

Letter to the editor

Open Access

High-quality chromosome-level genome assembly of Tibetan fox (*Vulpes ferrilata*)

DEAR EDITOR,

Using Oxford Nanopore and Hi-C sequencing technology, we successfully assembled a chromosome-level genome of the Tibetan fox (*Vulpes ferrilata*), with a total size of 2.38 Gb and N50 length of 133 960 477 bp. The 157 contigs were further assembled into 18 chromosomes with a sequence length of 2 378.42 Mb, accounting for 99.95% of the total length. A total of 21 715 protein-coding genes were predicted in the assembled genome, 86.47% of which were functionally annotated. Phylogenetic analysis showed that *V. ferrilata* and the red fox (*V. vulpes*) formed a clade, with an estimated divergence time of 3.27 million years ago (Ma). Significantly enriched pathways and Gene Ontology terms associated with the expanded gene families in the *V. ferrilata* genome were mainly related to hypoxia response and energy metabolism, indicating the mechanistic strategy of *V. ferrilata* for high-altitude adaptation. Furthermore, selection signature analysis identified genes associated with DNA damage repair and angiogenesis in *V. ferrilata*. Construction of the *V. ferrilata* genome provides valuable information for further genetic analysis of important biological processes, which will facilitate the study of genetic changes during evolution.

The Tibetan fox, which belongs to the family Canidae and order Carnivora, is widely distributed in the northern Qinghai-Xizang (Tibet) Plateau (QTP) at altitudes of more than 3 500 m a.s.l. (Clark et al., 2008). As a plateau-endemic species, it plays an important role in maintaining ecological balance (Harris et al., 2014). Since the 1990s, concerted conservation efforts have led to the rapid increase in *V. ferrilata* populations, which have played a vital role in preventing and controlling local rodent pests (Liu, 2013). Despite this, various questions remain to be clarified. For example, the divergence time between *V. ferrilata* and the red fox (*V. vulpes*) remains controversial (ranging from 2.43 to 4.74 Ma) (Fritz et al., 2009;

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2022 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Humphreys & Barraclough, 2014) and whether *V. ferrilata* has evolved a unique plateau adaptation mechanism that differs from other species is unclear. The assembly of high-quality genomes should help clarify the adaptation mechanisms of high-altitude animals and provide a useful resource for future research. Thus, we constructed a high-quality genome assembly of *V. ferrilata* by combining next-generation sequencing (NGS) short-read, Nanopore long-read, and Hi-C read sequencing. We compared the genomic features of *V. ferrilata* with those of 10 other species, focusing on hypoxia response- and energy metabolism-related gene families, which may contribute to high-altitude adaptation and resistance to hypoxic and low-temperature conditions.

We collected an adult male *V. ferrilata* sample for sequencing from Gande County, Golog Tibetan Autonomous Prefecture, Qinghai Province (N33°54'1", E99°48'54"). Permits for sample collection and use are provided in the Supplementary Materials. Leg muscle tissue was collected for DNA extraction and heart, liver, lung, gut, testis, and kidney tissues were collected for total RNA extraction. In this research, different strategies were used to obtain different omics data. For long-read genome sequencing (PromethION, library size: 20 kb), we obtained ~253.01 Gb of Nanopore reads (read number: 15 437 356; mean read length: ~16.39 kb; N50: ~23 kb) after base calling using Guppy v3.2.2 (-c dna_r9.4.1_450bps_fast.cfg). The MGISEQ2000 platform (library size: 400 bp) was used to obtain the NGS data. After quality control using fastp 0.20.0 (-n 0 -f 5 -F 5 -t 5 -T 5), 145 Gb of NGS short-insert reads were obtained (Supplementary Tables S1, S2). We also obtained ~264 Gb of Hi-C reads from the Illumina NovaSeq 6000 platform (pair-ended sequencing with a read length of 150 bases and library size of 350 bp). For transcriptome data, we obtained ~46.03 Gb of clean data (Supplementary Table S3). Additional details are provided in the Supplementary Materials and Methods. The 17-mer frequency distribution analysis was performed using Jellyfish v2.2.10, which revealed that the estimated genome size of *V. ferrilata* was ~2.3 Gb (Supplementary Figure S1). We also

Received: 31 December 2021; Accepted: 29 March 2022; Online: 31 March 2022

Foundation items: This work was supported by the National Natural Science Foundation of China (31872242, 32070405)

obtained a preliminary ~2.35 Gb genome assembly (contig N50: 61 Mb) using NextDenovo v2.0-beta.1 (reads_cutoff:1k, seed_cutoff:25k). Data correction was performed using Racon v1.3.1 and NextPolish v1.0.5, and the final polished genome assembly was ~2.38 Gb with a contig N50 of ~61.59 Mb (Supplementary Table S4). Genome completeness and accuracy were evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v4.0.5 (-l mammalia_odb10 -g genome), with 92.76% complete BUSCOs (8 558) identified (Supplementary Table S5). GC depth analysis was used to assess the presence of exogenous contamination in the genome. Results revealed a genome coverage depth of 102.73 X and average GC content of 41.25% (Supplementary Figure S2). The single nucleotide polymorphism (SNP) and indel tests identified 8 282 homozygous SNPs, accounting for 0.000348% of the genome (depth>=5X). The number of homozygous indels was 29 587, accounting for 0.001243% of the genome (depth>=5X). Thus, the genome showed 99.998409% base accuracy (depth>=5X) (Supplementary Table S6). Overall, the assessment results indicated high accuracy and completeness of the *V. ferrilata* genome assembly.

For chromosome-level assembly, HiC Pro v2.8.1 was used to control the quality of the raw Hi-C data. Due to interruption and error correction of contigs during Hi-C-assisted assembly, the number of contigs eventually changed from 359 to 379. Finally, 157 contigs (2 340 706 585 bp) were anchored onto 18 linkage groups with a mounting rate of 98.36% (Supplementary Table S7; Figure 1A). The final assembly results are presented in Supplementary Table S8. To better understand the *V. ferrilata* genome assembly, basic genomic information was compared with that of other *Vulpes* species (Table 1). Of the three species, the *V. ferrilata* genome had the largest contig N50 (Kukekova et al., 2018; Peng et al., 2021). Analysis showed that the *V. ferrilata* and *Canis lupus familiaris* genomes had a high degree of synteny (Figure 1B), consistent with their close phylogenetic relationship.

GMATA v2.2 and Tandem Repeats Finder v4.07b (2 7 7 80 10 50 500 -f -d -h -r) were used to identify tandem repeats (TRs). Transposable elements (TEs) were identified using MITE-Hunter (-n 20 -P 0.2 -c 3), RepeatModeler v1.0.11 (-engine wublast), and RepeatMasker v1.331 (nolow -no_is -gff -norna -engine abblast -lib lib). We identified a total of 4 545 261 repeats with a size of ~857.2 Mb using repeat annotation, accounting for 37.77% of the assembled *V. ferrilata* genome (Supplementary Table S9 and Figure S3).

Three independent methods were used for gene prediction: i.e., *ab initio* prediction, homology search, and transcriptome prediction. Specifically, for the prediction of RNA-seq-based genes, STAR v2.4 was used to align filtered mRNA-seq reads with the reference genome. The transcript assembled by Stringtie v1.3.4d and PASA v2.3.3 was used to predict open reading frames (ORFs). GeMoWa v1.6.1 was used to align homologous peptides of related species with the assembled transcript to obtain gene structure information for homologous prediction. For *de novo* prediction, RNA-seq-based sequences were assembled using Stringtie v1.3.4d, and a training set was generated using PASA v2.3.3. Augustus v3.3.1 was then used for *ab initio* gene prediction of the training set.

EvidenceModeler v1.1.1 (EVM) was used to combine the three results, for a total of 21 715 protein-coding genes (Supplementary Table S10).

After aligning the obtained genes in four databases, including Non-Redundant Protein Sequence Database (NR), Gene Ontology Resource (GO), Clusters of Orthologous Groups for Eukaryotic Complete Genomes (KOG), and Kyoto Encyclopedia of Genes and Genomes (KEGG), BUSCO evaluation showed that the genome benchmark reached 86.47%, suggesting an almost complete assembly. In total, 18 776 genes were functionally annotated (Supplementary Table S11 and Figure S4). We also annotated different types of non-coding RNAs (ncRNAs) based on the Rfam database. The prediction results indicated that the ncRNAs of the *V. ferrilata* genome included 143 320 transfer RNAs (tRNAs), 3 788 microRNAs (miRNAs), 295 ribosomal RNAs (rRNAs), and 585 regulatory RNAs (Supplementary Table S12). The annotated gene set of *V. ferrilata* was compared to that of other species. Analysis showed that the annotated genes of these species had similar distribution trends, indicating a reliable annotation of the *V. ferrilata* genome (Supplementary Figure S5).

Gene families for *V. ferrilata* and other species (i.e., *Ailuropoda melanoleuca*, *Canis lupus*, *Enhydra lutris*, *Felis catus*, *Homo sapiens*, *Leptonychotes weddell*, *Mus musculus*, *Mustela putorius*, *Panthera tigris*, and *V. vulpes*) were clustered using OrthoMCL v2.0.9 (Li et al., 2003). Finally, 5 835 single-copy genes were obtained and used for phylogenetic inference (Supplementary Figure S6). Although the phylogenetic tree topology is consistent with previous phylogenetic studies, the divergence time between *V. ferrilata* and *V. vulpes* differed (3.27 Ma) (Fritz et al., 2009; Humphreys & Barraclough, 2014) (Supplementary Figure S7). This divergence time coincides with the geological events related to the rapid uplift of the QTP three Ma, indicating a more reliable divergence time estimation for the *Vulpes* genus (Zhong & Ding, 1996).

We compared the expansion and contraction of gene families in the *V. ferrilata* genome with that of 10 other species (including eight species in the order Carnivora) using CAFÉ v4.2.1. Among the 419 expanded gene families in *V. ferrilata*, 79 changed significantly ($P < 0.05$), involving 538 genes (Figure 1C). After KEGG and GO enrichment analysis of the expanded orthologous groups, 67 significantly over-represented pathways and 32 significantly enriched GO terms were obtained (Figure 1D, E; Supplementary Tables S13, S14). Notably, extreme temperature response- and hypoxia-related gene families showed significant pathway expansion, including the oxidative phosphorylation (ko00190, 44 genes, $P = 1.32E-35$), thermogenesis (ko04714, 17 genes, $P = 1.81E-05$), mTOR signaling pathway (ko04150, 44 genes, $P = 2.02E-34$), HIF-1 signaling pathway (ko04066, 21 genes, $P = 7.28E-10$), vascular smooth muscle contraction (ko04270, 12 genes, $P = 2.24E-05$), and calcium signaling pathway (ko04020, 10 genes, $P = 0.005373021$) families. The 32 significantly enriched GO terms, including ATP hydrolysis-coupled proton transport (GO:0015991, 44 genes, $P = 2.18E-53$), oxidoreductase activity (GO:0016620, 33

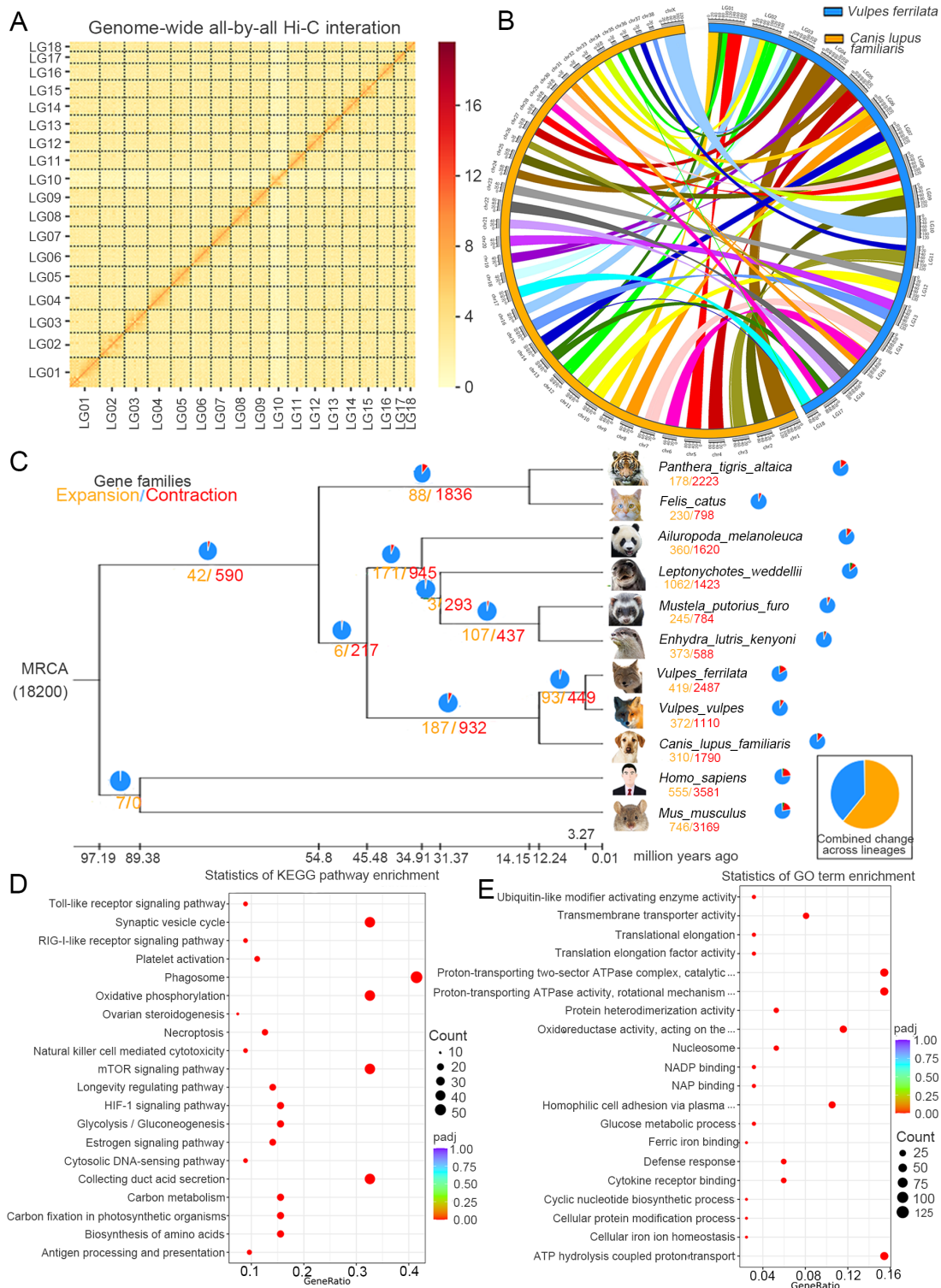


Figure 1 Statistics and data analysis of genome assembly of *Vulpes ferrilata*

A: Genome-wide all-by-all Hi-C interaction identified 18 linkage groups. B: Synteny analysis of *V. ferrilata* and *Canis lupus familiaris* genomes. C: Gene family evolution between genomes of *V. ferrilata* and 10 other species, orange and red numbers indicate gene family expansions and contractions, respectively. Length of branch indicates divergence time. MRCA: Most Recent Common Ancestor. Ma: Million years ago. D: Significant KEGG enrichment in expanded gene families. Value around each bar indicates number involved in each KEGG pathway. E: GO classification of expanded gene families, including top 20 significant GO categories ($P < 0.05$). BP: Biological process; CC: Cellular component; MF: Molecular function.

Table 1 Comparison of genome assemblies between *V. ferrilata* and other *Vulpes* species

Feature	<i>V. ferrilata</i>	<i>V. vulpes</i>	<i>V. lagopus</i>
Assembly level	Chromosome	Scaffold	Chromosome
Total length (bp)	2 379 658 649	2 421 568 072	2 345 550 353
Number of scaffolds	234	82 424	929
Scaffold N50 (bp)	133 960 477	12 472 085	131 537 142
Number of contigs	379	183 898	1 456
Contig N50 (bp)	52 909 674	55 450	33 460 300
GC content (%)	41.25	41.3	41.28

genes, $P=1.17E-23$), glucose metabolic process (GO:0006006, nine genes, $P=5.64E-13$), NAD binding (GO:0051287, nine genes, $P=5.84E-09$), cellular iron ion homeostasis (GO:0006879, seven genes, $P=4.92E-07$), and protein ubiquitination (GO:0016567, three genes, $P=0.025884$), were related to energy metabolism and transportation and may contribute to the adaptation of *V. ferrilata* to low oxygen and extreme temperature conditions. Thus, the significantly expanded genes associated with hypoxia and energy metabolism may reflect mechanisms underlying the adaptations of *V. ferrilata* to high-altitude environments.

The Codeml program in PAML v4.8 was employed to test for positively selected genes (PSGs) in the *V. ferrilata* genome using the branch site model (Yang, 2007). We identified 175 PSGs in *V. ferrilata* (Supplementary Table S15). Of these PSGs, nine may be related to the adaptation of *V. ferrilata* to low oxygen and high UV radiation, including vascular endothelial growth factor A (*VEGFA*) and mitochondrial genome maintenance exonuclease 1 (*MGME1*). More detailed information is provided in Supplementary Table S16. Compared with other native plateau species, the adaptation strategy of *V. ferrilata* is reflected more significantly at the expanded gene family level than in PSGs. This suggests that *V. ferrilata* may have a unique plateau adaptation mechanism, which requires further exploration.

To the best of our knowledge, this is the first report on the chromosome-level genome of *V. ferrilata*, a species unique to the QTP. We also determined the divergence time between *V. ferrilata* and *V. vulpes*. Furthermore, analysis of expanded and contracted gene families and PSGs revealed the possible adaptation strategies of *V. ferrilata* to the plateau environment. This high-quality genome provides a solid basis for future studies on the population and conservation of this species and will help to improve our understanding of the environmental adaptations of species native to the QTP.

SCIENTIFIC FIELD SURVEY PERMISSION INFORMATION

Permission for field surveys in Qinghai Province was granted by the Qinghai Forestry and Grassland Bureau. Project approval of administrative license (2021: No. 7) was issued by the Qinghai Forestry and Grassland Bureau. All sample collection procedures and experiments were approved by the Qinghai Forestry and Grassland Bureau and conformed to the guidelines established by the Ethics Committee for the Care and Use of Laboratory Animals of Qufu Normal University (Permit No. QFNU2018-013).

DATA AVAILABILITY

The genome assembly and sequenced data were submitted to the Science Data Bank databases (DOI 10.11922/sciencedb.01523), National Center for Biotechnology Information (NCBI: PRJNA762184, PRJNA768296, and JAJBZS000000000), and National Genomics Data Center (GSA: CRA006096; GW: GWHBHNE000000000).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

T.S.L. conceived the project. T.S.L., T.G., and L.D.W. collected the samples. Q.G.W., S.Y.Z., Y.H.D., and L.P.S. performed the genome assembly, gene annotation, and bioinformatic analysis. T.S.L. wrote the manuscript. W.L.S., H.S.D., and H.H.Z. revised the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We thank Mr. Ting-Bin Lyu for assistance with sample collection. We are particularly grateful to Jiangwen Cairen for assistance in this study. We would also like to thank the Qinghai Forestry and Grassland Bureau for support during this project.

Tian-Shu Lyu¹, Qin-Guo Wei¹, Li-Dong Wang¹, Sheng-Yang Zhou¹, Lu-Peng Shi¹, Yue-Huan Dong¹, Hua-Shan Dou², Wei-Lai Sha¹, Ta Ga³, Hong-Hai Zhang^{1,*}

¹ School of Life Science, Qufu Normal University, Qufu, Shandong 273165, China

² Hulunbuir Academy of Inland Lakes in Northern Cold & Arid Areas, Hulun Buir, Inner Mongolia 021000, China

³ Ecological Environment and Natural Resources Administration, Qumarléb, Xizang 815500, China

*Corresponding author, E-mail: zhanghonghai67@126.com

REFERENCES

Clark HO Jr, Newman DP, Murdoch JD, Tseng J, Wang ZH, Harris RB. 2008. *Vulpes ferrilata* (Carnivora: Canidae). *Mammalian Species*, **821**: 1–6.

Fritz SA, Bininda-Emonds ORP, Purvis A. 2009. Geographical variation in

- predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology Letters*, **12**(6): 538–549.
- Harris RB, Zhou JK, Ji YQ, Zhang K, Yang CY, Yu DW. 2014. Evidence that the Tibetan fox is an obligate predator of the plateau pika: conservation implications. *Journal of Mammalogy*, **95**(6): 1207–1221.
- Humphreys AM, Barraclough TG. 2014. The evolutionary reality of higher taxa in mammals. *Proceedings of the Royal Society B: Biological Sciences*, **281**(1783): 20132750.
- Kukekova AV, Johnson JL, Xiang XY, Feng SH, Liu SP, Rando HM, et al. 2018. Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. *Nature Ecology & Evolution*, **2**(9): 1479–1491.
- Li L, Stoekert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, **13**(9): 2178–2189.
- Liu N. 2013. Estimation of Tibetan Fox Population Size Using Fecal DNA Technology and Line Transects. Master thesis, East China Normal University, Shanghai. (in Chinese)
- Peng YD, Li H, Liu ZZ, Zhang CS, Li KQ, Gong YF, et al. 2021. Chromosome-level genome assembly of the Arctic fox (*Vulpes lagopus*) using PacBio sequencing and Hi-C technology. *Molecular Ecology Resources*, **21**(6): 2093–2108.
- Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586–1591.
- Zhong D, Ding, L. 1996. Process and mechanism of uplift of Qinghai-Tibet Plateau. *Science in China (Series D)*, **26**(4): 289–295.