scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A high-quality Chromosome-level reference genome assembly of white-lipped deer (*Przewalskium albirostris*)

Yuangang Yang[1,5], Bin Li[2,3,4,5], Wei Luo[1], Bo Xu[2,3,4], Peng Luo[1], Tongzuo Zhang[2,4] ✉ & Zhangqiang You[1] ✉

White-lipped deer (*Przewalskium albirostris*) is the sole species in the genus Przewalskium of the Cervidae family. And it is exclusively found in the Qinghai-Tibet Plateau, where it has evolved as a specialized deer on account of the extreme climates and geographical conditions. However, limited research on the genome of this species exists. Here, PacBio and Hi-C sequencing data were constructed to a high-quality Chromosome-level reference genome for the white-lipped deer. The assembly, totaling 2.99 GB, is composed of 34 chromosomes (32 autosomes and X, Y chromosomes), with a scaffold N50 length of 76.18 Mb. Furthermore, it achieved scores of 98.4% and 63.8 in Universal Single-Copy Ortholog (BUSCO) and quality value (QV), respectively. Subsequently, employing the egapx software for genome structural annotation, we identified a total of 21909 protein coding genes, annotating 21859 of them. The high-quality and completeness of our genome sequence and annotation was affirmed by these results. The assembly will significantly contribute to enhance the understanding of the adaptation mechanisms of the white-lipped deer to the extreme environments.

Background & Summary

The Cervidae family includes four subfamilies: Cervinae, Hydropotinae, Muntiacinae, and Odocoileinae, including approximately 16 genera and 56 species in total[1]. And, it trails just behind the bovids (Bovidae) as the second most rich family among ruminant animals. Species in Cervidae family exhibit typical ruminant characteristics, and most members of the family have evolved the unique ability of periodic antler regeneration[2,3]. The members of the Cervidae family are recognized as important species for research in ecology, evolutionary biology, and other fields.

White-lipped deer (*Przewalskium albirostris*) is the sole species in the genus Przewalskium of the Cervidae family, with no subspecies differentiation[4]. It is exclusively inhabiting the high altitudes of the eastern Qinghai-Tibet Plateau, ranging from 3,500 to 5,100 meters[4]. However, poaching and habitat reduction have significantly decreased the distribution range of the white-lipped deer on the Qinghai-Tibet Plateau compared to historical periods, resulting in a fragmented, island-like pattern. Therefore, the deer has been listed as Vulnerable (VU) on the IUCN Red List and as Endangered (EN) on the China Red List of Vertebrates[5,6]. Qinghai-Tibet Plateau, the highest in the world, has created a distinct environment and climate due to its unique geographical position[7]. The white-lipped deer is a specialized species that has evolved under the unique geographical conditions of the Qinghai-Tibet Plateau[8]. High altitude, strong ultraviolet radiation, and a cold, dry climate have driven the white-lipped deer to develop a series of unique survival strategies and physiological mechanisms to adapt to the extreme environment[9,10]. Despite being an ideal model for studying the genetic mechanisms of high elevation adaptation in deer species, research on the white-lipped deer has primarily focused on ecology,

[1]Ecological Security and Protection Key Laboratory of Sichuan Province, Mianyang Normal University, Mianyang, Sichuan Province, China. [2]Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, Qinghai, China. [3]University of Chinese Academy of Sciences, Beijing, 100049, China. [4]Qinghai Provincial Key Laboratory of Animal Ecological Genomics, Xining, 810008, Qinghai, China. [5]These authors contributed equally: Yuangang Yang, Bin Li. ✉e-mail: zhangtz@nwipb.cas.cn; youzq@mtc.edu.cn

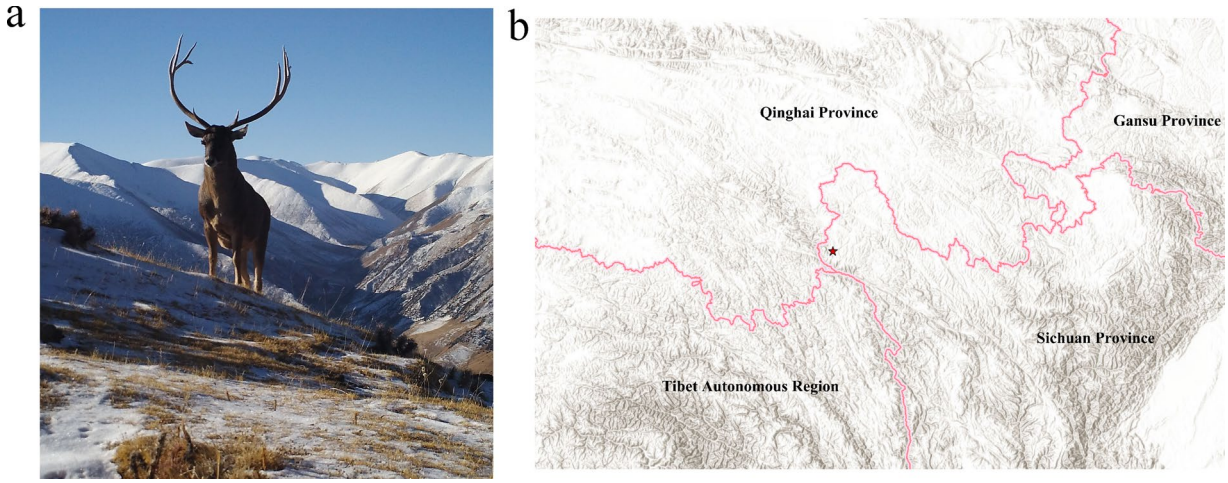

Fig. 1 Photograph and location of the White-lipped deer in present study. (**a**) photograph of an adult male white-lipped deer individual. (**b**) location of white-lipped deer muscle samples (marked with a red pentagram).

behavior, and physiology. A scaffold-level genome assembly published in 2019, includes 171,874 scaffolds and a scaffold N50 of 3.8 Mb, limit the studies of evolutionary on the deer[11].

With the advancement of sequencing technologies, third-generation sequencing technologies (SMRT and Nanopore sequencing) could offer advantages over traditional second-generation sequencing technologies (Illumina sequencing) by generating longer reads[12,13]. It could address challenges in sequence assembly and analysis of structural variations in complex genomes with an effective way[14]. In present study, the third-generation sequencing technology and Hi-C sequencing was utilized to construct a high-quality chromosome-level reference genome of the white-lipped deer with annotated function information. Our data not only establish a reference genome for the white-lipped deer but also offer valuable resources for future research on the molecular mechanisms of adaptation in deer to extreme environments, including facilitating studies on gene expression patterns, identifying genetic variations associated with adaptation, and exploring the evolutionary history of adaptations.

Methods

Sampling and sequencing. Muscle samples were collected from a freshly deceased adult male white-lipped deer (2n = 66) in Zhenda township, Luoxu town, Shiqu County, Sichuan Province, China (32°51′22 N, 97°38′18E; 4278 m) and immediately preserved in a liquid nitrogen tank at −196 °C (Fig. 1). DNA extraction was performed using the QIAamp Blood DNA Midi Kit (Qiagen, Valencia, CA, USA)[15]. The agarose gel electrophoresis and UV spectrophotometer (Thermo Fisher Scientific) were used to determine the quality and quantity of the extracted DNA[16]. Subsequently, a paired-end library with an insertion size of 1500 bp was prepared following the standard Pacbio Template Prep Kit 1.0 protocol (Template Preparation Using BluePippin Size Selection)[17]. Sequencing of the library was conducted using the CCS (Circular Consensus Sequencing) mode on the PacBio Revio sequencing platform, resulting in 107.24 Gb of HiFi long-read data, covering approximately 35.86 × of the genome[17]. Subsequently, following the steps of crosslinking, endonuclease digestion, end repair and biotin labeling, ligation, and enrichment of biotin-labeled fragments, a high-throughput chromatin conformation Hi-C library (with an insert size range of 300 to 700 bp) was prepared using the TruSeq DNA PCR-free prep kit from Illumina. The qualified library was sequenced on the Illumina NovaSeq X Plus platform, producing a total of 286.51 Gb of 150 bp long pass reads, covering approximately 95.82 × of the genome[18].

Genome size and assembly. The HiFi and Hi-C sequencing data were employed to perform contig assembly using hifiasm with default parameters, resulting in a primary assembly and two haplotype assemblies[19]. The primary assembly contained 3.02 Gb, consisting of 2,362 contigs and 42.52 Mb contig N50 (Table 1). The contigs from the primary assembly were subsequently anchored to chromosomes, serving as the chromosomal framework. Specifically, contig indexes were first established using samtools, followed by mapping of Hi-C data to the contigs using chromap[20,21]. The aligned SAM files were then converted to BAM files and BED files using chromap[21]. Following identification of interactions between contigs based on signal strength from the Hi-C data, the contigs were sorted, trimmed, and optimized using yahs with default parameters[22]. Finally, a Hi-C contact matrix was generated using juicer tools and manually corrected using juicer box (Fig. 2)[23]. The results showed a total of 2.99 Gb of data (99%) was mapped to 34 chromosomes (32 autosomes and the X and Y sex chromosomes), with a scaffold N50 of 76.18 Mb (Fig. 3, Table 1). The sex chromosomes (X and Y) of the white-lipped deer were identified by aligning the 34 chromosomes of the assembly with the genome of the closely related species of the wapiti (*Cervus canadensis*), using MCScanX in TBtools-II (Fig. 3)[24,25]. The completeness and quality value (QV) of the assembled genome were assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) and merqury, respectively[26,27]. The BUSCO analysis identified 98.4% complete BUSCOs in the mammalia_odb10 database, comprising 95.5% complete and single-copy BUSCOs, and 2.9% complete and duplicated BUSCOs.

Types	Number of Contig	Contig length (bp)	Number of Scaffold	Scaffold length (bp)
N10	4	82305322	3	115044080
N20	8	66499702	6	103774609
N30	13	56286797	8	99353339
N40	19	49163182	12	81950460
N50	26	42528409	16	76185968
N60	33	34619012	20	62505744
N70	44	23145596	25	54417921
N80	65	9574114	31	44709216
N90	237	796545	158	765213
Total	2362	3016971198	1973	2993228475

Table 1. Assembly statistics of White-lipped deer (*Przewalskium albirostris*).

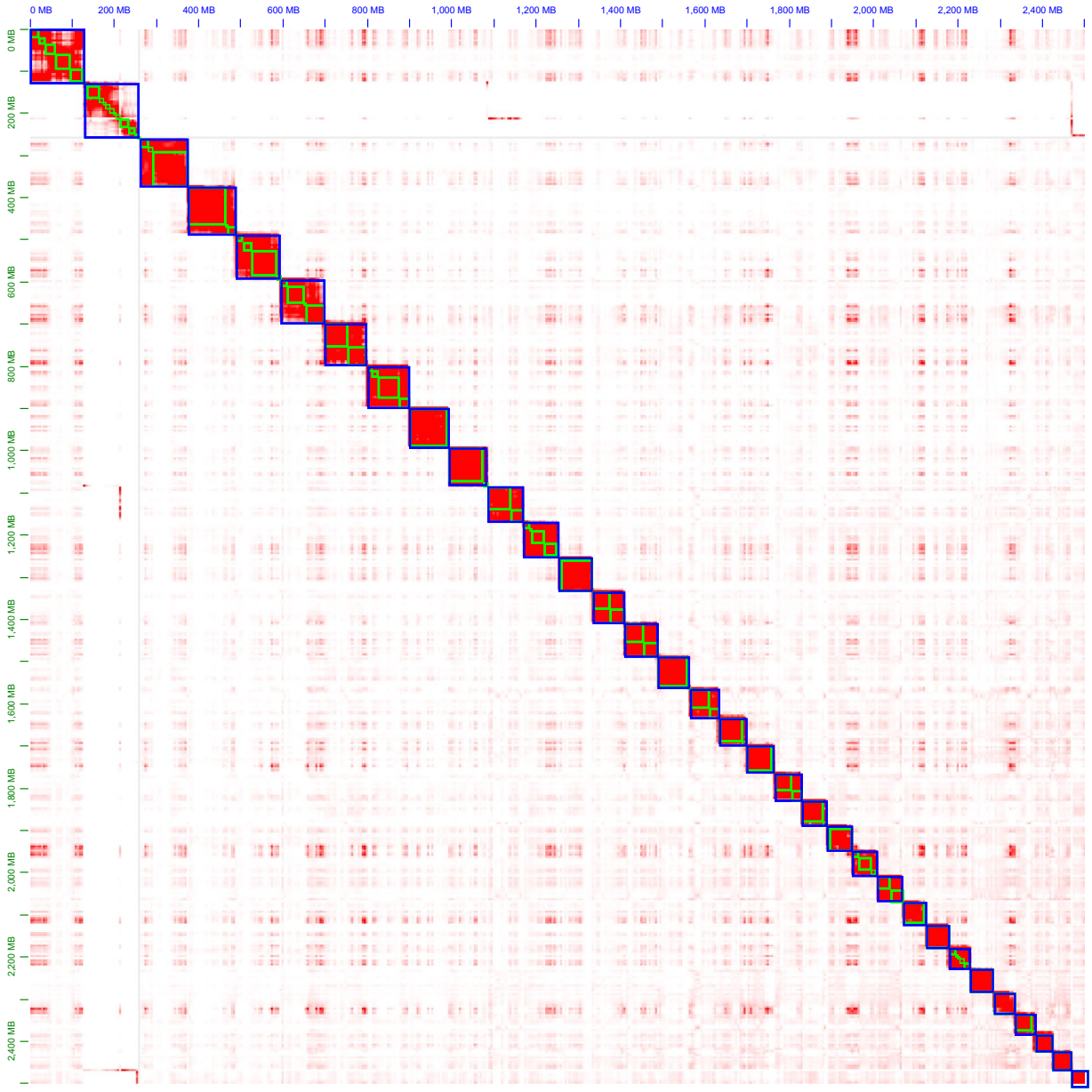

Fig. 2 The heat map of Hi-C contact matrix of the white-lipped deer genome assembly.

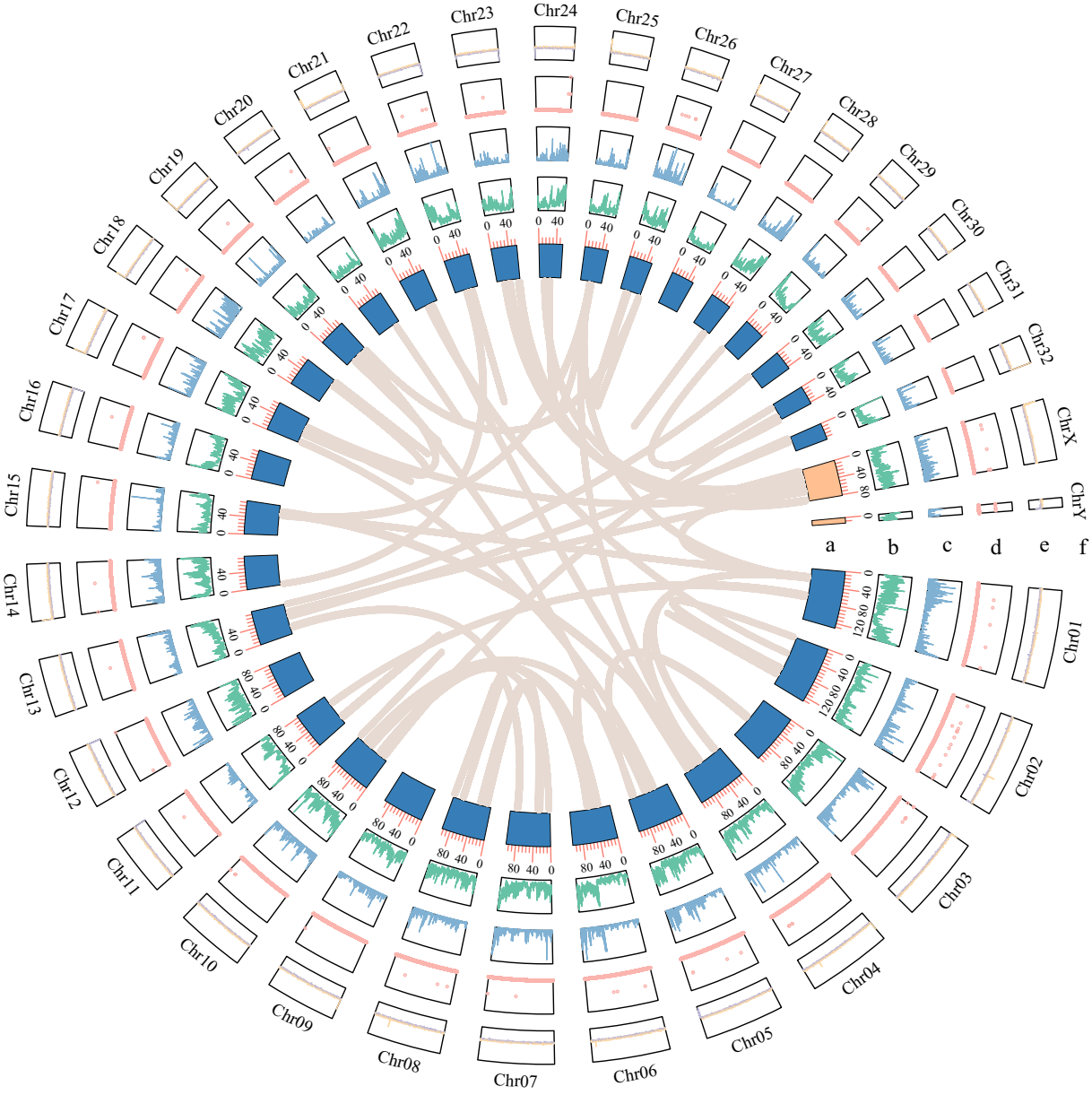

Fig. 3 Overview of the white-lipped deer assembled genome. The syntenic block links between different chromosomes were orange colored. From the inner to the outer layers. (**a**) Chromosome length (Mb); (**b**) line graph of GC content; (**c**) line graph of Gene density distribution within the genome; (**d**) Rate of unknown bases (N-rate distribution dot plot); (**e**) GC skew (line graph showing GC preference); (**f**) Chromosome name.

Additionally, the average quality value (QV) was 63.8, indicating high-quality assembly. Based on these results, we successfully assembled a high-quality chromosome-level reference genome for the white-lipped deer[26,27].

Genome repeats annotation. RepeatModeler and EDTA pipelines were used to generate two de novo libraries of transposable elements (TEs) and repetitive sequences for the white-lipped deer genome with default settings[28,29]. Subsequently, RepeatMasker was used to mask the repetitive sequences in the genome following the merging of the both TE libraries[30]. A total of 1.43 Gb repetitive sequences were annotated, accounting for approximately 47.36% of the entire genome. Specifically, the DNA transposons, Long Interspersed Nuclear Elements (LINEs), Long Terminal Repeat (LTR) elements, and Short Interspersed Nuclear Elements (SINEs) account for 26.74%, 12.56%, 5.08%, and 0.57% of the genome, respectively (Table 2).

Genome structure prediction and functional annotation. After masking the repetitive sequences in the genome, protein-coding genes were predicted using the NCBI eukaryotic genome annotation pipeline (egapx)[31]. Transcriptome data of sika deer (*Cervus nippon*) were download from the NCBI database[32]. Egapx was used to predict protein-coding genes in the masked genome of the white-lipped deer based on the transcriptome data[33]. A total of 21909 protein-coding genes were identified, with BUSCO results showing that 99.6%

Class	Subclass	Number	Total length (bp)	Percent of genome (%)
SINEs:		158081	17318785	0.57
LINEs:		991880	380447384	12.56
	L2/CR1/Rex	78954	10885445	0.36
	RTE/Bov-B	499893	205077646	6.77
	L1/CIN4	413033	164484293	5.43
LTR:		845057	154015442	5.08
	Ty1/Copia	1646	662512	0.02
	Gypsy/DIRS1	49424	20085418	0.66
	Retroviral	200294	47022436	1.55
DNA transposons:		5683845	810297825	26.74
	hobo-Activator	110322	15644727	0.52
	Tc1-IS630-Pogo	49390	9351464	0.31
	MULE-MuDR	272	81099	0.003
	PiggyBac	812	119704	0.004
Unclassified:		93436	71287581	2.35
Total interspersed repeats:			1433367017	47.31
Satellites:		46	350867	0.01
Simple repeats:		9064	1098174	0.04
Total bases masked:			1434816058	47.36

Table 2. Statistics of repetitive elements.

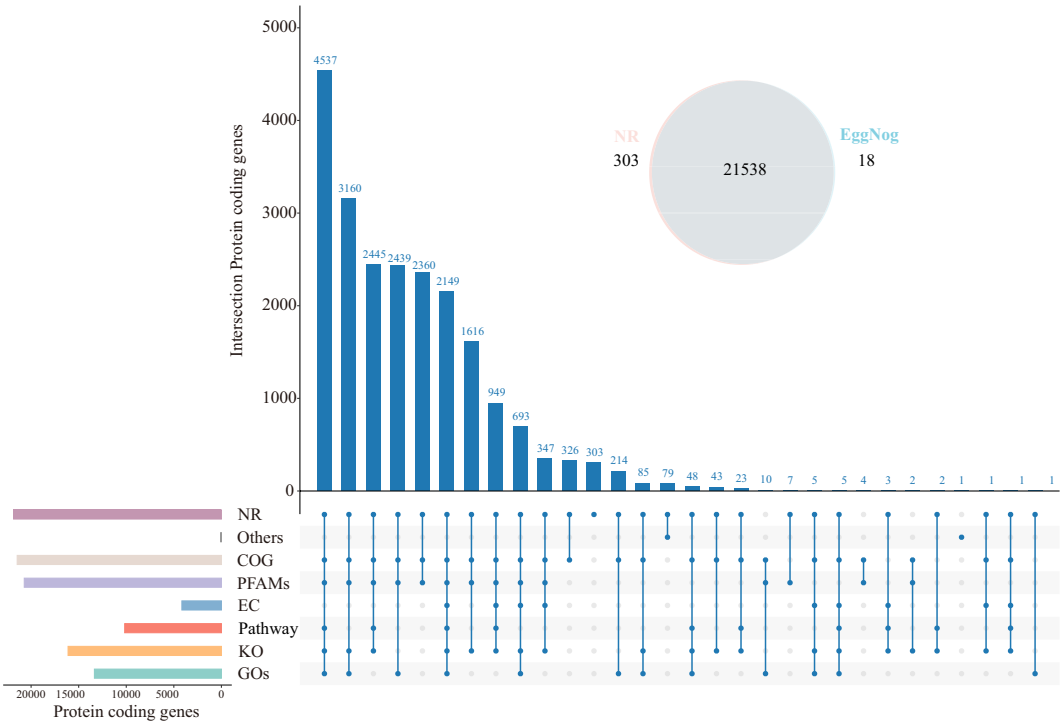

Fig. 4 Statistics of the functional annotation of the protein-coding genes in the genome of the white-lipped deer. Evolutionary genealogy of genes: Non-supervised Orthologous Groups (EggNOG) database (GOs means Gene Ontology database; KO means KEGG orthology; Pathway means KEGG Pathway database; EC means enzyme commission numbers; PFAMs means protein families; COG means clusters of orthologous groups; Others means protein-coding genes have been annotated to the EggNOG database but have not been classified into the sub-databases.) and Non-Redundant protein sequence (NR) database.

were complete BUSCOs (45.9% complete and single-copy, 53.7% complete and duplicated), while the remaining 0.1% were fragmented BUSCOs and 0.3% missing BUSCOs[26]. Subsequently, the functions of these coding genes were annotated using EggNOG-mapper pipeline and the Evolutionary genealogy of genes: Non-supervised Orthologous Groups (EggNOG) database (http://eggnog5.embl.de/#/app/downloads, Mammalia, 40674)[34,35]. The longest characterized transcripts were translated into protein sequences using TBtools-II, and these protein

Species	Sequencing/Assembly Technologies	Genome Size (ungaped) (Gb)	Scaffold N50(Mb)	Number of scaffolds
White-lipped deer (*Przewalskium albirostris*)	PacBio/hifiasm	3.02 (2.99)	76.18	1973
white-lipped deer (*Przewalskium albirostris*)	Illumina HiSeq/Platanus	2.7 (2.6)	3.8	171,874
white-tailed deer (*Odocoileus virginianus*)	PacBio Sequel II/WTDBG2	2.4 (2.4)	52.5	506
Red deer (*Cervus elaphus*)	PacBio/hifiasm	2.9 (2.9)	83.5	144
Père David's Deer (*Elaphurus davidianus*)	Illumina HiSeq/ SOAPdenovo	2.5 (2.5)	3	46,350
Hog deer (*Axis porcinus*)	Illumina HiSeq/ SOAPdenovo	2.7 (2.6)	20.8	136,093
Mule deer (*Odocoileus hemionus*)	Illumina HiSeq/BWA	2.3 (2.3)	0.83	16,207
Reeves'Muntjac (*Muntiacus reevesi*)	PacBio/Arima2	2.7 (2.7)	114.5	267
Red Muntjac (*Muntiacus muntjak*)	Illumina HiSeq/ SOAPdenovo	2.7 (2.6)	1.4	170,618
Chinese water deer (*Hydropotes inermis*)	Illumina HiSeq/Supernova	2.5 (2.5)	75	1712
Black muntjac (*Muntiacus crinifrons*)	Pacbio/Falcon	2.5 (2.5)	632.1	5860
Eurasian elk (*Alces alces*)	Illumina HiSeq/AllPaths	2.5 (2.4)	76.7	5939
Yarkand deer (*Cervus hanglu yarkandensis*)	Illumina NextSeq/Supernova	2.6 (2.6)	77.7	17,927
Sika deer (*Cervus nippon*)	PacBio/hifiasm	3.4 (3.4)	67.2	7,559

Table 3. Genomic statistics of the ten ruminants for gene structure prediction.

sequences were mapped to various sub-databases in the EggNOG database using the blsatp method in diamond with default parameters[24,36]. Results showed that 21556 protein-coding genes matched the EggNOG database, while 303 of the remaining 353 genes matched the Non-Redundant Protein (NR) database (Fig. 4)[37]. Overall, the functions of 21859 protein-coding genes (99.77%) were successfully annotated.

Data Records

The genome assembly and genomic sequencing data including, Hi-C reads by Illumina NovaSeq X Plus and PacBio HiFi reads by PacBio Revio were deposited at the Sequence Read Archive database of NCBI with accession numbers SRR29234840[38] and SRR29234841[39]. The genome assembly have been deposited at the GenBank database under the accession number GCA_046563475.1[40]. The repeats sequences, annotation results of gene structure, longest characterized coding sequences, protein sequences, functional prediction and run scripts were deposited at the Figshare database[41].

Technical Validation

The quality of DNA extraction was assessed using 0.8% agarose gel electrophoresis, and the concentration was quantified using a UV spectrophotometer in the extracting of total DNA[16]. For constructing the HiFi sequencing library, the library was first evaluated using the Agilent Bioanalyzer 2100 to determine the starting size for fragment selection[41]. Subsequently, the final library was quantified and checked for quality using Qubit and the Agilent Bioanalyzer 2100[42]. For constructing the Hi-C sequencing library, the library was first quality-checked using the Agilent High Sensitivity DNA Kit on the Agilent Bioanalyzer[42]. The library was then quantified using the Quant-iT PicoGreen dsDNA Assay Kit on the Promega QuantiFluor fluorescence quantification system[43]. Qualified libraries were then subjected to sequencing.

To evaluate the completeness of our assembly, the BUSCO and merqury were used to assess completeness and accuracy[26,27]. The presence of 98.4% complete BUSCOs and an average QV of 63.8 indicate a high level of integrity and quality in the genome assembly. Subsequently, we used a NCBI eukaryotic genome annotation pipeline homologous prediction method to predict protein-coding genes in the white-lipped deer genome, and the annotation results were also evaluated using BUSCO. The presence of 99.6% complete BUSCOs indicates good prediction of gene structures. In the KEGG functional annotation, a total of 21909 protein-coding genes were successfully annotated, accounting for 99.77% of the total protein-coding genes. These results strongly indicated that the annotated gene set of the white-lipped deer genome is fairly complete.

Our genome assembly has significantly improved in terms of contig and scaffold length by comparing the assembly metrics previous genome version of the white-lipped deer and other deer species with the quality assessment of the assembled genome results (Table 3)[11,44–55]. The assembly length and gap statistics suggest that our genome assembly is of high quality, enhancing our understanding of the white-lipped deer genome structure and function, and providing a reliable foundation for further biological research and applications.

Code availability

No specific code was developed in present study. Data analysis was conducted following the manuals and protocols provided by the developers of the respective bioinformatics tools, which are detailed in the "Methods" section. The versions of the software utilized are indicated below.
hifiasm v0.19.9: default.
samtools v1.20: default.
chromap v0.2.6: default.
yahs v1.2a.2: default.
juicer tools v1.22.01: default.
juicer box v 1.11.08: default.

TBtools-II: default.
BUSCO v5.7.1: default.
Merqury v1.3: default.
RepeatModeler v2.0.5: default.
EDTA v2.2.0: default.
RepeatMasker v4.1.5: default.
GeMoMa v1.9: default.
EggNOG-mapper v2.1.12: default.
diamond v2.1.9: default.

Received: 5 June 2024; Accepted: 11 March 2025;
Published online: 01 May 2025

References

1. Pitra, C., Fickel, J., Meijaard, E. & Groves, C. Evolution and phylogeny of old world deer. *Molecular Phylogenetics and Evolution* **33**, 880–895 (2004).
2. Slate, J. *et al*. A Deer (*Subfamily Cervinae*) Genetic Linkage Map and the Evolution of Ruminant Genomes. *Genetics* **160**, 1587–1597 (2002).
3. Rössner, G. E., Costeur, L. & Scheyer, T. M. Antiquity and fundamental processes of the antler cycle in Cervidae (Mammalia). *The Science of Nature* **108** (2020).
4. Li, B. *et al*. Captivity Shifts Gut Microbiota Communities in White-Lipped Deer (*Cervus albirostris*). *Animals* **12**, 431 (2022).
5. Kaji, K., Ohtaishi, N., Miura, S. & Wu, J. Distribution and status of White-lipped Deer (*Cervus albirostris*) in the Qinghai-Xizang (Tibet) Plateau, China. *Mammal Review* **19**, 35–44 (1989).
6. Hilton-Taylor, C. *2000 IUCN Red List of Threatened Species*. (IUCN, 2000).
7. Qiu, J. China: The third pole. *Nature* **454**, 393–396 (2008).
8. Jiang, Z. *et al*. Diversity and endemism of ungulates on the Qinghai-Tibetan Plateau: Evolution and conservation. *Biodiversity Science* **26**, 158–170 (2018).
9. Mao, K., Wang, Y. & Liu, J. Evolutionary origin of species diversity on the Qinghai–Tibet Plateau. *Journal of Systematics and Evolution* **59**, 1142–1158 (2021).
10. Liu, K., Yang, J. & Yuan, H. Recent progress in research on the gut microbiota and highland adaptation on the Qinghai-Tibet Plateau. *Journal of Evolutionary Biology* **34**, 1514–1530 (2021).
11. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_006408465.1 (2019).
12. Roberts, R. J., Carneiro, M. O. & Schatz, M. C. The advantages of SMRT sequencing. *Genome Biology* **14** (2013).
13. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nature Biotechnology* **34**, 518–524 (2016).
14. Bleidorn, C. Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity* **14**, 1–8 (2015).
15. A.S., P. *et al*. *Genomic DNA Extraction Protocol Using DNeasy Blood & Tissue Kit (QIAGEN) Optimized for Gram-Negative Bacteria V1*. https://doi.org/10.17504/protocols.io.paadiae (2018).
16. Khetan, D., Gupta, N., Chaudhary, R. & Shukla, J. Comparison of UV spectrometry and fluorometry-based methods for quantification of cell-free DNA in red cell components. *Asian Journal of Transfusion Science* **13**, 95 (2019).
17. Rhoads, A. & Au, K. F. PacBio Sequencing and its Applications. *Genomics, Proteomics & Bioinformatics* **13**, 278–289 (2015).
18. Eisenstein, M. Illumina faces short-read rivals. *Nature Biotechnology* **41**, 3–5 (2023).
19. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170–175 (2021).
20. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
21. Zhang, H. *et al*. Fast alignment and preprocessing of chromatin profiles with Chromap. *Nature Communications* **12** (2021).
22. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39** (2022).
23. Yardımcı, G. G. & Noble, W. S. Software tools for visualizing Hi-C data. *Genome Biology* **18** (2017).
24. Chen, C. *et al*. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Molecular Plant* **16**, 1733–1742 (2023).
25. Wang, Y. *et al*. Detection of colinear blocks and synteny and evolutionary analyses based on utilization of MCScanX. *Nature Protocols* **19**, 2206–2229 (2024).
26. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
27. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* **21** (2020).
28. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451–9457 (2020).
29. Ou, S. *et al*. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20** (2019).
30. Chen, N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Current Protocols in Bioinformatics* **5** (2004).
31. Thibaud-Nissen, F. *et al*. P8008 The NCBI Eukaryotic Genome Annotation Pipeline. *Journal of Animal Science* **94**, 184–184 (2016).
32. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra: SRP135837 (2018).
33. Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O. & Grau, J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* **19** (2018).
34. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825–5829 (2021).
35. Huerta-Cepas, J. *et al*. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309–D314 (2018).
36. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* **18**, 366–368 (2021).
37. Pruitt, K. D. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **33**, D501–D504 (2004).
38. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra: SRR29234840 (2025).
39. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra: SRR29234841 (2025).
40. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_046563475.1 (2025).
41. Li, B. A high-quality Chromosome-level reference genome assembly of white-lipped deer (*Przewalskium albirostris*). *figshare* https://doi.org/10.6084/m9.figshare.25971784.v2 (2024).

42. Panaro, N. J. *et al*. Evaluation of DNA Fragment Sizing and Quantification by the Agilent 2100 Bioanalyzer. *Clinical Chemistry* **46**, 1851–1853 (2000).
43. Anantanawat, K., Pitsch, N., Fromont, C. & Janitz, C. High-Throughput Quant-iT PicoGreen Assay using an Automated Liquid Handling System. *BioTechniques* **66**, 290–294 (2019).
44. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_023699985.2 (2023).
45. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_910594005.1 (2021).
46. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_021018665.1 (2021).
47. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_003798545.1 (2018).
48. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_004115125.1 (2019).
49. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_963930625.1 (2024).
50. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_008931205.1 (2019).
51. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_020226075.1 (2021).
52. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_020226055.1 (2021).
53. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_015832495.2 (2023).
54. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_010411085.1 (2020).
55. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_038088365.1 (2024).

Acknowledgements

This research was supported by the Scientific research initiation project (QD2023A02 and QD2023A01), Mianyang Science and Technology Program(2023ZYDF076), and Natural Science Foundation of Qinghai Province (2023-ZJ-901T).

Author contributions

Tongzuo Zhang and Zhangqiang You designed and supervised the project, revised the manuscript, and access fundings; Bin Li performed bioinformatics analysis and wrote the manuscript; Yuangang Yang collected the samples and wrote the manuscript; Wei Luo, Bo Xu, and Peng Luo extracted DNA and visualized pictures. All authors have read, revised, and approved the final manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.Z. or Z.Y.

Reprints and permissions information is available at www.nature.com/reprints.

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