Chromosome-level Genome Assembly of the High-altitude Leopard (*Panthera pardus*) Sheds Light on Its Environmental Adaptation

Chuang Zhou¹⁺, Yi Liu¹⁺, Rusong Zhang¹, Xiaofeng Zheng¹, Guangqing Zhao¹, Fengjun Li¹, Wei Liu², Bisong Yue¹⁺∗, Nan Yang³⁴⁺∗

¹Key Laboratory of Bioresources and Ecoenvironment (Ministry of Education), College of Life Sciences, Sichuan University, Chengdu, 610064, P.R. China
²College of Animal Science and Veterinary Medicine, Southwest Minzu University, Chengdu, P. R. China
³Institute of Qinghai-Tibetan Plateau, Southwest Minzu University, Chengdu, P. R. China
⁴Collaborative Innovation Center for Ecological Animal Husbandry of Qinghai-Tibetan plateau, Southwest Minzu University, China

†These authors contributed equally to this work.

* Corresponding authors: E-mails: bsyue@scu.edu.cn; yanggan0204@126.com.

E-mails:
Chuang Zhou: czhou_scu@163.com
Yi Liu: liuyi_shz@163.com
Rusong Zhang: 15940336815@163.com
Xiaofeng Zheng: xiaofengzheng2020@163.com
Guangqing Zhao: 419938295@qq.com
Fengjun Li: lifengjun1994@126.com
Wei Liu: liuwei19831108@163.com
**Abstract**

The leopard (*Panthera pardus*) has the largest natural distribution from low to high altitude areas of any wild felid species, but recent studies have revealed that leopards have disappeared from large areas, probably owing to poaching, a decline of prey species, and habitat degradation. Here we reported the chromosome-scale genome assembly of the high-altitude leopard (HL) based on nanopore sequencing and high-throughput chromatin conformation capture (Hi-C) technology. *Panthera* genomes revealed similar repeat composition, and there was an appreciably conserved synteny between HL and the other two *Panthera* genomes. Divergence time analysis based on the whole genomes revealed that the high-altitude leopard (HL) and the low-altitude leopard (LL) differentiate from a common ancestor approximately 2.2 Mya. Through comparative genomics analyses, we find molecular genetic signatures that may reflect high-altitude adaptation of the HL. Three HL-specific missense mutations were detected in two positively selected genes i.e. *ITGA7* (Ala112Gly, Asp113Val and Gln115Pro) and *NOTCH2* (Ala2398Ser), which is likely to be associated with hypoxia adaptation. The chromosome-level genome of the HL provides valuable resources for the investigation of high-altitude adaptation and protection management of the vulnerable leopard.

Key words: high-altitude leopard, chromosome-level genome, phylogenetic analysis, high-altitude adaptation, missense mutations

**Significance**

The leopard (*Panthera pardus*), widely distributed from low to high altitude areas, is under substantial threats like illegal poaching, prey loss, and habitat degradation, and human activities. In this study, we generated chromosome-level genome assembly from the high-altitude leopard, and investigated its genetic mechanism of high-altitude adaptation. The newly generated data should contribute to a better understanding of the high-altitude adaptation and conservation of the vulnerable leopard.
Introduction

Leopards (Panthera pardus) have pelage hues varying from pale yellow to deep golden and being patterned with black rosettes. They had the most expansive natural range of any large felid, occurring throughout sub-Saharan Africa, India and southern Asia (Hayward et al., 2006), and their coat and color patterns varied widely across various types of habitats. The leopards inhabiting snowy temperate forests with low winter temperatures could display a pale cream-colored long-hair winter coat, which can possibly cause its confusion with the snow leopard (Panthera uncia) (Uphyrkina et al., 2001). Leopards can occur at sea level, in foothill areas, in mountains, as well as in the Himalayas where they are sympatric with snow leopards up to 5,200 m. Leopards are currently listed as vulnerable by the International Union for the Conservation of Nature (IUCN) and in Appendix I of the Convention of International Trade in Endangered Species (CITES). Leopard are confronted with worldwide population declines as a result of substantial threat throughout its range like illegal poaching, prey loss, and habitat degradation, and anthropogenic disturbances (Han et al., 2019). Besides the protective legislation, habitat restoration, and reintroduction for conservation of endangered species, genomic approaches become more widely adopted in conservation. The reference genome is a key factor to investigate many biological problems which are crucial for species conservation like demography, inbreeding, hybridization, disease susceptibility, behavioral ecology, and adaptation (Brandies et al., 2019). For example, the reference genome of the Tasmanian devil (Sarcophilus harrisii), an endangered Australian marsupial, played an indispensable role in understanding an infectious clonal cancer devil facial tumor disease (DFTD) and the management of the species in the wild (Brandies et al., 2019). Therefore a high-quality reference genome is an important conservation and management tool for the protection and long-term survival of the vulnerable leopard.

High-altitude leopards (HLs) primarily live in the mountain ranges where they are sympatric with snow leopards and there are severe physiological challenges such as low
oxygen level, high ultraviolet (UV) radiation, aridity, and low primary productivity (Janecka et al., 2017; Verma et al., 2012). Advantageous genetic mutations and selective pressure were considered as the contributing factors of the adaptation to high-altitude environment. Many studies have sought to determine the molecular genetic basis for the successful adaptation of high-altitude animals like yak (Qiu et al., 2012), grey wolf (Zhang et al., 2014), Tibetan antelope (Ge et al., 2013), goat (Song et al., 2016), and chicken (Wang et al., 2015). Some responsible genes have been worked out by previous studies, among which \textit{EPAS1} (endothelial PAS domain protein1) and \textit{EGLN1} (egl-9 family hypoxia inducible factor 1) were the most prominent ones. \textit{EPAS1} and \textit{EGLN1} are key genes associated with the hypoxia inducible factor (HIF) pathway and possess functional mutations (Lorenzo et al., 2014; Xu et al., 2014). Convergent evolution has occurred in distantly related organisms under the same selective pressures to adapt to high-altitude environment. For example, \textit{EPAS1} gene was reported to be one vital influencing factor in high-altitude adaptation, which was shared by the snow leopard (Cho et al., 2013), Tibetans (Peng et al., 2011; Simonson et al., 2010; Wang et al., 2011), Tibetan grey wolf (Zhang et al., 2014), Tibetan mastiff (Gou et al., 2014; Li et al., 2014; Wang et al., 2014) and Tibetan goat (Song et al., 2016). On the contrary, different geographic populations of the same species could adapt to high-altitude conditions through different genes or functional pathways, such as human (from Tibet, Andes and Ethiopia) (Simonson et al., 2010; Bigham et al., 2010; Scheinfeldt et al., 2012) and Tibetan pig (from Tibet, Gansu, Sichuan and Yunnan province in China) (Ai et al., 2014). The genetic mechanism of high-altitude adaptation in the HL, however, remains perplexing. Thereby, understanding genetic factors that underlie adaptation to high-altitude conditions could fill important gaps in our use and understanding of conservation genetics to support HL conservation.

Currently, most of reference genomes were assembled based on short reads generated through second-generation sequencing technologies (such as Illumina or Roche 454) with
limited contiguity and quality. Compared to second-generation sequencing technologies, long-read sequencing technologies (such as Oxford nanopore and PacBio SMRT) can generate long reads and enables the assembly of a genome with a high level of completeness. These long reads can span complex regions that have not been sequenced. The value of the ultra-long reads produced by ONT has been reported in improving the contiguity and completeness of assembled genomes (Ge et al., 2019). In addition, high-throughput chromatin conformation capture (Hi-C) technology allows the genome sequences to be assembled to the scale of full chromosomes based on capture of all DNA interaction patterns in chromatin (Lieberman-Aiden et al., 2009). In this study, we combined nanopore long-read sequencing and Hi-C sequencing to generate a high-quality genome for the HL. This chromosome-scale genome can lay a strong foundation for fully understanding the diversity and population dynamics of the HL and can provide a sound support for breeding, functional genomic research and species conservation.

Results and Discussion
Genome Sequencing and Assembly
The detailed sequencing data was summarized in Table S1; In total, 694.64 Gb of sequencing data was generated, of which 164.16 Gb (67.19X coverage) was produced from Illumina reads, 264.53 Gb (98.52X coverage) from ONT reads and 265.95 Gb (108.77X coverage) from the Hi-C library. A k-mer depth of 50 was the highest peak in the figure, and the k-mer number 123,373,857,631 was used to calculate the HL genome size (fig. S1). The genome size of the HL was estimated to be 2.44 Gb, and the heterozygosity was about 0.20%. The repeatability was about 52.81% and the GC content was about 42.58%. The genome was assembled with ONT long reads and yielded 2.43 Gb of reference genome size with contig N50 length about 60.11 Mb and a longest length of 166.45 Mb. In the case of the Hi-C super-scaffolding, the total size of the genome was 2.43 Gb with the N50 value of 147.15 Mb in length. The HL genome was the best
assembled in the Panthera taxon (Table 1). Although the N50 scaffold of the lion was as high as 136.05 Mb, its N50 contig (0.29 Mb) was far lower than that of the HL (60.11 Mb). The genome size of the final assembled version is almost the same as that estimated by k-mer, indicating the accuracy of the assembly. In the Panthera that have been sequenced, the HL genome (2.43 Gb) was larger than the Panthera tigris (tiger) genome (2.39 Gb) and Panthera leo (lion) (2.41 Gb), smaller than the Panthera onca (jaguar) (2.50 Gb) and LL (2.58 Gb).

To assess the quality of chromosome-level genome assembly, a Hi-C heat map of the whole genome visualized it. There were 19 chromosomes in the figure, and the interaction signal strength of the two pairs of chromosomes around the diagonal was much stronger than the interaction strength of the chromosomes at other positions (fig. 1), which indicate that the quality of the genome assembly was very high. We evaluated the completeness of the complete genome assembly through calculating coverage for a set of single-copy orthologous genes in mammals using BUSCO, which revealed a genome coverage rate of 94.8%. Although the BUSCO values of near-source species were not much different: 96.1% for lion, 95.5% for LL, and 94.3% for tiger, the BUSCO value of jaguar was 81.9% due to the higher proportion of fragmented BUSCOs (12.3%) than that of other species (fig. S2). The lower the fragmented BUSCOs, the better the assembly quality. The possible reason is that although the fragmented gene reached the predicted score, the length did not meet the requirement. It is possible that the fragment was not assembled during the genome assembly process, or it may be that the gene was not completely predicted due to the particularity of the gene structure during gene prediction.

In order to evaluate the accuracy of assembly, the comparison rate of Illumina library reads to the assembled genome was about 98.84%, and the coverage rate was about 97.90%, indicating that the reads and the assembled genome have good consistency. There are a total of 1,104,323 SNPs in the HL genome (heterozygous SNPs 1,098,046, homozygous SNPs 6,277), accounting for 0.047% of the total genome.
Given the close relationship between HL and the other two *Panthera* species and the size and quality of the draft genome for the other two *Panthera* species, we performed genomic synteny analyses. Substantial genome-wide colinearity was illustrated between HL and other two *Panthera* species (fig. 2). For deeper understanding of chromosomal synteny, we took the HL and lion genomes for example to conduct one-to-one chromosomal alignment. The findings showed close overall genome synteny between HL and lion. The X, B4 and B1 chromosomes of the lion and the HL chromosomes 8, 10 and 12 have very good collinearity, which can be said to be completely coincident (fig. 3a). In addition, chromosomes D1, C2, and B2 of the lion and chromosomes 3, 4, and 6 of the HL showed a high degree of coincidence in addition to a small number of rearrangements (fig. 3b). Differences in genomic structure between closely related species are considered to be a major factor in species diversification, because gene requires recombination in collinear chromosomes (Zhang and Wang, 2004).

**Genome Characterization**

We found that 36.97% of the HL genome was composed of repetitive elements, including long interspersed elements (LINEs) (22.68%), short terminal repeat (SINE) (6.93%), long terminal repeat (LTR) (5.27%), and DNA elements (3.22%) (Table S2). A total of 1,323,721 perfect SSRs were identified, including 494,352 mono-, 518,799 di-, 58,490 tri-, 193,811 tetra-, 44,212 penta- and 14,057 hexa-nucleotide SSRs (Table S3). The obtained consensus gene set composed of a total of 19,120 protein-coding genes (PCGs), lower than that of Amur tiger (20,226). The longest chromosome of the HL was chromosome 5 (238.91Mb), and the chromosome with the most coding genes was chromosome 2 (9,972), which is far more than other chromosomes (Table 2). Homology-based and structure-based strategies were employed for functional annotation of protein-coding genes. We found functional annotation for 19,102 PCGs, which represents about 99.91% of all the genes (Table S4). For clarity, the distributions of gene density, GC
density and repeat density across the 19 HL chromosomes were further illustrated (fig. 4).

Generally, the regions with low gene density had high repeat content, and the regions with high repeat content usually had high GC content.

Gene Family, Phylogeny and Divergence

A gene family, a group of homologous genes with similar structures, generally has similar functions (Demuth et al., 2006). A total of 19,120 HL genes were clustered into 15,709 gene families with an average of 1.22 genes per gene family. We identified 51 HL-specific gene families which have 154 HL genes among the ten mammalian species included in the analysis. Furthermore, there were 802 HL-specific gene families among Panthera species (fig. 5a). These lineage-specific gene families possibly have contributed to the evolution of the HL. The phylogenetic tree constructed using single-copy orthologues revealed that leopard was most closely clustered with lion, and further, formed into a clade genus Panthera (fig. 5b), which was in line with previous studies (Bagatharia et al., 2013; Christiansen, 2008). The HL diverged approximately 2.2 million years ago from its common ancestor and the genus Panthera diverged about 15.5 million years ago from the genus Felis.

In order to examine the evolutionary history of gene families, the expansion and contraction in the HL genome were estimated in comparison to other mammals. We found that the HL genome composed of 885 expanded gene families and 2,882 contracted gene families (fig. 5b). The 885 HL expanded gene families contained 1,246 genes, while 2,249 genes for 2,882 contracted gene families. In this study, it was observed that many expanded gene families were distributed in pathways associated with high-altitude adaptation in the HL genome such as HIF (GO:0102113), response to hypoxia (GO:0001666), and response to UV (GO:0009411). The study of nucleotide polymorphisms in Tibetans reveals strong evidence that HIF2α and other genes in the HIF signaling pathway were positively selected (Rankin et al., 2007). In addition to
hypoxia, strong ultraviolet radiation environments in high altitude areas is also the main limiting factor for the successful colonization of animals. Over the higher altitude of Tibetan Plateau, the UV erythemal dose has a higher value with the multi-yearly mean value is about 5500 J • m$^{-2}$, and over some regions, the value is up to 6000 J • m$^{-2}$, while the low altitude area is only 1500 J • m$^{-2}$ (Xiao and Jiang, 2013). Many significantly contracted gene families in the HL were distributed in olfactory receptor activity (GO:0004984) and immune response (GO:0006955). Due to the thin air, single environment, and simple odor molecules in the air at high altitudes, the olfactory receptor genes of species usually shrink significantly compared to lower altitudes (Li et al., 2013). The whole genome sequencing of Tibetan chickens and ground fits living on the plateau showed a large number of olfactory receptor genes were lost in the genome, thereby reduced energy consumption during olfactory perception (Wang et al., 2015; Qu et al., 2013). In addition, the high-altitude environment has strong ultraviolet radiation and fewer types of pathogenic microorganisms, so a lot of immune-related genes have contracted or lost at high altitudes. However, the immune genes have adaptively evolved in Tibetan chickens may be related to artificial domestication pressure was greater than natural selection pressure (Zhang et al., 2016).

Positive Selection
Positive selection provides evolutionary innovation in specific adaptation, and the PSGs in HL had a significantly higher ratio of nonsynonymous substitutions to synonymous substitutions than other genes. Based on orthologous gene annotation by OrthoFinder2 across ten mammals, 6,463 single-copy genes were used for positive selection analyses. As a result, we identified 817 positive selected genes in the HL using the branch-site model in PAML. The GO and KEGG enrichment analyses showed significant terms and pathways were involved in high-altitude adaptation, such as vascular smooth muscle contraction (KEGG map04270), positive regulation of vascular endothelial growth factor...
signaling pathway (GO:1900748), angiogenesis (GO: 0001525); regulation of DNA recombination (GO:0000018), response to ionizing radiation (GO:0010212) (fig. 5c).

Hypoxia has a greater impact on the animal's cardiovascular system, and high-altitude species mainly improve hypoxic tolerance by enhancing myocardial contraction and angiogenesis. In the ground tit, PSGs (HIF1AN, hypoxia inducible factor 1 subunit alpha inhibitor; ANGP, angiopoietin; ADAM family, etc.) were involved in angiogenesis, cardiopulmonary development, and reactive oxygen metabolism (Wang et al., 2015). These genes also existed in high-altitude yaks and humans, indicating different animal groups have similar adaptation mechanisms to hypoxic response (Qiu et al., 2012; Simonson et al., 2010) VEGF is the main factor that induces angiogenesis in vivo. Studies on VEGF in yak lung blood vessels have found that VEGF was scattered only on the alveolar wall. Under normal oxygen concentration, the expression of VEGF was very low, which played a role in maintaining the balance of blood vessel density. While the expression increased significantly during hypoxia, because vascular endothelial cells secreted some growth factors to control the proliferation and growth of smooth muscle cells and fibroblasts, making the structure of yak lungs different from cattle at low altitudes. Through comparative genomics analysis, 8 common amino acid substitutions of 6 genes (RNASE4, DNAH11, CDT1, RTEL1, ARMC2, and NT5DC1) were found in 3 high-altitude golden monkey species, which were related to lung function, DNA repair, and angiogenesis. Ultraviolet irradiation experiments were carried out on the CDT1 gene related to DNA repair. The results showed that the mutant CDT1 (A537V) has stronger stability than the wild type. It is speculated that the mutation will help the golden monkey to resist ultraviolet rays in a high-altitude environment. Detection of the RNASE4 gene related to angiogenesis found that the mutant RNASE4 (N89K+T128I) has higher activity in inducing HUVEC cells to generate tubular structures. It is speculated that these two mutations may enhance the angiogenic ability of RNASE4 and help the golden monkey adapt high altitude environment (Yu et al., 2016).
Further examination of these 817 PSGs found that two genes (*ITGA7* and *NOTCH2*) functionally associated with hypoxia adaptation have four HL specific missense mutations i.e. *ITGA7* (Ala112Gly, Asp113Val and Gln115Pro) and *NOTCH2* (Ala2398Ser) (fig. 6), which suggested a role in the high-altitude adaptation of the HL. *ITGA7* participates in PI3K-akt and MAPK signaling pathways, indirectly regulates VEGF and HIF signaling pathways and enhances hypoxic adaptation (Zhang et al., 2015). Notch signal directly or indirectly participates in the regulation of angiogenesis through interactive dialogue with vascular endothelial growth factor, BMP-SMAD signaling pathway, extracellular matrix molecules, etc., and sustained Notch signal is to maintain the stability of the structure and function of adult vascular system. *NOTCH1* loss of function can cause zebrafish vascular remodeling disorders including the collapse of the dorsal large artery (Fish and Wythe, 2015). High concentration of vascular endothelial growth factor can induce the expression of arterial marker genes ephrinB2, *Dll4* and *NOTCH4* (Zhang et al., 2008).

**Conclusions**

Here we assembled the chromosome-level genome of the HL by *de novo* assembly of long reads produced by Nanopore sequencing and Hi-C. The reference quality genome (2.43 Gb) of the HL has the assembled contig N50 of 60.11 Mb and the longest contig of 166.45 Mb. By utilizing Hi-C technology, we assembled contigs into scaffolds that resulted in a chromosome-level genome assembly with 19 chromosomes and a scaffold N50 length of 46.03 Mb. The genome was functionally annotated to produce a total of 19,102 (99.91%) protein-coding genes. The phylogenetic analysis showed that the leopard was most closely related to the lion in *Panthera* species, and the HL diverged from the common ancestor 2.2 Mya. The functional enrichment analysis of PSGs in the HL revealed significant terms and pathways which were associated with high-altitude adaptation. We reported four HL-specific missense mutations detected in two positively
selected genes i.e. *ITGA7* (Ala112Gly, Asp113Val and Gln115Pro) and *NOTCH2* (Ala2398Ser), which possibly played a pivotal role in the hypoxia adaptation of HL. The chromosome-scale genome of the HL provides an invaluable resource for fully understanding the diversity and population dynamics of the HL and can provide a sound support for breeding, functional genomic research and species conservation of the leopard.

**Materials and methods**

**Sample Collection and Sequencing**

The muscle sample preserved in the Natural Museum of Sichuan University was collected from a naturally dead male HL in Baiyu county, Sichuan province where the altitude was almost above 3,500 m. SDS extraction method (50ml) was used to standardize the operation procedure to extract DNA. Agarose gel electrophoresis was used to analyze the purity and integrity of DNA. Nanodrop detected the purity of DNA (OD260/280 ratio). Qubit fluorimeter (Invitrogen, Carlsbad, CA, USA) accurately quantified the DNA concentration. Second-generation sequencing was performed on Illumina NovaSeq PE150. The DNA samples passed the sequencing test were randomly broken into fragments with a length of 350 bp by the Covaris breaker. NEB Next® Ultra DNA Library Prep Kit ((NEB, Ipswich, MA, USA) was used for library construction, and the reagents and consumables recommended in the instructions were strictly used.

ONT sequencing was conducted on flow cells of a PromethION 24 sequencer (version R9.4.1, ONT, UK). First use BluePippin (Sage Science, Beverly, MA, USA) to select DNA size selection (30–80 kb). The nanopore library was then constructed using the Ligation Sequencing Kit 1D (SQK-LSK109) (New England Biolabs). It mainly includes 3 steps. 1. Use NEBNext End repair / dA-tailing Module (E7546) (New England Biolabs) for DNA repair and NEBNext FFPE DNA Repair Mix (M6630) for end-prep. 2. Use NEBNext Quick Ligation Module (E6056) (New England Biolabs) for Adapter ligation
and clean-up. 3. Priming and loading the flow cell.

The Hi-C library was sequenced on Illumina NovaSeq PE150. The tissue was treated with paraformaldehyde, a cell cross-linking agent, to fix the chromatin conformation in the nucleus. After cell lysis, the above-mentioned cross-linked fixed chromatin was treated with restriction enzyme Hind III to create gaps on both sides of the cross-linking point. When the end was repaired, the biotin-14-dATP was added to label the end of the oligonucleotide. Nucleic acid ligase joined adjacent DNA fragments. The protease digested the protein at the junction to release the cross-linking state of the protein and DNA. Covaris M220 (Covaris, Woburn, MA) was used to randomly break the extracted genomic DNA into 350bp fragments. Under the adsorption of avidin magnetic beads, the DNA with biotin is captured, and the entire library preparation was completed in strict accordance with the steps of terminal repair, addition of A, linker connection, PCR amplification and purification of the DNA fragments.

Genome Assembly

Illumina data was used to estimate the genome size of the HL, and K-mers were counted by jellyfish version 2.2.9 (Marçais and Kingsford, 2011) with 17-base oligonucleotide. The depth of ONT data coverage was calculated according to the estimated genome of survey. Then nanoplot version 1.18.2 was employed for quality control of nanopore data. Finally, we used wtdbg version 1.2.8 (Ruan and Li, 2020) to preassemble nanopore data to obtain genome contigs sequence (parameter -p 19 -k 0). Two rounds of polishing were performed on the contigs obtained from the initial assembly. Based on the default parameters of Racon version 1.32, nanopore data was used to polish the contig sequences (Vaser et al., 2017), and according to pilon version 1.22 (Walker et al., 2014), Illumina data was used to polish the contig sequences (parameter: -Xmx30g --diploid --changes -- threads 8). In order to raise the assembled contigs to the chromosomal level, the ALLHiC (Zhang et al., 2019) was used to uniquely map Hi-C reads into contigs, and to retain the
500 bp flanking region of the restriction site (MboI) for further analysis. The number of chromosomes was 19 and the restriction sites were 50. In addition, low-quality matches were filtered according to the pipeline (https://github.com/tangerzhang/ALLHiC/wiki). Subsequently, the genetic algorithm optimized in ALLHiC was used to order and orient the contigs. In order to assess the completeness and accuracy of genome assembly, BUSCO version 2.0 (Simão et al., 2015) evaluated the quality of the genome. Furthermore, we selected Illumina library reads using BWA-MEM v.0.7.17 (Li and Durbin, 2009) to compare to the assembled genome, counted the comparison rate of reads, the extent of genome coverage and the distribution of depth, and evaluated the integrity of the assembly and the uniformity of sequencing. Then we used samtools v1.3.1 (Durbin et al., 2009) to sort the results of BWA by chromosome coordinates, removed duplicate reads, performed SNP Calling, filtered the original results, and finally got the SNP statistical results.

Genome Synteny Analysis
To visualize the concordance between the HL and other two chromosome-level Panthera genomes, we conducted genomic syntenic analyses with MCscan (Python version) (Tang et al., 2008) and LAST (Kielbasa et al., 2011) v746.

Characterization of Repeats
There are a large number of repeat elements in the nuclear DNA in most eukaryotic genomes, which have been indicated to have structural and functional roles (Biscotti et al., 2015). Repeatmodeler v1.0.3 (Benson, 1999), including RECON v1.08 (Bao and Eddy, 2002) and repeatscout v1.0.5 (Price et al., 2005), was employed to construct a de novo repeat library based on the complete genome with default parameters. Then the repeat consensus database with classification information (Tarailo-Graovac and Chen, 2009) was merged with the Repbase (Jurka et al., 2005) database to construct a repeat
consensus database. Repeatmasker v4.0.6 (Tarailo-Graovac and Chen, 2009) was adopted to predict interspersed repeat elements in the whole genome against the Repbase and *de novo* repeat libraries with default parameters. Krait tool (Du et al., 2018) was employed to predict and characterize genome-wide microsatellite (SSR) loci of the whole genome, which can identify the loci that could be used for population genetic studies.

Gene Prediction and Annotation

Using the repeat-masked genome, we combined *de novo* and homology-based approaches to predict gene models. In the *de novo* method, the software packages GENSCAN v3.1 (Burge and Karlin, 1997), AUGUSTUS v2.4 (Stanke and Waack, 2003), GLIMMERHMM v3.0.4 (Majoros et al., 2004) were employed for predicting the protein-coding genes of the HL genome with internal gene models. For the homology-based prediction, protein sequences from six mammals (lion, LL, tiger, cat, dog and human) were aligned to the HL genome with TBLASTN. High-scoring segment pairs (HSPs) were concatenated using solar (version 0.9.6). GENEWISE (Birney et al., 2004) (version 2.2.0) was used to analyze these alignments to determine accurately spliced alignments. Finally, EVidenceModeler v1.1.1 was used to integrate the above evidences with different weights for each to generate a consensus gene set (Haas et al., 2008). To functionally annotate the predicted genes of the HL, their protein sequences were used as queries to search against the swissprot and trembl protein databases (Bairoch and Apweiler, 2000) using BLASTP with an E-value cutoff of $1 \times 10^{-5}$. In order to annotate the functional motifs and protein domains, we employed the InterproScan tool (version 5.27) (Hunter et al., 2009) in coordination with publicly available databases including Gene3D, PRINTS, Pfam, CDD,SMART, MobiDBLite and PROSITE. Gene ontology (GO) IDs for each gene were assigned based on the results of SwissProt. To find the best match and the involved pathway for each gene, all genes were uploaded to KAAS (Moriya et al., 2007), a Web server for functional annotation of genes against the manually corrected KEGG.
gene database by BLAST, using the bi-directional best hit (BBH) method.

Gene Family Identification and Phylogenetic Analysis

Orthologous gene clusters and 1:1 orthologous genesets among 10 mammals (HL, LL, lion, tiger, jaguar, cat, dog, cattle, horse and human) were identified using the package Orthofinder2 (Emms and Kelly, 2018) with Diamond as protein aligner (Buchfink et al., 2015). The software prank v3.8.31 was used to align the sequences of 1:1 orthologous genes from 10 mammals. The alignment outputs were treated with Gblocks (Castresana, 2000) to remove poorly aligned sequences. Then, the coding sequences of 1:1 orthologous genes were concatenated for each species to develop a super-sequence for building the tree. The maximum likelihood (ML) phylogenetic tree was constructed using RAxML (Stamatakis, 2014) with the GTRGAMMA model and 1,000 rapid bootstrap replicates. Human was set as the outgroup. McMctree as implemented in the PAML package (Yang, 1997) was used to predict divergence times. Five calibration times were obtained from the TimeTree database (http://www.timetree.org/), including divergence times of cat and lion (12.2-16.6 Mya), cat and dog (51.0-56.0 Mya), cat and horse (70.2-79.0 Mya), cat and human (91.0-101.0 Mya), and cattle and horse (76.0-82.0 Mya). Café tool (version 4.0) was used to analyze the expansion and contraction of orthologous gene families between ancestor and each of the ten species (Han et al., 2013). The gene family results from Orthofinder2 (Emms and Kelly, 2018) and a tree with estimated divergence times between species were used as inputs. We used a criterion of $P < 0.05$ for significantly changed gene families, and the separate birth ($\lambda$) and death ($\mu$) rates were estimated with the same program using the lambda/mu command with –s and –t options.

Positive Selection Analysis

The above one-to-one gene clusters and the phylogenetic tree between ten mammals were used to identify positively selected genes (PSGs). The lineage HL was designated as
‘foreground’ phylogeny. The software prank was used to align the coding sequences and poorly aligned sequences with gaps were removed with a codon model using Gblocks. The values of dN, dS, and ω were estimated with the codeml program implemented within the PAML package (Yang, 2007). The basic and branch-site models were tested, and genes under relaxation of selective pressure were eliminated by Likelihood Ratio Tests (LRTs). We then identified the PSGs of the HL by means of FDR adjustment with Q values < 0.05. The functional categories and pathways enriched in the PSGs were analyzed by Metascape (Zhou et al., 2019). The HL-specific missense mutations were validated by comparison with more mammals with publicly available genomes.

Acknowledgements

We thank prof. Xiuyue Zhang for valuable advice. This study was supported by the Second Tibetan Plateau Scientific Expedition and Research Program (STEP), Grant No.2019 QZKK0501; and “the Fundamental Research Funds for the Central Universities”, Southwest Minzu University (2020PTJS22001).

Data Availability

The high-altitude leopard whole-genome sequencing data and the chromosome-level genome assembly have been deposited to the CNSA (CNGB Nucleotide Sequence Archive) with accession CNP0001205 (https://db.cngb.org/cnsa/), and the NCBI under the accessions SRR13500268, SRR13500269, SRR13500270, SRR13500271, SRR13500272, SRR13500273, SRR13500274, SRR13500275, SRR13500276, and SRR13500277.
Literature Cited

13. Christiansen P. 2008. Phylogeny of the great cats (Felidae: Pantherinae), and the influence of fossil taxa and missing characters. Cladistics: the international journal

18


Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting
genome annotation and pathway reconstruction server. Nucleic Acids Res. 35(Web
Peel PE, Hogg CJ, Belov K. 2019. The Value of Reference Genomes in the
Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large
Qu Y, et al. 2013. Ground tit genome reveals avian adaptation to living at high altitudes in
the Tibetan plateau. Nat Commun. 4:2071.
Rankin EB, et al. 2007. Hypoxia-inducible factor-2 (HIF-2) regulates hepatic
17(2):155-158.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO:
assessing genome assembly and annotation completeness with single-copy
Song S, et al. 2016. Exome sequencing reveals genetic differentiation due to high-altitude
adaptation in the Tibetan cashmere goat (Capra hircus). BMC Genomics 17:122.


Xu XH, et al. 2014. Two functional loci in the promoter of EPAS1 gene involved in high-


Figure Legends

**Figure 1.** Hi-C interaction heat map between 19 chromosomes for the HL genome.

**Figure 2.** Examples of chromosome rearrangements. The lines between the two horizontal lines link the alignment blocks.

**Figure 3.** Alignments of the HL chromosomes to lion chromosomes. (a) examples of high synteny and high assembly correctness of HL chromosomes. (b) examples of chromosomal rearrangements between the HL and lion. Forward alignments are in blue, and reverse alignments are in red.

**Figure 4.** The genome landscape of HL. From outer to inner circles: the 19 chromosomes at the Mb scale, gene density, GC density and repeat density across the genome, respectively, drawn in 1 Mb non-overlapping windows.

**Figure 5.** Comparative genomics analysis related to the HL. (a) Orthologous gene clusters among *Panthera* species. (b) Phylogenetic tree constructed using one-to-one orthologous genes. The time lines indicate divergence times among the species. (c) Enrichment analysis of the PSGs of the HL.

**Figure 6.** Missense mutations in the HL. (a) Amino acid sequence alignment of *ITGA7*. The HL-specific missense mutations in *ITGA7* is marked in red. (b) Amino acid sequence alignment of *NOTCH2*. The HL-specific missense mutation in *NOTCH2* is marked in red.
Table 1 Assembly statistics of the HL and other Panthera species

<table>
<thead>
<tr>
<th>Species</th>
<th>HL</th>
<th>P. tigris</th>
<th>P. leo</th>
<th>P. onca</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequencing technology</td>
<td>Illumina; Oxford Nanopore; Hi-C</td>
<td>Illumina</td>
<td>Illumina; Oxford Nanopore; 10X Genomics</td>
<td>Illumina</td>
<td>Illumina</td>
</tr>
<tr>
<td>Assembly size (Gb)</td>
<td>2.43</td>
<td>2.39</td>
<td>2.41</td>
<td>2.51</td>
<td>2.58</td>
</tr>
<tr>
<td>Number of scaffolds</td>
<td>3,384</td>
<td>1,479</td>
<td>8,060</td>
<td>438,792</td>
<td>50,377</td>
</tr>
<tr>
<td>N50 scaffold (Mb)</td>
<td>147.15</td>
<td>8.86</td>
<td>136.05</td>
<td>0.12</td>
<td>21.70</td>
</tr>
<tr>
<td>Number of contigs</td>
<td>4,607</td>
<td>157,032</td>
<td>23,775</td>
<td>460,121</td>
<td>265,330</td>
</tr>
<tr>
<td>N50 contig (Mb)</td>
<td>60.11</td>
<td>0.03</td>
<td>0.29</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>GC%</td>
<td>41.7%</td>
<td>41.5%</td>
<td>41.6%</td>
<td>41.7%</td>
<td>41.9%</td>
</tr>
</tbody>
</table>

Note: The assembly statistics of other Panthera genomes were from NCBI assembly database. The GenBank assembly accession numbers were as follows: P. tigris (GCA_000464555.1), P. leo (GCA_008795835.1), P. onca (GCA_004023850.1), LL (GCA_001857705.1).

Table 2 The statistics and characteristics of the HL chromosomes

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Chromosome size (Mb)</th>
<th>Anchored scaffold number</th>
<th>Anchored gene number</th>
<th>Percentage repetitive sequences (%)</th>
<th>GC content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr1</td>
<td>93.18</td>
<td>50</td>
<td>672</td>
<td>36.83</td>
<td>43.08</td>
</tr>
<tr>
<td>Chr2</td>
<td>39.30</td>
<td>42</td>
<td>643</td>
<td>37.00</td>
<td>48.49</td>
</tr>
<tr>
<td>Chr3</td>
<td>157.78</td>
<td>15</td>
<td>877</td>
<td>39.45</td>
<td>39.65</td>
</tr>
<tr>
<td>Chr4</td>
<td>222.61</td>
<td>103</td>
<td>1,730</td>
<td>38.01</td>
<td>41.46</td>
</tr>
<tr>
<td>Chr5</td>
<td>238.91</td>
<td>129</td>
<td>1,265</td>
<td>38.43</td>
<td>39.80</td>
</tr>
<tr>
<td>Chr6</td>
<td>205.14</td>
<td>59</td>
<td>987</td>
<td>40.24</td>
<td>39.04</td>
</tr>
<tr>
<td>Chr7</td>
<td>86.62</td>
<td>4</td>
<td>630</td>
<td>36.99</td>
<td>43.27</td>
</tr>
<tr>
<td>Chr8</td>
<td>81.73</td>
<td>100</td>
<td>370</td>
<td>38.25</td>
<td>40.37</td>
</tr>
<tr>
<td>Chr9</td>
<td>150.65</td>
<td>79</td>
<td>1,047</td>
<td>39.13</td>
<td>40.15</td>
</tr>
<tr>
<td>Chr10</td>
<td>147.15</td>
<td>74</td>
<td>1,194</td>
<td>39.94</td>
<td>41.87</td>
</tr>
<tr>
<td>Chr11</td>
<td>141.31</td>
<td>55</td>
<td>1,172</td>
<td>39.00</td>
<td>41.31</td>
</tr>
<tr>
<td>Chr12</td>
<td>139.24</td>
<td>44</td>
<td>1,096</td>
<td>37.61</td>
<td>42.89</td>
</tr>
<tr>
<td>Chr13</td>
<td>119.75</td>
<td>130</td>
<td>785</td>
<td>53.80</td>
<td>40.62</td>
</tr>
<tr>
<td>Chr14</td>
<td>93.56</td>
<td>66</td>
<td>753</td>
<td>37.65</td>
<td>42.60</td>
</tr>
<tr>
<td>Chr15</td>
<td>168.59</td>
<td>68</td>
<td>1,613</td>
<td>37.26</td>
<td>42.27</td>
</tr>
<tr>
<td>Chr16</td>
<td>69.36</td>
<td>73</td>
<td>657</td>
<td>37.40</td>
<td>42.77</td>
</tr>
<tr>
<td>Chr17</td>
<td>108.80</td>
<td>11</td>
<td>1,286</td>
<td>38.89</td>
<td>42.53</td>
</tr>
<tr>
<td>Chr18</td>
<td>62.86</td>
<td>116</td>
<td>1,047</td>
<td>33.97</td>
<td>47.50</td>
</tr>
<tr>
<td>Chr19</td>
<td>60.34</td>
<td>40</td>
<td>920</td>
<td>37.84</td>
<td>46.17</td>
</tr>
<tr>
<td>Sum</td>
<td>2,387.18</td>
<td>1,258</td>
<td>18,744</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 1
287x250 mm (.32 x DPI)
Figure 2

161x87 mm (32 x DPI)
Figure 3

Lion chromosome Y
Lion chromosome B4
Lion chromosome B1

(b)

Lion chromosome D1
Lion chromosome C2
Lion chromosome B2

(a)

Lion chromosome 8
Lion chromosome 10
Lion chromosome 12

559 x 362 mm (.32 x DPI)
Figure 5

400x317 mm (.32 x DPI)
<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HL</strong></td>
<td>AYPFPAPSVGKYPITPPSQHYSASSNAERTPSSHSHLGHEHPYLYTPSPESPDPQ</td>
</tr>
<tr>
<td><strong>LI</strong></td>
<td>AYPFPAPSVGKYPITPPSQHYSASSNAERTPSSHSHLGHEHPYLYTPSPESPDPQ</td>
</tr>
<tr>
<td><strong>Lion</strong></td>
<td>AYPFPAPSVGKYPITPPSQHYSASSNAERTPSSHSHLGHEHPYLYTPSPESPDPQ</td>
</tr>
<tr>
<td><strong>Tiger</strong></td>
<td>AYPFPAPSVGKYPITPPSQHYSASSNAERTPSSHSHLGHEHPYLYTPSPESPDPQ</td>
</tr>
<tr>
<td><strong>Cat</strong></td>
<td>AYPFPAPSVGKYPITPPSQHYSASSNAERTPSSHSHLGHEHPYLYTPSPESPDPQ</td>
</tr>
<tr>
<td><strong>Dog</strong></td>
<td>AYPFPAPSVGKYPITPPSQHYSASSNAERTPSSHSHLGHEHPYLYTPSPESPDPQ</td>
</tr>
<tr>
<td><strong>Cattle</strong></td>
<td>AYPFPAPSVGKYPITPPSQHYSASSNAERTPSSHSHLGHEHPYLYTPSPESPDPQ</td>
</tr>
<tr>
<td><strong>Horse</strong></td>
<td>AYPFPAPSVGKYPITPPSQHYSASSNAERTPSSHSHLGHEHPYLYTPSPESPDPQ</td>
</tr>
<tr>
<td><strong>Human</strong></td>
<td>AYPFPAPSVGKYPITPPSQHYSASSNAERTPSSHSHLGHEHPYLYTPSPESPDPQ</td>
</tr>
</tbody>
</table>

Figure 6

228x128 mm (.32 x DPI)