminal repeat (LTRs,9.23%) and long interspersed elements (LINEs, 33.95%) comprising the two major retrotransposon classes, the characteristics of repeat sequences were consistent between Cni-M and updated Cni-F 1.0 (Cni-F 2.0) [14] (Supplementary Table S9, Fig. 1E, F). For clarity, the distributions of gene density, repeat density, GC density and collinearity across the chromosomes between Cni-M and Cni-F 2.0 [14] were further illustrated in Fig. 1A.

### Gene family and genome evolution traits between sika deer and red deer

To explore the genetic basis of hybrid fertility between sika deer and red deer, 6 homolog species was selected to gene family clustering analysis. The results showed that 21,938 gene families were identified, including 9,127 shared single-copy gene families. The shared single-copy genes were used for phylogenetic analysis with a maximum-likelihood method (Supplementary Fig.S7). To further research gene family's evolution characteristic of sika deer and red deer, core and pan gene families were identified in sika deer and red deer base on gene family cluster results, and found that sika deer and red deer have 11,931 core gene families (Fig. 2A, B). GO and KEGG enrichment analysis showed that the 11,931 core gene families were significantly involved in the binding (GO:0005488, $P=1.39\times10^{-150}$), ion binding (GO:0043167, $P=3.45\times10^{-119}$), protein binding (GO:0005515, $P=9.83\times10^{-115}$), catalytic activity (GO:0003824, $P=1.16\times10^{-69}$), anion binding (GO:0043168, $P=1.43\times10^{-57}$), transferase activity (GO:0016740, $P=2.83\times10^{-42}$), regulation of cell communication (GO:0010646, $P=3.37\times10^{-16}$), localization (GO:0051179, $P=7.41\times10^{-12}$), kinase activity (GO:0016301, $P=3.64\times10^{-22}$), MAPK signaling pathway ($P=0.0002$), Wnt signaling pathway ($P=0.0009$), Toll-like receptor signaling pathway ($P=0.016$) and other functions necessary to maintain essential life activities (Supplementary Table S12-13).

The results of synteny analysis revealed a highly homologous synteny relationship between the genomes of sika deer and red deer, with an average synteny region size of 2.4G and sequence identity of up to 99% (Supplementary Table S10). It is noteworthy that during the process of species evolution, chromosomes will diverge amongst various species. Our study revealed that chromosome fusion took place during the evolution from red deer to sika deer, with red deer's Chr23 and Chr4 fusing with sika deer's Chr1 (Fig. 2D, E, F). The microcolinearity analysis of red and sika deer (Fig. 2C) corroborated this conclusion as well. It is also consistent with previous reports [9].

### Transcription factors identification and expression analysis

Transcription factors (TFs) are responsible for phenotypic changes by controlling the expression of trait-associated genes [15, 16]. To study the regulation of TFs on the development of antler, we reanalysed the RNA-seq data published by zhang et al. [17]. The hmmsearch program was used to identify TFs in all protein sequences of sika deer based on AnimalTFDB (v2.0) database. A total 607 expressed TFs were identified during the different development stages of sika deer antler (Supplementary Table S15), of which all TFs showed significantly different expression (Supplementary Fig.S8). Subsequently, these TFs were classified into 57 families based on the animalTFDB database. zf-C2H2 (158), Homeobox (64), ZBTB (53), bHLH (51), TF_bZIP (29), HMG (28), Forkhead box (24), ETS (17), MYB (15), and THR-like (11) were possessed more than 10 copies. Notably, the zf-C2H2 exhibited a remarkable 158 copies. Therefore, we hypothesize that transcription factors such as zf-C2H2, Homeobox, and ZBTB play crucial regulatory roles during the antler development. Indeed, zf-C2H2 family has been reported to be associated with development and differentiation of organs/tissues in the early embryonic stages [18, 19]. Furthermore, using 15 days of deer antler development as a control, we conducted a comparative

---

(See figure on next page.)

**Fig. 1** Overview of *Cervus nippon* genome assembly and repeat sequence annotation. **A** Genomic landscape of the sika deer. From outer to inner circles: a, the chromosomes at the Mb scale (red color is Cni-F 2.0 chromosomes, blue color is Cni-M chromosomes); b-d, repeat density, gene density, and GC density across the genome, respectively, drawn in 5 Mb non-overlapping windows; e, collinearity of the chromosome-level genome of Cni-M and Cni-F 2.0. **B** Heatmap of chromosomal telomeric motif density distribution. **C** Hi-C interaction heatmap of the Cni-M. The darker the colour, the stronger the interaction signal, the closer the interaction regions. **D** Comparison of the newly assembled sika deer genome with 8 published deer genomes based on BUSCO and Contig N50. **E**, **F** Classification and substitution level of TEs in the Cni-M and Cni-F 2.0. The x-axis represents the kimura substitution level (CpG adjusted) of TEs. The y-axis represents the percentages of TEs in the genome. The different colours represent different types of transposons
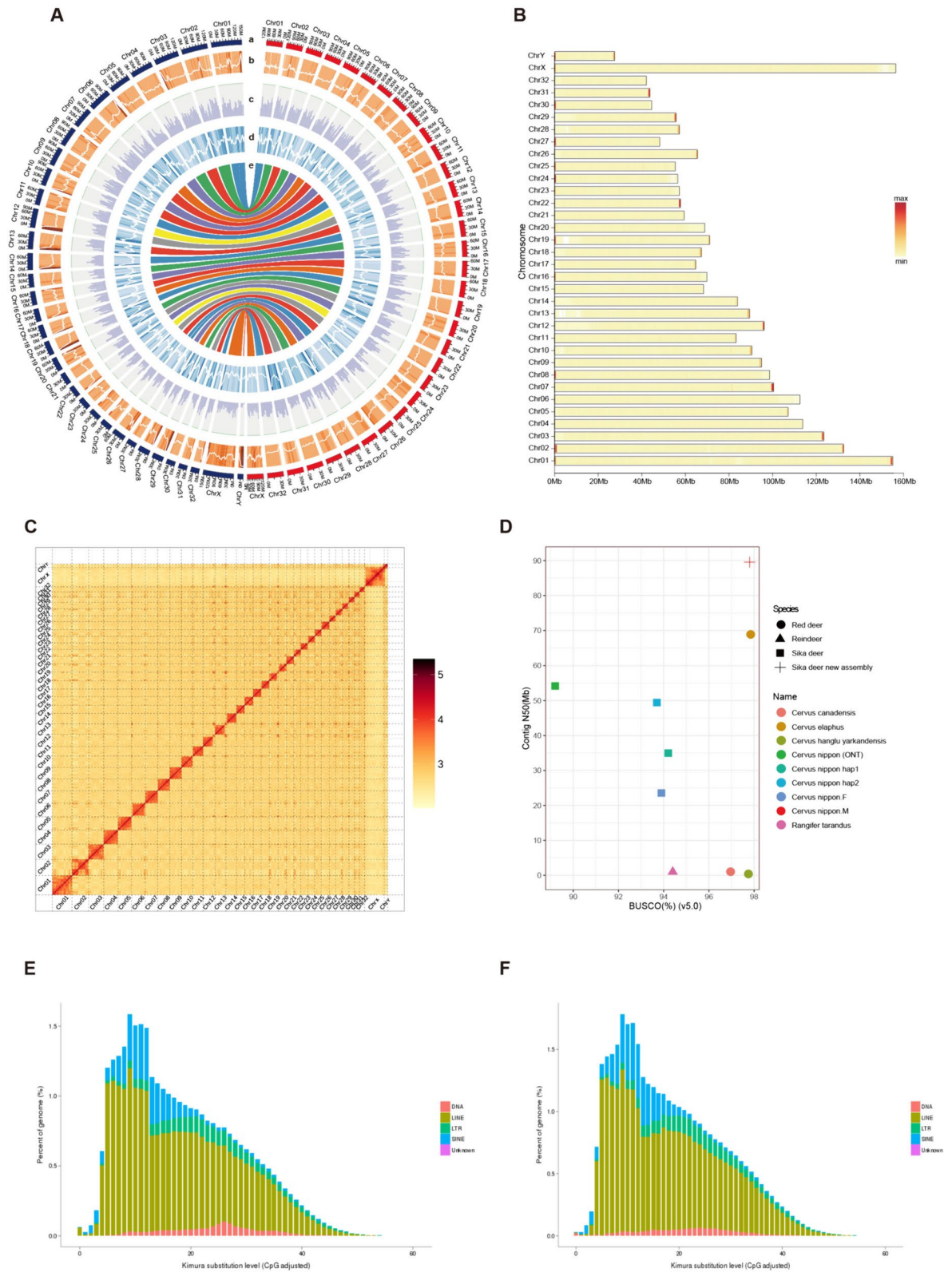
**Fig. 1** (See legend on previous page.)

Wang *et al. BMC Genomics*　　(2024) 25:617

Page 5 of 13
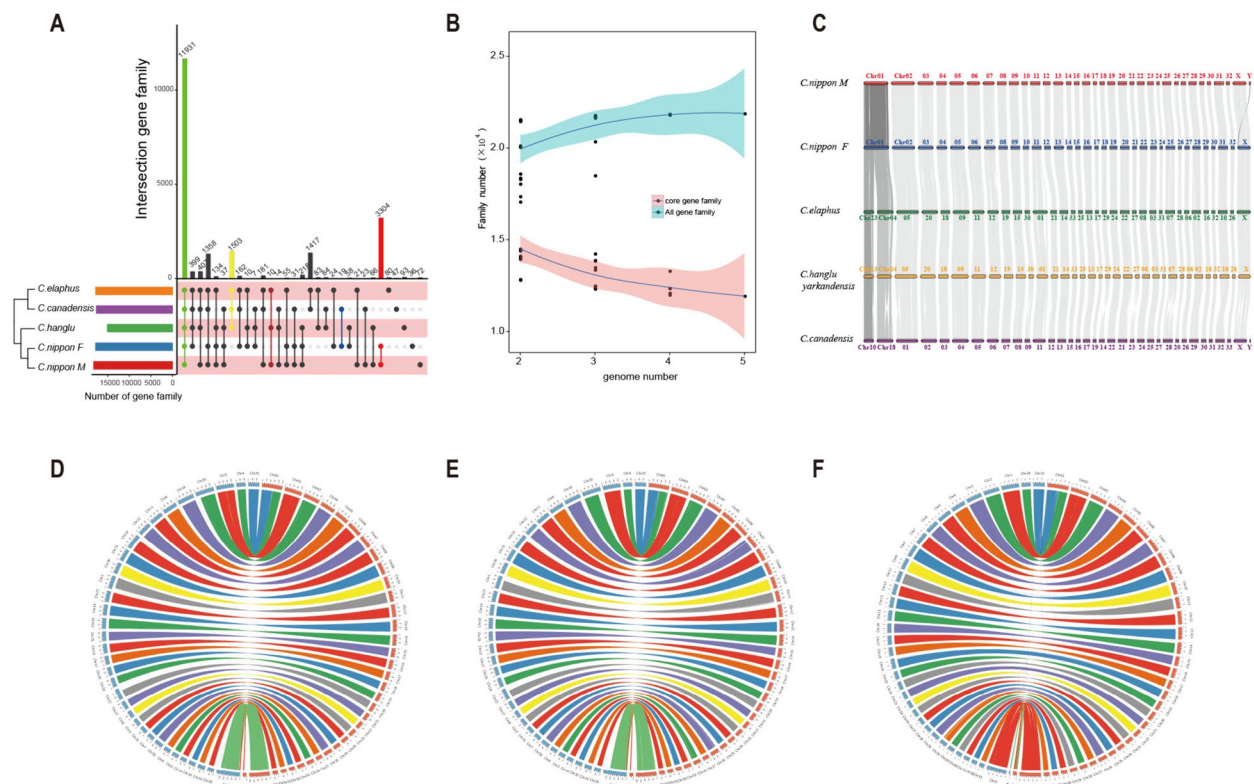


**Fig. 2** Evolution of gene families and chromosome between sika deer and red deer. **A** Sika deer and red deer gene families cluster result. **B** Increasing trend of all gene families and decreasing trend of core gene families as number of deer genomes increases. **C** Microsynteny of the chromosome-level genome of sika deer and red deer. Red deer chromosome 23 and chromosome 4 synteny with sika deer chromosome 1. **D** Circos plots showing genome collinearity between Cni-M and *Cervus elaphus*. **E** Circos plots showing genome collinearity between Cni-M and *Cervus hanglu yarkandensis*. **F** Circos plots showing genome collinearity between Cni-M and *Cervus canadensi*

analysis of the expression levels of transcription factors at different developmental stages in the base, middle, and tip tissues. In the sika deer antler tip tissue, we found that the expression of SIX2, SIX4, SOX6, SOX9, ZHX3 FOXC2 at 25 days are significantly higher than at 15 days, the expression of PAX8, FOXP4 significantly increased at 45 days, the expression of RUNX1, ALX1, JUN, KLHL4 significantly increased at 45 days, the expression of FOSB, SOX15, IKZF2, KLF5 significantly increases starting at 100 days, the expression of TWST2 always lower than 15 days at the pre 100 days of sika deer antler development. In the sika deer antler middle tissue, we found that the expression of SIX1, SOX2, SOX8 and IRX5 at 25, 45, and 65 days are significantly higher than at 15 days. the expression of SATB2 significantly higher than 15 days at pre 100 days of sika deer antler development, the expression of RUNX2 significantly increases starting at 45 days, the expression of HMX1, MUSC and BARX1 significantly increases starting at 130 days, the expression of TWST2 always lower than 15 days at the pre 100 days of sika deer antler development. In the sika deer antler base tissue, we found that the expression of RUNX2, IRX5, SPI1, SATB2,

and KLHL6 significantly increases starting at 25 days, the expression of TWST2 always lower than 15 days at the pre 100 days of sika deer antler development, the expression of TWST1 significantly increases starting at 65 days (Fig. 3).

To further explore the regulatory effects of TFs on gene expression, 6 co-expression modules were clustered by gene expression at different stages of antler development (Supplementary Fig.S9). The blue module contained 3,702 genes, the brown module contained 799 genes, the green module had 73 genes, the turquoise module included 3,844 genes, the yellow module consisted of 513 genes and grey module consisted of 35 genes. Enrichment analysis found that genes in the yellow module significantly involved in ossification (GO:0001503, $P=0.001$), regulation of signal transduction (GO:0009966, $P=0.001$), key signaling pathways or functional pathways related to cell communication (GO:0010646, $P=0.001$), cell differentiation (GO:0030154, $P=0.02$), Osteoclast differentiation (map04380, $P=0.0006$), Rap1 signaling pathway (map04015, $P=0.001$), Regulation of actin cytoskeleton (map04810, $P=0.0009$), and other processes associated

Wang *et al. BMC Genomics*     (2024) 25:617

Page 6 of 13
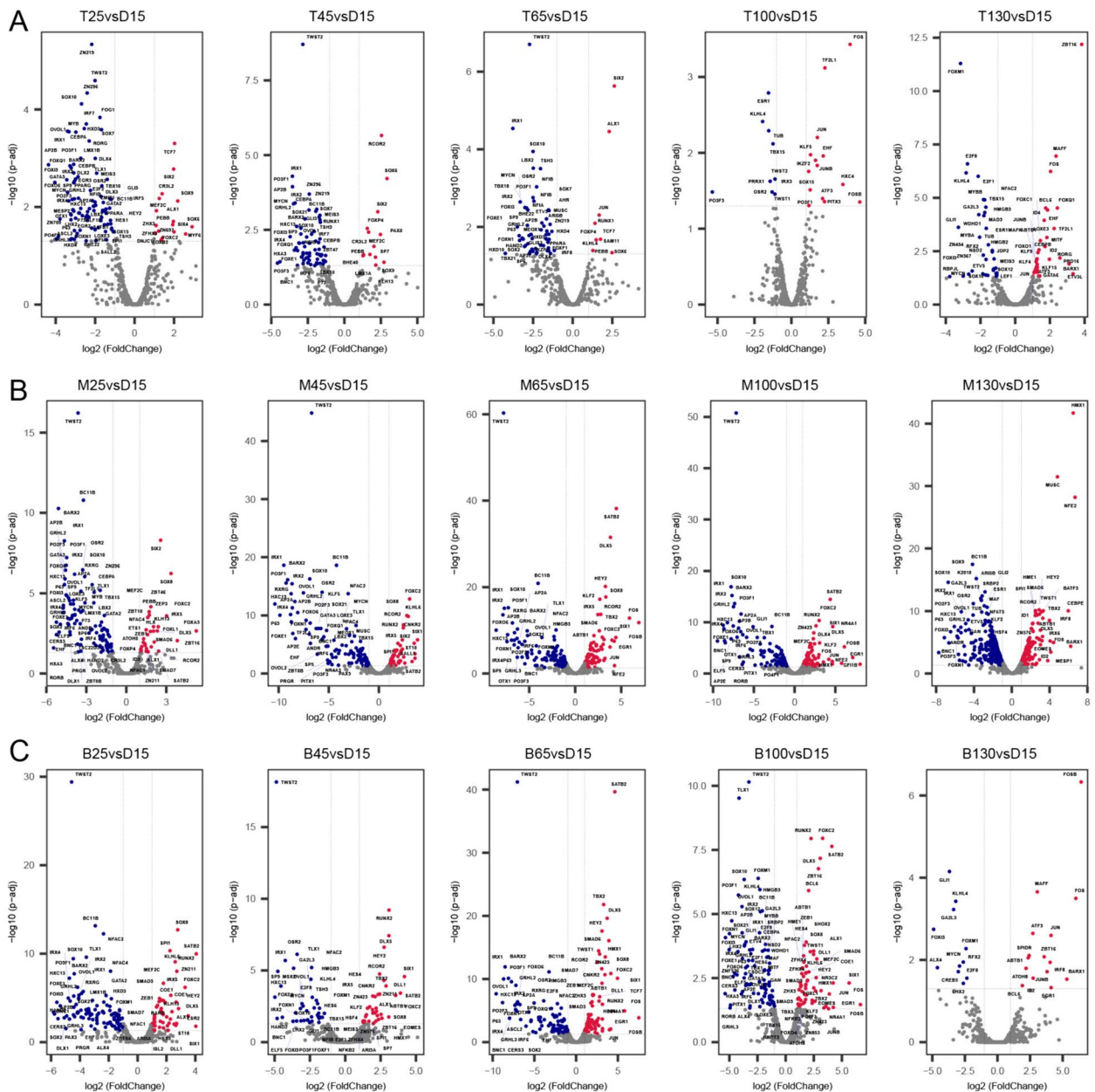


**Fig. 3** Volcano plot of TFs expression at different developmental stages in sika deer antlers. **A** TFs expression differences at different developmental time points in sika deer antler tip tissue compared to 15d. **B** TFs expression differences at different developmental time points in sika deer antler middle tissue compared to 15d. **C** TFs expression differences at different developmental time points in sika deer antler base tissue compared to 15d

with antler growth and development. Meanwhile, 19 TFs, including IRX5, ZBT46, ZEB1, COE1, JDP2, RHBT1, SPI1, SAM11, ZN710, SMAD7, NFAC1, VDR, DDIT3, RUNX2, KLHL8, RARB, GLIS2, and TSC22 were found in the yellow module. Among them, SPI1, JDP2, ZNF170, NFAC1, and KLHL8, recognized as hub genes, as depicted in the co-expression interaction networks (Fig. 4).

## Discussion

In this study, we reported a consensus genome of male sika deer. The results of Contig N50 and BUSCO were both much better than the previously published sika deer genomes [3, 5, 9] and slightly surpassing the recently published red deer genome [20] (Fig. 1D), that exhibits higher continuity and completeness. To acquire a gap-free
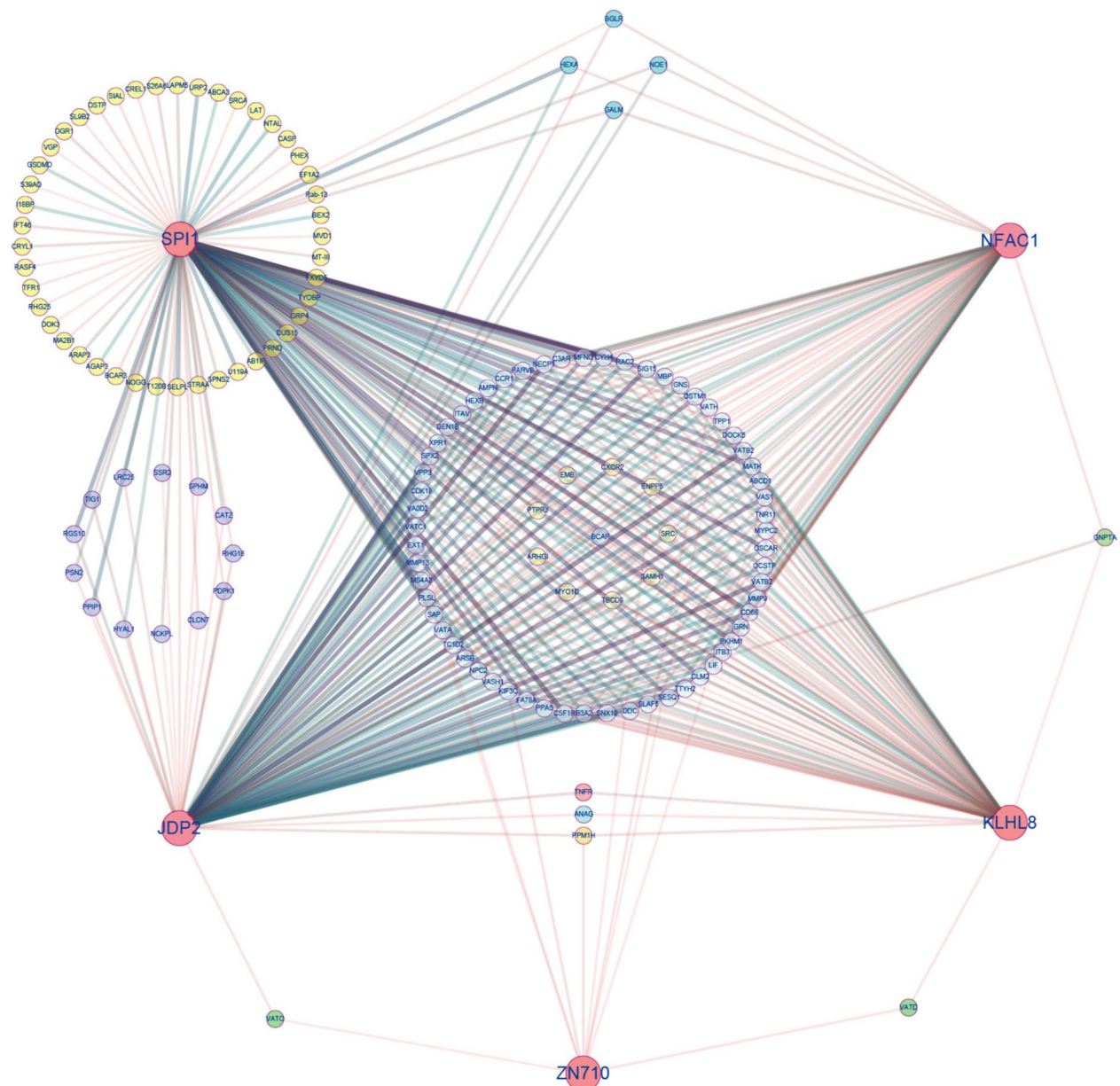
**Fig. 4** Co-expression network diagram of key transcription factors and target genes controlling osteoclast formation and mineralization. Red edge is key TFs in yellow module. Yellow edge is a gene regulated by SPI1. Purple edge is a gene co-regulated by SPI1 and JDP2. Green edge is a gene regulated by two TFs. Light blue edge is a gene co-regulated by SPI1, JDP2, KLHL8 and NEAC1. Sky blue edge is a gene co-regulated by SPI1, JDP2 and NFAC1. The thickness of the lines represents the strength of the interactions between TFs and target genes

consensus genome of sika deer, long reads sequencing data from two different individuals was applied to assembly consensus genome, Hifi long reads was used assembly backbone of male sika deer genome and ONT long reads was employed to fill gaps (Supplement Fig.S2). Actually, 24 contigs had been assembled into chromosomes level from HiFi long reads (Chr04, Chr05, Chr06, Chr08, Chr11, Chr12, Chr13, Chr14, Chr15, Chr16, Chr17, Chr18, Chr19, Chr20, Chr21, Chr22, Chr23, Chr24, Chr25, Chr26, Chr27, Chr29, Chr30, Chr31, Chr32), only 10 chromosomes (Chr01, Chr02, Chr03, Chr07, Chr09, Chr10, Chr28, ChrX, ChrY) had a few filled-gaps. Illumina short reads was mapped to consensus genome, over 99% mapping rate and coverage reach to 98.89% (Supplementary Table S6). Notably, Filled-gap regions was completely covered by short reads (Supplementary Fig.S3) and

Wang *et al. BMC Genomics*     (2024) 25:617

Page 8 of 13

Hi-C interactions same as expected (Fig. 1C), that indicated the reliability of filling gaps. Additionally, the 34 centromeric regions and 45 telomeres were identified in the consensus genome (Supplementary Table S7, Fig.S4). It showed that consensus genome was currently stands as the highest standard in genome assembly within the Cervidae [3, 5, 7, 9, 13]. This holds great significance for deeply analysis of the sika deer's genetic information. Furthermore, we have corrected and reassembled also a female sika deer genome (Cni-F 2.0) [14], primarily rectifying errors associated with chromosomes. This correction eliminates confusion for future researchers, greatly improving work efficiency (Supplementary Fig.S1).

The sika deer and red deer are two entirely different species, exhibiting clearly differences in both their phenotypic characteristic and genomic karyotype. Red deer have 68 chromosomes (2n = 68), while the sika deer have 66 chromosomes (2n = 66) [9]. Interestingly, hybridization between the sika deer and red deer can produce fertile offspring with significantly improved productivity, a characteristic that has captivated breeders of the sika deer. In this study, we specifically focused on the gene families and chromosomal evolution characteristics of the sika deer and red deer. We found that the genomes of the sika deer and red deer are highly conserved during evolution, with collinear regions covering approximately 2.4G and accounting for over 90% of the genome. Moreover, the sequence similarity between collinear regions exceeds 99% (Supplementary Table S10). This suggests that the core conserved regions between the red deer and sika deer are approximately 2.4G, which same to previous genome surveys of both species [3, 7]. This is the genetic basis for the ability of the sika deer to hybridize and produce offspring with the red deer.

As important regulation factors, TFs plays a critical role in the development of several organs, including kidney, skull and stomach [21]. We found that 607 TFs significantly different expression at different development stages of antler. These TFs encompass the regulation of various cellular developmental processes, such as cell proliferation, differentiation, apoptosis, growth inhibition, osteoclast formation and ossification, all of which are associated with the development of deer antlers. Among them, In the middle tissue of deer antlers, RUNX2 expression level starts to increase compared to the 15d stage at 45d, while in the base tissue, its expression level higher than the 15d at 25d. This pattern is in complete accordance with the characteristics of ossification in deer antlers, and it also agree with previous research conclusions regarding the role of RUNX2 in the ossification process of deer antlers [22]. SOX9 and RUNX2 are essential for the differentiation of mesenchymal stem cell-derived osteochondro-progenitors towards chondrogenesis and osteogenesis, respectively, but SOX9 is dominant over RUNX2 function in mesenchymal precursors that are destined for a chondrogenic lineage during endochondral ossification [22]. TWST2 belongs to a class of important transcription factors that inhibit the premature or ectopic differentiation of preosteoblast cells during osteogenesis [23]. After 25d, the expression level of TWST2 consistently remains lower than that at 15d and it increased in 130d. This aligns perfectly with the endpoint of deer antler growth, which ultimately results in ossification and the formation of antlers. Therefore, TWST2 relieves its inhibitory effect on deer antler ossification by gradually increasing its expression. PAX2, SIX4, SIX6, SOX8, SPI1, JDP2, ZNF170, NFAC1 are all important regulatory factors involved in the process of cell proliferation and development [24–28], playing crucial roles in the development of deer antlers as well. Compared to the transcriptome analysis conducted by Zhang et al. [17], our re-analysis the RNA data has yielded novel insights. Firstly, our re-analysis has once again confirmed the reliability of the consensus genome, identifying a total of 24,127 expressed genes (FPKM > 0.1), including 3,126 novel genes (12.96%). In contrast, Zhang et al.'s analysis identified 24,856 expressed genes (FPKM > 0.1), with 8,778 being novel genes (35.32%). This result indicates that we utilized a more comprehensive reference genome and achieved a higher gene annotation rate. Secondly, our focus on transcription factors (TFs) revealed additional TFs involved in antler development beyond SOX9, IRX1 and FOXL2. Notably, SPI1, JDP2, ZNF170 NFAC1 and KLHL8 were newly discovered as important regulators of antler osteoclast formation. Furthermore, a key co-expression network (Fig. 4) was found in our study. Zhang et al. used only DEGs (7,417) in their WGCNA analysis and obtained 15 co-expression modules, whereas our analysis was based on all expressed genes (24,127) and acquired 6 co-expression modules. It is worth noting that the WGCNA guidelines strongly do not recommend using DEGs for co-expression modules analysis. The main reason is that analysis based on DEGs will worsen the stability of the analysis. It is recommended to use all expressed genes for cluster analysis. Hence, it is evident that we have conducted a more comprehensive analysis.

In conclusion, this study has acquired a consensus genome for male sika deer. Through the analysis of gene families and genomic evolution in red deer and sika deer, we have found a clear understanding of the reasons for the fertility of red deer and sika deer hybrids. Based on transcriptome and comparative genomes analysis, we discovered that 607 TFs are involved in the regulation of antler growth and development.

Wang *et al. BMC Genomics*    (2024) 25:617

Page 9 of 13

## Methods

### A high-quality consensus reference genome assembly and quality assessment

To acquired high-quality consensus genome of sika deer, a new assembly strategy was designed by combining Illumina short reads, PacBio HiFi longs reads, Oxford Nanopore longs reads, and Hi-C paired-end reads, those data obtaining from the recently published genomes of two male sika deers [5, 9] (Supplementary Table S1). Firstly, fastp software (v0.20.1) [29] was used to remove low-quality reads and sequencing-adaptor-contaminated reads for Illumina short reads and Hi-C reads, high quality sequencing data were acquired. Then, we adopt three steps assembly a gap-free consensus genome (Supplementary Fig.S2). 1) De novo assembly: for the PacBio assemblies, consensus reads (HiFi reads) were generated using CCS software (https://github.com/pacificbiosciences/unanimity) with the default parameter. These long and highly accurate (>99%) HiFi reads were assembled using Hifiasm (v0.19.1) [30] with default parameters to generate a draft contig genome as contig v1. For the ONT assemblies, Nextdenovo (v2.4.0) (https://github.com/Nextomics/NextDenovo) [31] with default parameters was applied to assemble another draft contig genome as contig v2. Two sets of primary contig genomes were generated. 2) Chromosome levels genome: Hi-C data were used to anchor and remove some short contigs. Hi-C data were classified as valid or invalid interaction pairs using Hicup (v0.8.1) [32], and only valid interaction pairs were retained for subsequent assembly, All-HIC software (v0.9.13) [33] was used to cluster, order, and orient the contigs base on contig v1 and acquired 34 pseudomolecule chromosomes. 3) Gap-free genome assemble: Using the numcer of MUMmer software packages (v4.0) [34], we aligned contig v2 to the preliminary scaffolded chromosome-level genome. Based on the alignment results, we scripted the process to fill the regions containing gaps with sequences from contig v2, resulting in a gap-free reference genome. Subsequently, NextPolish software (v2.4.0) (https://github.com/Nextomics/NextPolish) [31] was employed to perform error correction on the gap-free sequence using HiFi long reads and Illumina short reads. Finally, Hi-C data was used to further improve the genome assembly and detect chromatin interactions, ultimately resulting in the final gap-free reference genome (Supplementary Table S2) and heatmap of genomic interactions was plotted (Fig. 1C) with juicebox software (v2.20.00) [35].

For telomere identification, animal telomeric sequences (TTAGGG/CCCTAA) were identified, and 45 of the expected 68 telomeres (34 chromosomes) were identified using the perl script (Supplementary Table S7). In addition, we applied the quarTeT pipeline [36] (http://www.atcgn.com:8080/quarTeT/home.html) to identify centromere candidate regions, resulting in the identification of centromere candidate regions on all 34 chromosomes (Supplementary Fig. S4).

To evaluate the accuracy of the assembly at the single base level, short Illumina reads were mapped to the male sika deer genome using BWA (v0.7.8) [37] with parameter settings of '-k 32 -w 10 -B 3 -O 11 -E 4', and variant calling was performed using SAMTOOLS (v0.1.19) [38]. Meanwhile, the package Merqury was used for assessing the quality of genome assemblies using a reference-free and k-mer–based approach [10, 11]. Besides, assembly completeness was assessed by using the 248 core genes in the Core Eukaryotic Genes Mapping Approach (CEGMA) [39] and the 3354 vertebrata gene set from OrthoDB 10 and the Benchmarking Universal Single-Copy Orthologs (BUSCO) (v5.4.5) [40] (Supplementary Table S4-6). To validate the accuracy of gap filling, we examined the alignment of Illumina short reads to the gap regions using IGV tools (Integrative Genomics Viewer) to confirm the accuracy of gap filling (Supplementary Fig.S3).

### Genome annotation

A combined method of homologous comparison and de novo prediction were applied to detect the repeated sequences within the Cni-M. RepeatMasker (v3.3.0) [41] and the associated RepeatProteinMask [41] were used for homologous comparison to align against the Repbase database [42] to identify transposable elements. For de novo prediction, LTR_FINDER (v1.0.7) [43], RepeatScout (v1.0.5) [44] and RepeatModeler (v2.1) were first used for de novo candidate database construction of repetitive elements. Subsequently, the genome was soft masked using RepeatMasker using the newly created species-specific repeat libraries mentioned above. Tandem repeat sequences were de novo predicted using Tandem Repeats Finder (v4.0.9) [45].

Gene prediction was performed through a combination of homology-based prediction, de novo prediction and transcriptome-based prediction methods. For homologous annotation, protein sequences including *Cervus elaphus* (GCF_910594005.1), *Cervus hanglu yarkandensis* (GCA_010411085.1), *Cervus nippon* (GWHA-NOY00000000, https://ngdc.cncb.ac.cn/gwh), *Bos taurus* (male, GCF_002263795.1), *Ovis aries* (Oar_v3.1.94) and *Rangifer tarandus* (gigaDB http://gigadb.org/) [13] were aligned against the male Sika deer genome using TBLASTN (v2.2.29) [46]. High-quality blast hits that corresponded to reference proteins were concatenated by Solar software (v0.9.6) [47] after filtering the low-quality records. The genomic sequence of each reference protein was extended upstream and downstream

Wang *et al. BMC Genomics*     (2024) 25:617

Page 10 of 13

by 1000 bp to represent a protein-coding region. Gene-Wise (v2.2.0) [48] software was used to predict the gene structure contained in each protein region. Homology predictions were denoted as "Homology-set". The RNA-seq clean data from sika deer were first de novo assembled using Trinity (v2.0) [49], and the assembled sequences were then aligned against their respective genomes using the PASA pipeline (v2.0.2) [50] with BLAT as the aligner. Gene models created by PASA were denoted as PASA-T-set. We simultaneously employed five tools, Augustus (v3.0.2) [51], GeneID (v1.4) [52], GeneScan (v1.0) [53], GlimmerHMM (v3.0.2) [54], and SNAP [55], for ab initio prediction, in which Augustus, SNAP, and GlimmerHMM were trained by PASA-T-set gene models. In addition, RNA-seq reads were directly mapped to the genomes using Tophat (v2.0.9) [56], and subsequently, the mapped reads were assembled into gene models (Cufflinks-set) by Cufflinks (v2.1.1) [57]. According to these three approaches, all the gene models were finally integrated by EvidenceModeler (v1.1.1) [50]. Weights for each type of evidence were set as follows: PASA-T-set > Homology-set > Cufflinks-set > Augustus > GeneID = SNAP = GlimmerHMM = GeneScan. To obtain the untranslated regions (UTRs) and alternative splicing variation information, PASA2 was used to update the gene. To achieve functional annotation, the predicted protein sequences were aligned against public databases, including SwissProt [58], Gene Ontology [59], NR database (from NCBI), InterPro [60] and KEGG pathway [61]. The InterproScan tool (v4.7) [62] together with the InterPro database were applied to predict protein function based on the conserved protein domains and functional sites. The KEGG pathway and SwissProt databases were mainly mapped by the constructed gene set to identify the best match for each gene.

### Gene family cluster and evolution analysis

Gene families were constructed using the OrthoMCL pipeline (http://orthomcl.org/orthomcl/) with the parameter "-inflation 1.5" [63]. The protein-coding sequences of 6 homolog species, three highest quality genomes in published red deer genomes, including *Cervus elaphus* (female, GCF_910594005.1), *Cervus hanglu yarkandensis* (female, GCA_010411085.1), *Cervus canadensi* (male, GCF_019320065.1), together with our assembly Cni-M, Cni-F 2.0 [14] and *Bos taurus* (male, GCF_002263795.1) for gene family clustering. Only the longest transcripts were selected to represent the gene. Subsequently, the all-against-all search algorithm with a cut-off of 1e-7 was carried out to identify gene orthology relationships between sika deer and other species. The alignments with high-scoring segment pairs were conjoined for each gene pair. To identify orthologous gene pairs, more than 30% coverage of the aligned regions in both orthologous genes was needed. To detailed understanding of the evolutionary characteristics of sika deer and red deer gene families, Perl scripts was used to statistical analysis of shared and unique gene families in sika deer and red deer, resulting in the identification of pan and core gene families. EnrichPipeline [64] was applied to KEGG and GO enrichment analysis for the core gene families.

The phylogenetic relationship of sika deer with other species was reconstructed using the shared single-copy orthologous genes. The protein-coding sequences of genes were aligned by the MUSCLE (v3.8.31) tool with default parameters [65]. Sequences were then concatenated to one supergene sequence for each species and formed a data matrix. Then, phylogenetic analysis was performed using the maximum-likelihood (ML) algorithm in RAxML (v8.0.19) [66] with the GTR + GAMMA substitution model. The best-scoring ML tree was inferred by the rapid BP algorithm and ML searches after performing 1,000 rapid bootstraps.

### Chromosome synteny analysis of sika deer and red deer

A collinearity analysis between sika deer and red deer was conducted using the MUMmer package (v4.0) [34] with the following parameters: nucmer: –mum -maxgap = 500 -mincluster = 100; delta-filter: -l 200 -1 -q; show-coords: -rTH. Furthermore, to identify the synteny block among Cni-M and Cni-F 2.0 [14] and other three red deer, we used MCScan (python version) [67] to search and visualize intragenomic syntenic regions. A homologous synteny block map between sika deer and red deer was plotted by Circos.

### Transcription factor identification and classification

To identify TFs of male sika deer consensus genome, we built HMM profiles using the sequences in AnimalTFDB (v2.0) [68] and applied the hmmsearch program in HMMER (v3.0) [69] to search all the protein sequences of sika deer against the HMM profiles with an E-value of 0.0001 as the cut-off. Then, we assigned the TFs into different families according to AnimalTFDB (v2.0) [68].

### Analysis of transcriptome at different antler development stages

We download 48 samples transcriptome data (Supplementary Table S16) of antler different development stages from NCBI database and fastp software (v0.20.1) [29] was used to acquire high quality reads (Clean data). Clean data were mapped on the reference genome of a male sika deer by Hisat2 (v2.0.4) [70]. HTseq (v0.6.0) [71] was used to calculate read count, and finally, gene expression levels in terms of Gene expression levels were calculated as fragments per kilobase of transcript per million

Wang *et al. BMC Genomics*     (2024) 25:617

Page 11 of 13

mapped fragments (FPKM) were estimated according to the formula "FPKM = (number of reads in gene × $10^9$) / (number of all reads in genes × the gene length)." Differentially expressed genes (DEGs) were defined using DEseq2 (v1.28.1) [72] with a threshold of FDR < 0.05 and | log2 (foldchange) | > 1. Co-expression gene networks were constructed by implementing all gene using the R package WGCNA (v1.63) [73]. KEGG and GO enrichment (Supplementary Table S16-17) analyses of each module in the networks were conducted using EnrichPipeline [64]. Cytoscape (v3.8.0) [74] was employed for the visualization of co-expression networks in the selected modules.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10522-9.

Supplementary Material 1.

Supplementary Material 2.

### Code availability

Data processing was carried out using the protocols and manuals of the relevant bioinformatics software. The scripts required for specific processing steps can be obtained from this link. (https://figshare.com/s/bca41fbb99f125ad390f).

### Authors' contributions

Conceptualization and supervision: H.L., Q.W. Software: Q.W., R.H., and H.X. Investigation: R.H., Q.W. Writing-original draft preparation: Q.W. Writing-review & editing: Q.W., R.H., and H.L. Visualization: Q.W., H.X. Funding acquisition: H.L.

### Availability of data and materials

The *Cervus nippon* genome assembly project has been deposited at DDBJ/ENA/GenBank under the accession JAYKZF000000000. All data generated or analyzed during this study are included in this published article and its supplementary information files. All sequencing Reads in this study are available at the NCBI database (The genomic sequencing and transcriptomic sequencing data details are available in Supplementary Table S1, S14.). Other datasets (genome assembled sequence, official gene sets, gene models, transcriptome assembly) are available at the Figshare database (https://doi.org/10.6084/m9.figshare.23641179.v2) [75, 76].

## Declarations

### Ethics approval and consent to participate

All experimental designs and animal handling were approved by the Institutional Animal Care and Use Committee of Northeast Forestry University, China (2022049).

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

## References

1. IUCN, The IUCN Red List of Threatened Species, Version 2017-3. (2017); https://www.iucn.org
2. Cap H, Aulagnier S, Deleporte P. The phylogeny and behaviour of Cervidae (Ruminantia Pecora). Ethol Ecol Evol. 2002;14(3):199–216. https://doi.org/10.1080/08927014.2002.9522740.
3. Xing X, Ai C, Wang T, 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.
4. 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. https://doi.org/10.1016/j.jep.2012.12.008. (Epub 2012 Dec 12 PMID: 23246455).
5. Qin T, et al. A population of stem cells with strong regenerative potential discovered in deer antlers. Science. 2023;379:840–7. https://doi.org/10.1126/science.add0488.
6. Li C, Zhao H, Liu Z, et al. Deer antler – a novel model for studying organ regeneration in mammals. Int J Biochem Cell Biol. 2014;56:111–22. https://doi.org/10.1016/j.biocel.2014.07.007.
7. Chen L, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science. 2019;364:eaav6202. https://doi.org/10.1126/science.aav6202.
8. A M F, A S B, A J C B, et al. New physiological insights into the phenomena of deer antler: A unique model for skeletal tissue regeneration. J Orthop Transl. 2021;27:57-66. https://doi.org/10.1016/j.jot.2020.10.012.
9. Han R, Han L, Zhao X, et al. Haplotype-resolved Genome of Sika Deer Reveals Allele-specific Gene Expression and Chromosome Evolution. Genomics Proteomics Bioinformatics. 2023;21(3):470-82. https://doi.org/10.1016/j.gpb.2022.11.001.
10. Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526(7571):75–81.
11. Rhie A, Walenz BP, Koren S, et al. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020;21(1):1–27.
12. A reference standard for genome biology. Nat Biotechnol. 2018;36(12):1121. https://doi.org/10.1038/nbt.4318.
13. Li Z, Lin Z, Ba H, et al. Draft genome of the reindeer (Rangifer tarandus). Gigascience. 2017;6(12):gix102.
14. Qianghui, Wang. Refined and Annotated Genome Results for Female Sika Deer. figshare. Dataset. 2023. https://figshare.com/s/9d055902907319be4345.
15. Langer BE, Hiller M. TF forge utilizes large-scale binding site divergence to identify transcriptional regulators involved in phenotypic differences. Nucleic Acids Res. 2019;47(4):e19–e19.
16. Kuang JF, Wu CJ, Guo YF, et al. iphering transcriptional regulators of banana fruit ripening by regulatory network analysis. Plant Biotechnol J. 2021;19(3):477–89.
17. Zhang R, Dong Y, Xing X. Comprehensive transcriptome analysis of sika deer antler using PacBio and Illumina sequencing. Sci Rep. 2022;12:16161.
18. Nakamoto T, Yamagata T, Sakai R, et al. CIZ, a zinc finger protein that interacts with p130 cas and activates the expression of matrix metalloproteinases. Mol Cell Biol. 2000;20(5):1649–58.
19. Bai SW, Herrera-Abreu MT, Rohn JL, et al. Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. BMC Biol. 2011;9(1):1–18.
20. Pemberton J, Johnston SE, Fletcher TJ, et al. The genome sequence of the red deer, Cervus elaphus Linnaeus 1758 [version 1; peer review: 1 approved, 1 approved with reservations]. Wellcome Open Res 2021, 6:336. https://doi.org/10.12688/wellcomeopenres.17493.1.
21. Huilgol D, Venkataramani P, Nandi S, Bhattacharjee S. Transcription factors that govern development and disease: an achilles heel in cancer. Genes (Basel). 2019;10(10):794.
22. Ba H, Wang X, Wang D, et al. Single-cell transcriptome reveals core cell populations and androgen-RXFP2 axis involved in deer antler full regeneration. Cell Regeneration. 2022;11(1):1–19.
23. Lee MS, Lowe G, Flanagan S, Kuchler K, Glackin CA. Human Dermo-1 has attributes similar to twist in early bone development. Bone. 2000;27(5):591–602.
24. Barua M, Stellacci E, Stella L, Weins A, Genovese G, Muto V, Caputo V, Toka HR, Charoonratana VT, Tartaglia M, Pollak MR. Mutations in PAX2 associate with adult-onset FSGS. J Am Soc Nephrol. 2014;25(9):1942–53.

25. Grifone R, Demignon J, Houbron C, Souil E, Niro C, Seller MJ, Hamard G, Maire P. Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. Development. 2005;132(9):2235–49.

26. Schepers GE, Bullejos M, Hosking BM, Koopman P. Cloning and characterisation of the Sry-related transcription factor gene Sox8. Nucleic Acids Res. 2000;28(6):1473–80.

27. Nishiyama C, Masuoka N, Nishiyama M, Ito T, Yamane H, Okumura K, Ogawa H. Evidence against requirement of Ser41 and Ser45 for function of PU.1 – molecular cloning of rat PU.1. FEBS Lett. 2004;572(1–3):57–64.

28. Decker CE, Yang Z, Rimer R, Park-Min KH, Macaubas C, Mellins ED, Novack DV, Faccio R. Tmem178 acts in a novel negative feedback loop targeting NFATc1 to regulate bone mass. Proc Natl Acad Sci U S A. 2015;112(51):15654–9.

29. Chen S, Zhou Y, Chen Y. Jia Gu; fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90. https://doi.org/10.1093/bioinformatics/bty560.

30. Cheng H, Concepcion GT, Feng X, et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18:170–5. https://doi.org/10.1038/s41592-020-01056-5.

31. Hu, J. et al. An efficient error correction and accurate assembly tool for noisy long reads. bioRxiv 2023.03.09.531669. 2023. https://doi.org/10.1101/2023.03.09.531669.

32. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. HiCUP: pipeline for mapping and processing Hi-C data. F1000Res. 2015;20(4):1310. https://doi.org/10.12688/f1000research.7334.1. (PMID: 26835000; PMCID: PMC4706059).

33. Zhang Xingtan, Zhang Shengcheng, Zhao Qian, Ming Ray, Tang Haibao. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nature Plants. 2019;5(8):833–45.

34. Marçais G, et al. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol. 2018;14:1–14.

35. Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. Juicebox.js provides a cloud-based visualization system for hi-C data. Cell Syst. 2018;6(2):256-258.e1. https://doi.org/10.1016/j.cels.2018.01.001.

36. Lin Y, Ye C, Xingzhu Li , Chen Q, Wu Y, Zhang F, Pan R, Zhang S, Chen S, Wang X, Cao S, Wang Y, Yue Y, Liu Y, Junyang Yue J. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. Hortic Res 2023;10(8):uhad127. https://doi.org/10.1093/hr/uhad127

37. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

38. Li H, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.

39. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007;23:1061–7.

40. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO:assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.

41. Bergman CM, Quesneville H. Discovering and detecting transposable elements in genome sequences. Brief Bioinform. 2007;8:382–92.

42. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:4–9.

43. Xu Z, Wang H. LTR-FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:265–8.

44. Price AL, Jones NC, De Pevzner PA. novo identification of repeat families in large genomes. Bioinformatics. 2005;21:351–8.

45. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.

46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

47. Yu XJ, Zheng HK, Wang J, Wang W, Su B. Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. Genomics. 2006;88:745–51.

48. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res. 2004;14:988–95.

49. Grabherr MG, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.

50. Haas BJ, et al. Automated eukaryotic gene structure annotation using evidencemodeler and the program to assemble spliced alignments. Genome Biol. 2008;9:1–22.

51. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005;33:465–7.

52. Guigó R, Knudsen S, Drake N, Smith T. Prediction of gene structure. J Mol Biol. 1992;226:141–57.

53. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol. 1997;268:78–94.

54. Majoros WH, Pertea M, Salzberg SL. TigrScan and glimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20:2878–9.

55. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59.

56. Kim D, et al. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:1–13.

57. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7:562–78.

58. UniProt Consortium T. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2018;46(5):2699. https://doi.org/10.1093/nar/gky092.

59. Carbon S, et al. Expansion of the gene ontology knowledgebase and resources: the gene ontology consortium. Nucleic Acids Res. 2017;45:D331–8.

60. Finn RD, et al. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 2017;45:D190–9.

61. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–62.

62. Jones P, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.

63. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.

64. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1–13.

65. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

66. Stamatakis A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22:2688–90.

67. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49. https://doi.org/10.1093/nar/gkr1293. (Epub 2012 Jan 4. PMID: 22217600; PMCID: PMC3326336).

68. Zhang H, et al. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. Nucleic Acids Res. 2015;43:D76–81.

69. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:29–37.

70. Kim D, Paggi JM, Park C, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.

71. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.

72. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:1–21.

73. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

74. Shi Z, Derow CK, Zhang B. Co-expression module analysis reveals biological processes, genomic gain, and regulatory mechanisms associated with breast cancer progression. BMC Syst Biol. 2010;4:74.

75.  Wang Q. A high-quality reference genome of a male sika deer and annotation results. figshare. 2023. https://figshare.com/s/fff85e8a33 46a9dfc2d6

76.  Qianghui W. The contig level assembly result of male sika deer genome. Figshare. 2023. https://figshare.com/s/74030e97bc2226b 7fd11

## Publisher's Note

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