ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/359593567

Comparative genomic analysis of high-altitude adaptation for Mongolia Mastiff, Tibetan Mastiff, and Canis Lupus

Article *in* Genomics · March 2022

DOI: 10.1016/j.ygeno.2022.110359

CITATIONS
0

READS
132

6 authors, including:

Chunmei Cai
Qinghai University
21 PUBLICATIONS **285** CITATIONS

SEE PROFILE

All content following this page was uploaded by Chunmei Cai on 21 July 2022.

The user has requested enhancement of the downloaded file.

Genomics 114 (2022) 110359

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Comparative genomic analysis of high-altitude adaptation for Mongolia Mastiff, Tibetan Mastiff, and Canis Lupus

Chunmei Cai [a,b], Yingzhong Yang [a,b], Qin Ga [a,b], Guocai Xu [a,b], Rili Ge [a,b,*], Feng Tang [a,b,*]

[a] Research Center for High Altitude Medicine, School of Medical, Qinghai University, Xining 810016, PR China
[b] Key Laboratory of Application and Foundation for High Altitude Medicine Research in Qinghai Province, Xining 810016, PR China

ARTICLE INFO

Keywords:
Evolution
High-altitude adaptation
Genomics
Tibetan mastiff
Mongolia mastiff
Canis Lupus

ABSTRACT

Tibetan Mastiff has adapted to the extreme environment of the Qinghai-Tibetan Plateau. Yet, the underlying mechanisms of its high-altitude-adaptation and origin remains elusive. Here, we generated the draft genomes of Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus*. The phylogenetic tree uncovered that Tibetan Mastiff and Mongolia Mastiff were derived from *Canis Lupus* species. The comparative genomic analyses identified that the expansion of gene families related to DNA repair and damage response, and contraction related to ATPase activity revealed the genetic adaptations of Tibetan Mastiff and *Canis Lupus* to high altitude. In addition, the Tibetan Mastiff and *Canis Lupus* had signals of positive selection for genes involved in fatty-acid α/β- oxidation for highland adaptation. Notably, the positively selected *TERT* of Tibetan Mastiff should be an adaptive trait for correcting DNA damage. These findings suggested that the Tibetan Mastiff and *Canis Lupus* evolves basic strategies for adaptation to high altitude.

1. Introduction

With the "Tibetan Mastiff fever" in the 1980s, the massive drain of purebred Tibetan mastiffs led to the loss of germplasm resources and the complexity of breeds. Hence, the protection, utilization and research of Tibetan Mastiff are extremely urgent. Tibetan mastiff, one of the oldest and rarest dog breeds in the world, has been modified by deliberate selection in a relatively short period of time to adapt to the extremely high altitude of the Tibetan Plateau (typically 4500 m). The Tibetan Plateau is an extreme environment with a low-atmospheric oxygen pressure, a high level of ultraviolet radiation, a cold climate and limited resources [1]. With whole-genome sequencing technology, recent studies have indicated that the Tibetan Mastiff acquired various adaptive traits, such as a thick coat, enhanced metabolic capacity, high Hb–O_2 affinity, and low hemoglobin levels [2–9]. However, the genetic mechanisms of its adaptation to the high altitude are still limited to a few genes related to hypoxia response. It is reported that introgression could contribute to adaptive evolution [10]. Advance studies showed that the *Canis Lupus* (typically living in the Tibetan Plateau with 4500 m altitude) could provide a genetic source for the adaptation of Tibetan Mastiff by gene introgression [3,4]. Besides, recent mtDNA evidence revealed that the Tibetan Mastiff shares a common ancestor with other dogs deriving from the *Canis Lupus* [11,12]. However, several groups have demonstrated that the Tibetan Mastiff is more closely related with other Chinese native dog rather than *Canis Lupus* [3,13]. Therefore, how Tibetan Mastiff acquired the adaptive traits in high altitude remains elusive, and its origin also remains controversial.

Comparison of the breeds from different altitudes could powerfully dissect the physiological and genetic mechanisms of the Tibetan Mastiff adaptation to the high-altitude environment [6]. The Mongolia mastiff, living in the Hulunbuir Grassland (typically 800 m), is similar to the Tibetan Mastiff in terms of morphological characteristics, sports performance, *etc.* Here, we used the Mongolia Mastiff, resided at lower altitude, as a control for identifying the biological basis of Tibetan Mastiff adaptation to high altitude. There is a proposition that the Tibetan Mastiff was isolated in the Himalayan until they were enrolled into the Mongolian army of Genghis Khan, and then quickly spread across Eurasia in the early 13th century [13]. It is generally theorized that a consequence of this expansion promoted the germplasm of many famous large breeds such as mastiff-like dogs [12]. However, there is still no evidence showing that the germplasm of Mongolia Mastiff is also affected by Tibetan mastiff.

In this study, we *de novo* assembled the genome sequence of *Mongolia Mastiff*, *Tibetan Mastiff*, and *Canis Lupus* with excellent continuity at the

* Corresponding authors at: Research Center for High Altitude Medicine, School of Medical, Qinghai University, Xining 810016, PR China.
E-mail addresses: geriligao@hotmail.com (R. Ge), leileitang1984@163.com (F. Tang).

https://doi.org/10.1016/j.ygeno.2022.110359
Received 11 December 2021; Received in revised form 3 March 2022; Accepted 27 March 2022
Available online 29 March 2022
0888-7543/© 2022 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

contig and scaffold levels. In addition, we uncovered that Tibetan Mastiff was grouped with Mongolia Mastiff, and those two samples were more closely related with *Canis Lupus* than other species belonging to *Canidae, Canis*. In particular, InterPro classification of genes from significantly expanded and contracted gene family and from positively selected genes demonstrated significant enrichment in terms involved in DNA repair and damage response, ATPase activity, and fatty-acid α/β-oxidation for Tibetan Mastiff and *Canis Lupus* adaptation to high altitude environment with low oxygen, low temperature and strong UV.

2. Results

2.1. Genome assembly

We carried out *de novo* sequencing of *Mongolia Mastiff*, *Tibetan Mastiff*, and *Canis Lupus* to generate 229.60, 236.92, and 253.54 Gb data, respectively, using Illumina's paired-end sequencing platform (Supplementary Table 1). We filtered the raw data to obtain high-quality reads by discarding low-quality reads, including reads containing ambiguous bases more than 20% of total length, and reads containing adaptor sequences. After filtering, more than 96% and 90% of bases had Phred quality scores higher than 20 and 30, respectively. Besides, 203.88, 208.93, and 222.77 Gb clean bases were obtained for *de novo* assembly of *Mongolia Mastiff*, *Tibetan Mastiff*, and *Canis Lupus* genome, respectively (Supplementary Table 1). By employed supernova, Mongolia Mastiff had an estimated genome size of 2.31Gb with 1.49 K scaffolds (scaffold N50 length of 36.42 Mb and a contig N50 length of 107.07 Kb), Tibetan Mastiff had an estimated genome size of 2.32Gb with 1.49 K scaffolds (scaffold N50 length of 40.78 Mb and a contig N50 length of 107.94 Kb), and *Canis Lupus* had an estimated genome size of 2.32Gb with 1.49 K scaffolds (scaffold N50 length of 38.73 Mb and a contig N50 length of 116.24 Kb) (Table 1). The assemblies' statistics were similar to domestic dog genome [14]. In particular, the whole-genome assembly of a female *Canis Lupus* had been submitted to NCBI. We found that our assembly quality was slightly better than submitted with longer contig N50 and more accurate GC content on the initial assembly (Supplementary Table 2). Our smaller assembly size might be attributed to different size of sex chromosomes of *Canis Lupus*, as well as shorter scaffold N50, probably resulting from adopting different assembly software (Supplementary Table 2).

To assess assembly accuracy, we remapped filtered reads to the assembled genome. The Total mapping Rate of reads was above 99%, and the high Genome Coverage Rate indicated that the reads were evenly covered for *Mongolia Mastiff, Tibetan Mastiff,* and *Canis Lupus* genome (Supplementary Table 3). This implies that the current assembly covered almost all unique genomic regions. About 97.03%, 97.19%, and 97.28% of *Mongolia Mastiff, Tibetan Mastiff,* and *Canis Lupus* genome, respectively, were covered by at least 30× reads, which guaranteed a highly accurate assembly at the single-nucleotide level (Supplementary Table 3). The GC content of our three species were almost all concentrated between 20% and 80% (Supplementary Fig. 1). Thus, the *de novo* genomic assembly for Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus* were not strongly affected by GC-biased non-random sampling. The completeness of gene regions was further assessed using BUSCO, which demonstrated that 95.4%, 95.6%, and 95.7% of 4104 highly conserved single-copy orthologous genes were present in Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus* assembly, respectively (Supplementary Table 4). These results suggested that the assembly quality of Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus* were markedly high, thus ensuring the reliability of subsequent comparative genomic analyses.

Table 1
Genome assembly of *M. mastiff, T. mastiff,* and *C. lupus.*

	Mongolia_mastiff	Tibetan_mastiff	Canis_lupus
Number of long scaffolds(K)	1.49	1.76	1.85
Edge N50(Kb)	17.52	18.16	14.24
Contig N50(Kb)	107.07	107.94	116.24
Phaseblock N50(Mb)	3.59	4.93	1.30
Scaffold N50(Mb)	36.42	40.78	38.73
Missing 10Kb(%)	1.83	1.81	1.59
Assembly size(Gb)	2.31	2.32	2.32

2.2. Genome annotation

We identified repetitive regions in the genomes to discern assembly quality and to prepare the genome for annotation. We found that about 762, 789, and 889 Mb repeat sequences were finally revealed, which accounted for 31.26%, 32.23%, and 36.48% of *Mongolia Mastiff*, *Tibetan Mastiff*, and *Canis Lupus* genome, respectively (Supplementary Table 5). Repeat content of our three assemblies were qualitatively similar to domestic dog genome [14]. The long interspersed nuclear elements (LINEs) are the top category of repetitive elements among those three genomes. We employed *de novo*, homology- and transcriptome sequencing- based methods to predict the protein-coding genes in *Mongolia Mastiff*, *Tibetan Mastiff*, and *Canis Lupus* genome. We successfully generated 17,675/ 17,953/ 17,668 annotated genes of Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus*, respectively, which were very similar between our three individuals and the domestic dog (Table 2) [14]. We also annotated the gene structures of noncoding RNAs, including microRNA, rRNA, and tRNA, *etc.* The ratios of all categories were nearly consistent among our three assemblies (Supplementary Table 6). According to homologous searches against five databases, a total of 37,807 genes (99.17%), 38,492 genes (98.86%), and 36,986 genes (99.20%) were functionally annotated using at least one public database for *Mongolia Mastiff*, *Tibetan Mastiff*, and *Canis Lupus* genome, respectively (Table 3). Gene Ontology (GO) was adopted to annotate protein function based on conserved protein domains and functional sites, which was involved in biology process (BP), component of cell (CC), or molecule function (MF). The GO annotation showed that the categories were nearly consistent among three species, except that there were one unique term involved biological phase for *Canis Lupus*, as well as two unique terms involved in virion and nutrient reservoir activity for both Tibetan Mastiff and *Canis Lupus* (Supplementary Fig. 2). In addition, the gene number in each common category is also similar among Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus*.

2.3. Phylogenetic tree construction

Using the sequence similarities among protein- coding genes among the species, we identified 17,674/ 17,953/ 17,669 genes in 14,845/ 15,085/ 14,833 families of Mongolia mastiff/ Tibetan mastiff/ *Canis Lupus*, respectively (Fig. 1A). We classified 3658/ 3658/ 3658 single-copy genes and 6749/ 6789/ 6785 multiple-copy genes in the Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus*, respectively (Fig. 1A). Using the 3658 orthologous clusters that were shared among all the organisms to generate a maximum likelihood phylogenetic tree, our genomic analysis denoted that Tibetan Mastiff and Mongolia Mastiff belonged to the *Carnivora, Canidae, Canis, Canis Lupus* in the animal taxonomic status (Fig. 1B). Strikingly, we found that Tibetan Mastiff was most closely related to Mongolia mastiff, and these two species in turn formed a clade only with *Canis Lupus* (Fig. 1B). This result suggested that though the natural habitats of these three species were significant diverse, they were obviously most close among *Canis Lupus familiaris* with genome annotation.

2.4. Gene family expansion and contraction

The entire process of species evolution is accompanied by the contraction and expansion of gene families, which has been proposed as a major mechanism underlying adaptive evolution [15]. Since the phylogenomic datasets provide better precision and accuracy in estimating the timescale of placental mammal phylogeny, we adopted the

Table 2
Protein-coding gene annotation in the *M. mastiff, T. mastiff,* and *C. lupus* genome.

Type	Gene number	mRNA number	avRNA Length(bp)	Exon number	AvExon length(bp)	Intron number	avIntron Length(bp)
Mongolia_mastiff							
De novo	81,597	81,597	12,078.8	298,624	238.4	217,027	4213.3
Transcript-ome	55,280	55,280	73,968.5	590,155	164.6	534,875	5323.4
EVM integration	17,877	17,877	33,203.1	162,121	180.6	144,244	3912.0
PASA update	17,675	38,121	61,163.1	465,957	164.2	427,836	4644.9
Tibetan_mastiff							
De novo	81,223	81,223	11,878.5	296,030	237.4	214,807	4164.3
Transcript-ome	55,781	55,781	73,539.6	585,616	162.2	529,835	5316.7
EVM integration	18,173	18,173	32,029.0	162,198	180.1	144,025	3838.6
PASA update	17,953	38,934	59,841.9	465,344	164.7	426,410	4593.1
Canis_lupus							
De novo	77,776	77,776	15,103.4	294,666	237.6	216,890	5093.3
Transcript-ome	54,080	54,080	73,885.7	562,582	165.2	508,502	5325.9
EVM integration	17,874	17,874	33,198.6	160,355	180.6	142,481	3961.4
PASA update	17,668	37,283	59,956.7	443,828	164.2	406,546	4666.8

Table 3
Functional annotation of protein-coding genes for *M. mastiff, T. mastiff,* and *C. lupus* genome.

Species	Annotation database	Annotated Number	Percentage (%)*
Mongolia_mastiff	NR	36662	96.17%
	Swiss-prot	34773	91.22%
	INTERPRO	37700	98.89%
	GO	28373	74.43%
	KEGG	25091	67.94%
	All Annotated	37807	99.17%
Tibetan_mastiff	NR	37367	95.98%
	Swiss-prot	35442	91.04%
	INTERPRO	38378	98.57%
	GO	28951	74.36%
	KEGG	25583	67.23%
	All Annotated	38492	98.86%
Canis_lupus	NR	35844	96.14%
	Swiss-prot	34130	91.54%
	INTERPRO	36875	98.90%
	GO	27872	74.76%
	KEGG	24623	66.05%
	All Annotated	36986	99.20%

* Percentage(%): The percentage of genes annotated to the database to the total number of genes.

whole genome sequence data of eight mammal species (including *Odobenus Rosmarus Divergens, Neomonachus Schauinslandi, Felis Catus, Pantholops Hodgsonli, Bos Mutus, Homo Saplens, Mus Musculus,* and *Sarcophilus Harrisii*), *Boxer Dog* (reference genome), and our three samples to encompass the breadth of mammalian phylogeny (Supplementary Table 7–8) [16]. Then, we integrated the fossil-based temporal constraints to generate a phylogenetic tree to obtain contracted and expanded gene families and genes using CAFÉ [17]. We identified 225, 262, and 265 gene families that were expanded and 1574, 1897, and 1687 contracted in Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus*, respectively (Supplementary Table 9). In addition, 27, 14, and 23 gene families were significantly expanded and 14, 14, and 10 contracted in the Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus*, respectively (Supplementary Table 9).

Based on GO annotations, expanded gene families of Tibetan Mastiff and *Canis Lupus* were highly enriched in DNA repair and damage response, and contracted were highly enriched in ATPase activity (Fig. 2 and Table 4). However, there was not enrichment on above categories in Mongolia Mastiff (Fig. 2 and Table 4). The expansion in DNA repair and damage response should be an adaptive trait for Tibetan Mastiff and *Canis Lupus* exposure to strong UV radiation of high altitude. In addition, expanded and contracted gene families of both Tibetan Mastiff and Mongolia Mastiff were highly enriched in multiple receptor activity and neurological system process (Fig. 2 and Table 4). Interestingly, the expanded and contracted gene families of Tibetan Mastiff were specifically enriched in ion channel activity, while Mongolia Mastiff were particularly enriched in sensory system, tyrosine kinase activity, and transducer activity, *etc.* (Fig. 2).

2.5. Positive selection on single- copy genes

If the non-synonymous to synonymous substitution rates (Ka/Ks) is bigger than 1, it is supposed that this gene is subject to positive selection in the evolution, and non-synonymous mutation of this gene is retained as an advantage for adaptation. We select the Boxer Dog genomic sequences as a reference for positive selection analysis. We classified 9480 contained single-copy orthologous genes in the Mongolia Mastiff/ Tibetan Mastiff/ *Canis Lupus* (Supplementary Table 10). Estimating the Ka/Ks based on orthologous, we identified 316, 338, and 317 positively selected genes (PSGs) in the Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus* genome, respectively (Supplementary Table 11).

Functional annotation of the most significant PSGs related to adaptation in the Mongolia Mastiff genome were enriched in the regulation of RNA polymerase II transcription and ubiquitin system (Table 5). PSGs classified in the regulation of RNA polymerase II transcription included *p65, ATF3, CUX2, EHF, DDN, TLR4* precursor*, ASXL1, IGBP1, AKIRIN2, Loc4981817,* and *GNL3*. The proteins coding by *p65, ATF3,* and *CUX2* are the transcription factors [18–20], by *EHF* and *ASXL1* could modulate the transcription of certain genes [21,22], by *TLR4* precursor and *IGBP1* contributed to signal transduction [23,24], as well as the protein coding by *GNL3* probably interacts with p53 to affect tumorigenesis and stem cell proliferation [25]. Besides, another PSG in Mongolia Mastiff genome was *AMFR*, which encodes a receptor as a member of the E3 ubiquitin ligase family to catalyze ubiquitination and endoplasmic reticulum-associated degradation of specific proteins [26].

Ka/Ks analyses of *Canis Lupus* adaptation for different GO categories revealed an enrichment in lipid metabolism, olfactory receptor activity and ubiquitin system (Table 5). PSGs classified in the lipid metabolism are related to fatty-acid catabolism and anabolism, and peroxisomal α/β-oxidation of fatty-acid. The proteins coding by *P450* and *ACSL1* participate in lipid biosynthesis [27,28], as well as by *PECR* and TECR are the key enzyme of fatty-acid chain elongation in peroxisome and endoplasmic reticulum, respectively [29,30]. Besides, the *trans*-2-enoyl-CoA reductase (TER), encoded by the *PECR* gene, is also involved in peroxisomal α-oxidation system *via* cacatalyzing phytenoyl-CoA into phytanoyl-CoA [31]. Strikingly, the peroxisomal bifunctional enzyme isoform X1 (XP_545234.1 in NCBI), positively selected, is known as a

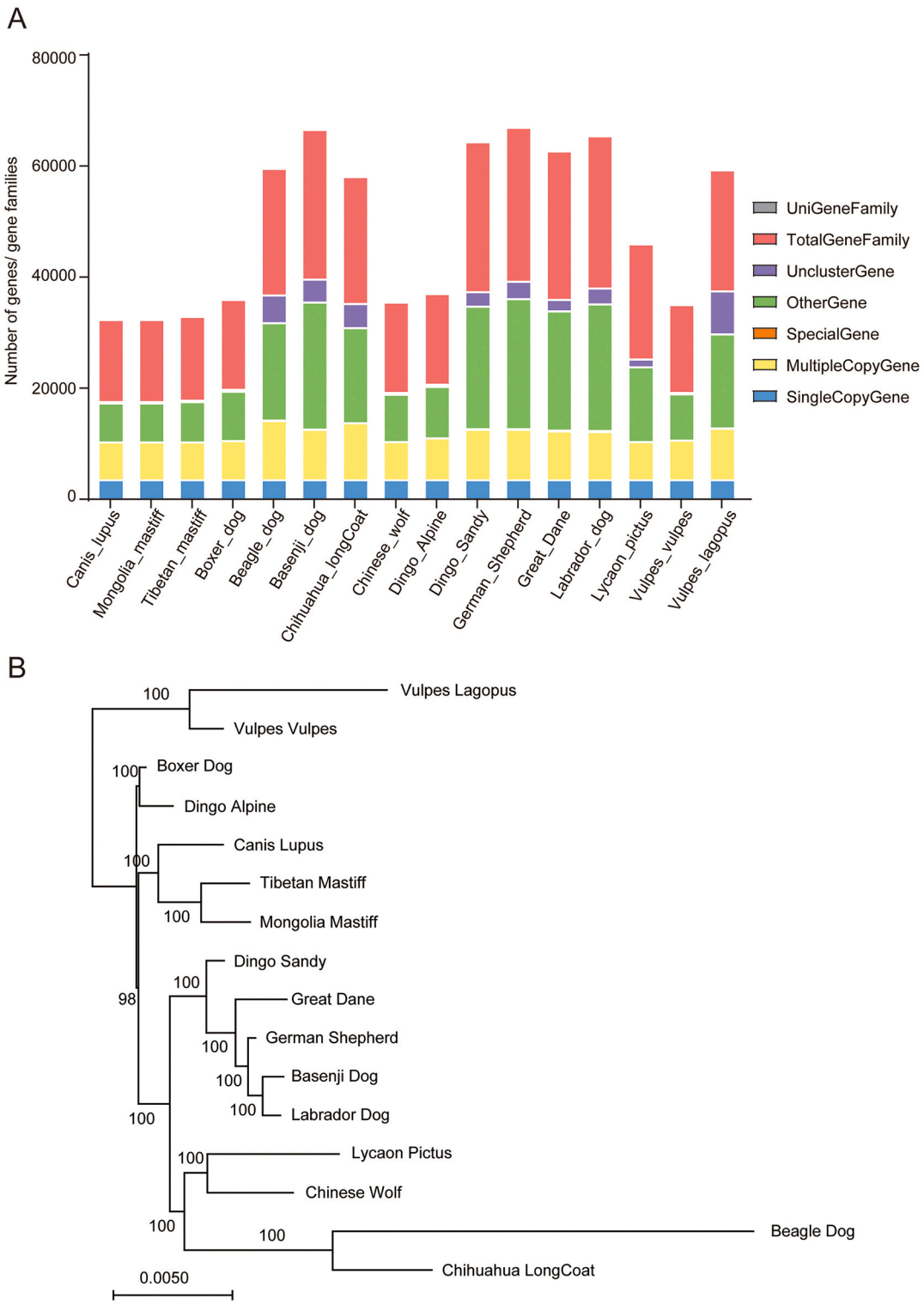

Fig. 1. Evolutionary analyses of the *M. mastiff, T. mastiff,* and *C. lupus* genome. (A) Distribution of the number of genes and gene families in each species. The horizontal axis is the species, and the vertical axis is the number of corresponding genes. The blue column represents single-copy genes, orange represents multiple-copy genes, red represents species-specific genes, green represents genes other than species-common and unique genes, and purple represents genes that are not clustered into groups. (B) Phylogenetic analysis of *M. mastiff, T. mastiff, C. lupus,* and closely related dog species. The branch length represents the rate of evolution, and the value on the branch represents the number of Bootstrap support (when tree-making, set the Bootstrap to 100, and the support number is more than 70 to be more reliable). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

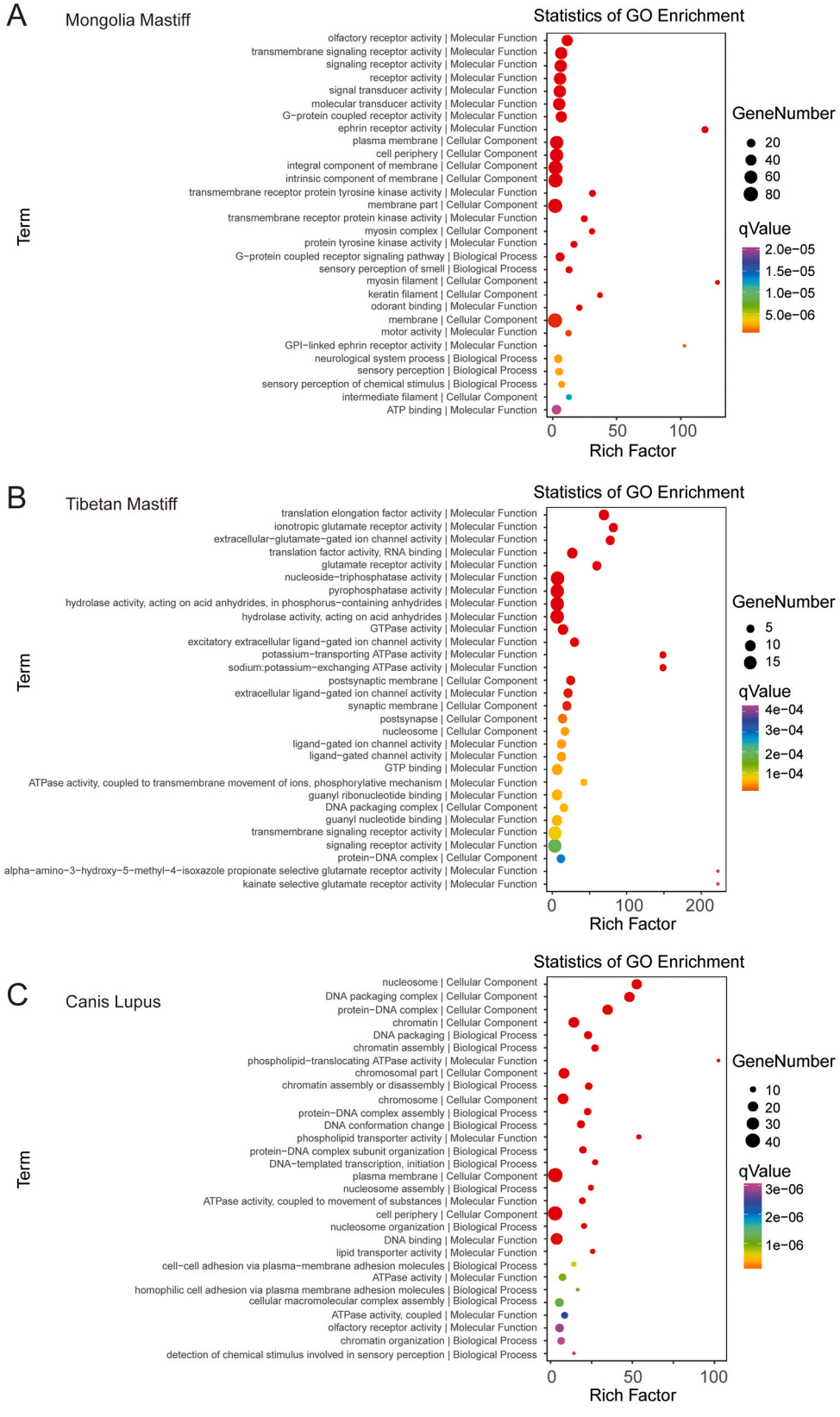

Fig. 2. Comparative genomic analysis of *M. mastiff, T. mastiff,* and *C. lupus*. Functional enrichment annotation of the most significantly expansive and contract gene families in *M. mastiff* (A), *T. mastiff* (B), and *C. lupus* (C) genome. The horizontal axis Rich Factor represents the ratio of Input frequency/Background frequency in the enrichment analysis, and the vertical axis represents GO term. The size of the bubble indicates the number of mutant genes annotated to this GO term, and the color corresponds to the q-value in the enrichment analysis.

crucial enzyme that participates in hydration and dehydrogenation of long-chain fatty acid β-oxidation [32]. Besides, some PSGs related to olfactory receptor activity and ubiquitin system were classified into integral component of membrane in Table 5.

A GO functional classification of the most significant PSGs related to adaptation in the Tibetan Mastiff genome were enriched in the regulation of RNA polymerase II transcription and response to hypoxia (Table 5). PSGs classified in the regulation of RNA polymerase II transcription included *p65*, *ATF3*, *EHF*, Loc4981817, *ZNF280D, MYPOP, ATF5,* and *FOXI2*. It is notable that the *p65, ATF3*, *EHF*, and *Loc4981817*

Table 4
Functional annotation of the most significantly expansive and contract gene families related to adaption for *M. mastiff, T. mastiff,* and *C. lupus* genome.

Species	GO terms	Input no.	Backgro-und no.	*P* value
Mongoliamastiff	receptor activity	58	1280	2.17E-31
	integral component of membrane	79	4417	1.21E-17
	ATP binding	27	1208	4.17E-07
	protein autophosphorylation	5	132	3.68E-03
	axon	5	154	7.03E-03
	neuromuscular junction	2	18	8.45E-03
	wound healing	5	190	1.63E-02
	visual perception	3	83	2.68E-02
	digestive tract morphogenesis	2	33	2.70E-02
	cell-substrate adhesion	4	165	3.98E-02
Tibetan mastiff	receptor activity	17	1245	2.65E-05
	neuronal cell body	3	108	1.27E-02
	blood microparticle	2	61	3.08E-02
	DNA binding	10	1374	8.63E-02
	sequence-specific DNA binding transcription factor activity	6	808	1.55E-01
	transcription, DNA-templated	11	1912	2.36E-01
	integral component of membrane	23	4418	2.39E-01
	regulation of neuron death	1	106	3.81E-01
	axon	1	155	5.04E-01
	lipid binding	2	386	5.21E-01
Canis lupus	DNA binding	29	1342	6.00E-11
	integral component of membrane	42	4414	4.93E-05
	embryo implantation	3	25	3.51E-04
	nucleus	36	4015	9.13E-04
	RNA polymerase II core promoter proximal region sequence-specific DNA binding	6	202	9.15E-04
	receptor activity	16	1255	1.37E-03
	transcription, DNA-templated	20	1855	2.50E-03
	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription	4	134	6.64E-03
	ATP binding	13	1175	1.28E-02
	intracellular membrane-bounded organelle	44	6247	2.73E-02

Table 5
Functional categories of the most significant PSGs related to adaption in *M. mastiff, T. mastiff,* and *C. lupus* genome.

Species	GO terms	Input no.	Backgro-und no.	*P* value
M_B*	cerebral cortex development	4	58	1.63E-02
	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription	4	72	3.29E-02
	RNA polymerase II transcription regulatory region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription	5	110	3.80E-02
	artery smooth muscle contraction	1	4	6.56E-02
	positive regulation of transcription from RNA polymerase II promoter	12	456	8.42E-02
	peptidyl-proline 4-dioxygenase activity	1	6	9.67E-02
	negative regulation of appetite	1	6	9.67E-02
	RNA polymerase II core promoter proximal region sequence-specific DNA binding	6	200	1.22E-01
	receptor activity	27	1280	1.30E-01
	positive regulation of pri-miRNA transcription from RNA polymerase II promoter	1	10	1.56E-01
T_B	hypoxia-inducible factor-1alpha signaling pathway	1	3	5.19E-02
	RNA polymerase II transcription regulatory region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription	4	110	1.29E-01
	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription	3	72	1.33E-01
	cellular response to interleukin-1	2	42	1.69E-01
	sequence-specific DNA binding RNA polymerase II transcription factor activity	10	415	1.97E-01
	cellular response to hypoxia	2	53	2.39E-01
	glucose transport	2	54	2.46E-01
	sequence-specific DNA binding transcription factor activity	17	808	2.57E-01
	receptor activity	25	1245	2.73E-01
	response to hypoxia	3	104	2.77E-01
C_B	monoterpenoid metabolic process	1	2	3.27E-02
	mitochondrion	23	1022	8.07E-02
	positive regulation of phosphatidylinositol 3-kinase signaling	2	38	1.30E-01
	drug metabolic process	1	9	1.39E-01
	integral component of membrane	81	4414	1.42E-01
	cell-substrate adhesion	5	167	1.43E-01
	fatty acid beta-oxidation	2	41	1.47E-01
	fatty acid transport	2	43	1.58E-01

(*continued on next page*)

Table 5 (*continued*)

Species	GO terms	Input no.	Backgro-und no.	*P* value
	regulation of catalytic activity	16	773	2.09E-01
	mitochondrion organization	8	354	2.32E-01

[*] M_B: Functional categories of the positive selected genes between Mongolia_mastiff and Boxer_dog.

were positively selected in both Mongolia Mastiff and Tibetan Mastiff. In addition, several genes involved in response to hypoxia were identified as being under positive selection pressure in Tibetan Mastiff. The liver fatty acid binding protein (FABPL), encoded by the *FABPL* gene, is known as a major fatty acids chaperone protein to facilitate uptake, intracellular transport, as well as mitochondrial and peroxisomal α/β-oxidation of fatty acids [33]. Another PSG, *TERT* encodes telomerase reverse transcriptase (TERT), which is one of the subunits of telomerase, the ribonucleoprotein complex [34]. The TERT expression is responsible for telomerase activity [35]. The classical activity of telomerase is to add the guanine-rich repetitive sequences to the chromosome ends and thus maintains telomere length to ensure genome stability in normal cells. The uncanonical role of telomerase is related to tumor development, apoptosis, and DNA repair. Shin et al. have revealed that the telomerase could accelerate the efficiency of DNA repair for different types of DNA damage by recruiting DNA repair proteins to the damaged DNA sites [36].

3. Discussion and conclusion

The Tibetan Plateau is an extreme environment with dry, cold climate, low oxygen, and strong UV radiation. The harsh conditions result in the evolution of organisms to well- adapted to high altitude [37]. Tibetan Mastiff takes a relatively short period to adapt to the extremely high altitude of the Tibetan Plateau. The physiological and genetic mechanisms of its adaptation to high altitude have always been the focus. However, the genetic mechanisms of its adaptation are still limited to a few genes related to hypoxia adaptation, such as *EPAS1* and *HBB* [2–9]. Besides, the origin of Tibetan Mastiff also remains controversial [38]. In this study, we provided a *de novo* genome of Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus*, as well as carried out a series of comparisons with other species to reveal the origin and underlying mechanisms for Tibetan Mastiff adaptation to high altitude.

we *de novo* assembled the genome sequence of *Mongolia Mastiff*, *Tibetan Mastiff*, and *Canis Lupus* with excellent continuity at the contig/scaffold N50, mapping rate, GC depth, and completeness. The ratio/number of repetitive sequences, predicted genes, and annotated protein sequences in Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus* were nearly consistent with domestic dog. The high- quality genome could provide a solid foundation for population and conservation studies, as well as for investigation of how the species acquired the adaptive traits.

Several groups have indicated that the Tibetan Mastiff derives from the *Canis Lupus* with mtDNA sequencing, while others revealed that the Tibetan Mastiff is more closely related with other Chinese native dog rather than *Canis Lupus* with mtDNA and whole-genome sequencing [3,11–13]. It is reported that modern wolves probably arise from the expansion of a population in Northeast of Sibera, which replaced other ancient wolf populations worldwide [39]. The dogs might derive from the ancient wolf populations, and interacted with the modern wolf population through admixture to obtain the adaptive traits in high altitude [39]. Our phylogenic tree indicated that the Tibetan Mastiff clustered with Mongolia Mastiff, and both species were most closely related to *Canis Lupus*. Furthermore, the Tibetan Mastiff and Mongolia Mastiff clade was clustered with currently annotated species belonged to the *Carnivora, Canidae, Canis, Canis Lupus,* which suggested that Tibetan Mastiff and Mongolia Mastiff were derived from *Canis Lupus*.

The UV radiation on Qinghai-Tibet Plateau is among the highest worldwide, and the strong UV radiation could induce direct damage to DNA [40]. In addition, hypoxia at high altitude causes increased production of reactive oxygen species (ROS), which directly damage the DNA. Functional categories showed that the expanded gene families of Tibetan Mastiff and *Canis Lupus* were highly enriched in DNA repair and damage response, which should be an adaptive trait for correcting UV- and ROS- induced DNA damage. Besides, upon hypoxia, the ATP production is limited for reducing mitochondrial oxidative phosphorylation [41]. Maintaining ATP homeostasis to adapt hypoxia, cells might suppress the ATPase activity to decrease ATP consumption [42]. Our research revealed that the contraction in ATPase activity of both Tibetan Mastiff and *Canis Lupus* was conducive to the survival and adaptation in hypoxia stress. These results were consistent with our previous finding in Tibetan Antelope that positively selected genes were enriched in the ATPase and DNA repair categories to adapt to harsh highland environments [43].

Population and comparative genomic studies of high-altitude animals and humans have revealed numerous genes for high-altitude adaptation, many of which are involved in hypoxia adaptation, related to hypoxia-inducible factor (HIF) pathway and hemoglobin (Hb) variants [44,45]. Studies about the HIF pathway are focused on *EPAS1*, encoding endothelial PAS domain containing protein 1 (EPAS1) that is a O2-regulated subunit of the HIF-2α, while about the Hb variants are focused on amino acid mutations of *HBB* (β-globin) gene cluster that encodes the β-type subunits of Hb isoforms. Previous genomic studies of Tibetan Mastiff also revealed clear evidence for positive selection on *EPAS1* and β-globin gene cluster [5,6]. The striking signature of positive selection on both *EPAS1* and β-globin gene cluster of Tibetan Mastiff are attributable to introgressive hybridization from *Canis Lupus* [3,4,9]. Notably, our study identified a novel PSG, *TERT*, in the Tibetan Mastiff genome for hypoxia adaptation. The protein encoded by *TERT* is crucial for accelerating the efficiency of DNA repair for different types of DNA damage, caused by strong UV radiation and hypoxia on Qinghai-Tibet Plateau, *via* recruiting DNA repair proteins to the damaged DNA sites. In addition, the protein encoded by *TERT* is one of the subunits of telomerase, the classical activity of which is to maintain telomere length for ensure genome stability. Interestingly, previous study showed that the reduced telomere length was negatively related to high-altitude pulmonary edema /maladaptation, which indirectly support that the positive selection of *TERT* contributes to hypoxia adaptation of Tibetan Mastiff to high altitude [46]. Additional confirmatory studies are required to determine the role and mechanism of TERT in high-altitude adaptation.

The cold climate and low oxygen are hostile to sustain prolonged thermogenesis of mammal native to high altitude by increasing thermoregulatory costs and limiting both thermogenesis and aerobic exercise capacity, respectively. Previous studies have shown that the capacity of fatty-acid oxidation is significantly enhanced in high-altitude deer mice and ground tit to promote thermogenic endurance under cold stress conditions [37,47]. The enhanced capacity of thermogenesis should be another adaptive trait for Tibetan Mastiff and *Canis Lupus* exposure to hypoxic cold stress of high altitude. The PSGs analysis of *Canis Lupus* genome revealed for the first time that the positively selected genes, encoding TER and peroxisomal bifunctional enzyme isoform X1, are critical for promoting peroxisomal fatty-acid α/β-oxidation to sustain prolonged thermogenesis. In addition to enhanced capacity of fatty-acid oxidation, an enhanced capacity to uptake and transport fatty acids also greatly elevate thermogenic capacity and endurance [48]. Our study first demonstrated that the positively selected FABPL of Tibetan Mastiff combines with fatty acids to facilitate fatty-acid uptake, intracellular transport, as well as mitochondrial and peroxisomal α/β-oxidation, and thus promote the thermogenic capacity for a prolonged period. Under conditions of chronic O2 deprivation and cold climate at high altitude, the strong selection in enhancement in

fatty-acid oxidation capacity may also contributes to high-altitude adaptation of *Canis Lupus* and Tibetan Mastiff.

Our study obtained a draft genome for Mongolia Mastiff, Tibetan Mastiff, and *Canis Lupus*. The comparative genomic analyses have suggested that both Mongolia Mastiff and Tibetan Mastiff belonged to *Canis Lupus* species, as well as our three samples were closely related. Although further experimental verification is needed, our three assembled genomes provided useful genomic resources for further investigating adaptation to harsh and extreme highland environments.

4. Materials and methods

4.1. Ethical statements

Animals were handled and blood samples were collected in accordance with regulations of the Animal Experimental and Medical Ethics Committee of the Qinghai University Medical College, Qinghai University.

4.2. Materials and genome sequencing

The blood sample used for the 10× Genomics library construction was acquired from male *Mongolia Mastiff*, *Tibetan Mastiff*, and *Canis Lupus* captured in Hulun Buir Grassland (800 m), Inner Mongolia, China, in Yushu (3780 m), Qinghai, China, and in Hoh Xil National Nature Reserve (4560 m), Qinghai, China, respectively. Blood DNA was isolated using Qiagen DNA purification kit (Qiagen, Valencia, CA, USA) following the manufacturer's instructions. Subsequent sequencing, assembly and analysis were conducted at Capital Bio Corporation (Beijing, China). A paired-end sequencing library with an insertion length of 150 bp was constructed for each individual genome sequencing according to the Illumina protocol (Illumina, San Diego, CA, USA). The HTQC package was adopted to filter low-quality bases and reads [49].

4.3. Scaffold assembly and assembly quality assessment

Illumina paired-end reads were assembled into contigs by employing supernova 2.0 assembler [50]. The sequencing coverage and GC content distribution of the assembled genome sequence were evaluated by mapping all sequencing reads back to the scaffolds using BWA [51]. The completeness of genome assembly was evaluated using BUSCO by searching the single copy orthologs set against the assembled genome [52].

4.4. Repeat identification, gene prediction, and annotation

RepeatMasker was performed for homologous comparison by searching against Repbase database to identify known repeats and low complexity DNA sequences [53]. The repeat elements of assembled genome were *de novo* predicted using RepeatModeler [54]. Employing a comprehensive strategy to annotate protein-coding genes by combining homology-based predictions, *ab initio* predictions, and RNA seq-based prediction methods. Augustus, SNAP, GeneMark-ET were used for *ab initio* prediction, Genewise was adopted for homology annotation, and PASA was employed to assembly RNA-seq reads into transcripts [55–57]. The gene sets predicted by above three approaches were integrated with EVM, and then update the gene sets by PASA. Based on the Rfam database, we adopted the cmscan program in the software INFERNAL to perform a genome-wide comparison to identify non-coding RNAs (ncRNAs) [58]. The structures of tRNAs and rRNAs were predicted by tRNAscan-SE and RNAmmer, respectively [59]. We functionally annotated the protein-coding genes according to homologous searches against five databases of NR, Swiss-prot, Interpro, KEGG and Gene Ontology [60–64].

4.5. Gene family construction

The genomes of 13 species of vertebrate were downloaded from NCBI. Only the longest transcript was selected for each gene locus with alternative splicing variants. According to the filtered sequence, we performed the all-*vs*-all BLAST to obtain the similarity of protein sequences among 13 species and our three samples. Orthologous groups were constructed by ORTHOMCL using the default settings based on the filtered BLAST results [65].

4.6. Phylogenetic tree construction

Single-copy gene families were retrieved from the ORTHOMCL results and used for phylogenetic tree construction. Before building a phylogenetic tree, it is necessary to employ the software ProtTest to select the best amino acid substitution model [66]. ProtTest can estimate the maximum likelihood of model parameters through AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) or DT (Decision Theory Criterion) to sort various models to find the best model. The best evolution model for this study is "VT + I + G + F". Finally, a phylogenetic tree was constructed by RAxML [67]. The model is "PROTGAMMAIVTF model", and the bootstrap is 100.

4.7. Gene family expansions and contractions

Expansion and contraction of gene clusters were determined using CAFÉ [17]. The phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny [16]. Hence, a phylogenetic tree was used in CAFÉ by adopting the whole genome sequence data of eight mammal species, Boxer Dog (reference genome), and our three samples to encompass the breadth of mammalian phylogeny, together with fossil-based temporal constraints.

4.8. Positively selected genes

Use ORTHOMCL to obtain the orthologous genes of Mongolia Mastiff, Tibetan Mastiff, *Canis Lupus* and Boxer Dog (reference genome), and compare and analyze the Mongolia Mastiff/ Tibetan Mastiff/ *Canis Lupus* to Boxer Dog according to the following steps. First, use the MUSCLE software to compare and analyze the amino acid sequences corresponding to genes that are orthologous to each other between the two samples, and obtain the corresponding CDS sequence comparison results [68]. Then, use the yn00 tool in PAML to calculate the Ka and Ks values between gene pairs that are orthologous genes [69]. Finally, if $Ka/Ks > 1$, it is considered that the evolution of this gene is affected by positive selection, and the non-synonymous mutation of this gene is retained as an advantage.

Data availability

The sequencing data generated in this study are available at NCBI under BioProject ID PRJNA826916.

CRediT authorship contribution statement

Chunmei Cai: Formal analysis, Visualization, Writing – original draft, Funding acquisition. **Yingzhong Yang:** Resources. **Qin Ga:** Resources. **Guocai Xu:** Resources. **Rili Ge:** Supervision, Writing – review & editing, Funding acquisition. **Feng Tang:** Supervision, Writing – review & editing, Funding acquisition.

Acknowledgment

This work was supported by National Natural Science Foundation of China (No. 82072107, 31571231, 81860299, 81960292).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2022.110359.

References

[1] T. Wu, The Qinghai-Tibetan plateau: how high do Tibetans live? High Alt. Med. Biol. 2 (2001) 489–499, https://doi.org/10.1089/152702901753397054.

[2] Y. Li, et al., Population variation revealed high-altitude adaptation of Tibetan mastiffs, Mol. Biol. Evol. 31 (2014) 1200–1205, https://doi.org/10.1093/molbev/msu070.

[3] B. Miao, Z. Wang, Y. Li, Genomic analysis reveals hypoxia adaptation in the Tibetan mastiff by introgression of the gray wolf from the Tibetan plateau, Mol. Biol. Evol. 34 (2017) 734–743, https://doi.org/10.1093/molbev/msw274.

[4] A.V. Signore, et al., Adaptive changes in hemoglobin function in high-altitude Tibetan canids were derived via gene conversion and introgression, Mol. Biol. Evol. 36 (2019) 2227–2237, https://doi.org/10.1093/molbev/msz097.

[5] G.D. Wang, et al., Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the tibetan plateau, Genome Biol Evol 6 (2014) 2122–2128, https://doi.org/10.1093/gbe/evu162.

[6] X. Gou, et al., Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia, Genome Res. 24 (2014) 1308–1315, https://doi.org/10.1101/gr.171876.113.

[7] M.C. van Patot, M. Gassmann, Hypoxia: adapting to high altitude by mutating EPAS-1, the gene encoding HIF-2alpha, High Alt. Med. Biol. 12 (2011) 157–167, https://doi.org/10.1089/ham.2010.1099.

[8] J.F. Storz, G.R. Scott, Z.A. Cheviron, Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates, J. Exp. Biol. 213 (2010) 4125–4136, https://doi.org/10.1242/jeb.048181.

[9] B. vonHoldt, Z. Fan, D. Ortega-Del Vecchyo, R.K. Wayne, EPAS1 variants in high altitude Tibetan wolves were selectively introgressed into highland dogs, PeerJ 5 (2017), e3522, https://doi.org/10.7717/peerj.3522.

[10] P.W. Hedrick, Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation, Mol. Ecol. 22 (2013) 4606–4618, https://doi.org/10.1111/mec.12415.

[11] J.F. Pang, et al., mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves, Mol. Biol. Evol. 26 (2009) 2849–2864, https://doi.org/10.1093/molbev/msp195.

[12] Q. Li, et al., Origin and phylogenetic analysis of Tibetan mastiff based on the mitochondrial DNA sequence, J Genet Genomics 35 (2008) 335–340, https://doi.org/10.1016/S1673-8527(08)60049-1.

[13] L. Yan, Z. Yaping, High genetic diversity of Tibetan mastiffs revealed by mtDNA sequences, Chin. Sci. Bull. 57 (2012) 1483–1487, https://doi.org/10.1007/s11434-012-4995-4.

[14] K. Lindblad-Toh, et al., Genome sequence, comparative analysis and haplotype structure of the domestic dog, Nature 438 (2005) 803–819, https://doi.org/10.1038/nature04338.

[15] J.T. Li, et al., Comparative genomic investigation of high-elevation adaptation in ectothermic snakes, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) 8406–8411, https://doi.org/10.1073/pnas.1805348115.

[16] M. dos Reis, et al., Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny, Proc. Biol. Sci. 279 (2012) 3491–3500, https://doi.org/10.1098/rspb.2012.0683.

[17] T. De Bie, N. Cristianini, J.P. Demuth, M.W. Hahn, CAFE: a computational tool for the study of gene family evolution, Bioinformatics 22 (2006) 1269–1271, https://doi.org/10.1093/bioinformatics/btl097.

[18] S. Ghosh, M.J. May, E.B. Kopp, NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses, Annu. Rev. Immunol. 16 (1998) 225–260, https://doi.org/10.1146/annurev.immunol.16.1.225.

[19] H. Gingras, O. Cases, M. Krasilnikova, G. Berube, A. Nepveu, Biochemical characterization of the mammalian Cux2 protein, Gene 344 (2005) 273–285, https://doi.org/10.1016/j.gene.2004.11.008.

[20] M. Rohini, A. Haritha Menon, N. Selvamurugan, Role of activating transcription factor 3 and its interacting proteins under physiological and pathological conditions, Int. J. Biol. Macromol. 120 (2018) 310–317, https://doi.org/10.1016/j.ijbiomac.2018.08.107.

[21] M. Trojanowska, Ets factors and regulation of the extracellular matrix, Oncogene 19 (2000) 6464–6471, https://doi.org/10.1038/sj.onc.1204043.

[22] O. Abdel-Wahab, A. Dey, The ASXL-BAP1 axis: new factors in myelopoiesis, cancer and epigenetics, Leukemia 27 (2013) 10–15, https://doi.org/10.1038/leu.2012.288.

[23] M. Kong, D. Ditsworth, T. Lindsten, C.B. Thompson, Alpha4 is an essential regulator of PP2A phosphatase activity, Mol. Cell 36 (2009) 51–60, https://doi.org/10.1016/j.molcel.2009.09.025.

[24] A. Ciesielska, M. Matyjek, K. Kwiatkowska, TLR4 and CD14 trafficking and its influence on LPS-induced pro-inflammatory signaling, Cell. Mol. Life Sci. 78 (2021) 1233–1261, https://doi.org/10.1007/s00018-020-03656-y.

[25] R.Y. Tsai, R.D. McKay, A nucleolar mechanism controlling cell proliferation in stem cells and cancer cells, Genes Dev. 16 (2002) 2991–3003, https://doi.org/10.1101/gad.55671.

[26] Z. Chen, S. Du, S. Fang, gp78: a multifaceted ubiquitin ligase that integrates a unique protein degradation pathway from the endoplasmic reticulum, Curr. Protein Pept. Sci. 13 (2012) 414–424, https://doi.org/10.2174/138920312802430590.

[27] C. Handschin, et al., Cholesterol and bile acids regulate xenosensor signaling in drug-mediated induction of cytochromes P450, J. Biol. Chem. 277 (2002) 29561–29567, https://doi.org/10.1074/jbc.M202739200.

[28] K. Nakahara, et al., The Sjogren-Larsson syndrome gene encodes a hexadecenal dehydrogenase of the sphingosine 1-phosphate degradation pathway, Mol. Cell 46 (2012) 461–471, https://doi.org/10.1016/j.molcel.2012.04.033.

[29] A.K. Das, M.D. Uhler, A.K. Hajra, Molecular cloning and expression of mammalian peroxisomal trans-2-enoyl-coenzyme a reductase cDNAs, J. Biol. Chem. 275 (2000) 24333–24340, https://doi.org/10.1074/jbc.M001168200.

[30] Y.A. Moon, J.D. Horton, Identification of two mammalian reductases involved in the two-carbon fatty acyl elongation cascade, J. Biol. Chem. 278 (2003) 7335–7343, https://doi.org/10.1074/jbc.M211684200.

[31] J. Gloerich, et al., Peroxisomal trans-2-enoyl-CoA reductase is involved in phytol degradation, FEBS Lett. 580 (2006) 2092–2096, https://doi.org/10.1016/j.febslet.2006.03.011.

[32] S.M. Houten, et al., Peroxisomal L-bifunctional enzyme (Ehhadh) is essential for the production of medium-chain dicarboxylic acids, J. Lipid Res. 53 (2012) 1296–1303, https://doi.org/10.1194/jlr.M024463.

[33] B.P. Atshaves, et al., Liver fatty acid-binding protein and obesity, J. Nutr. Biochem. 21 (2010) 1015–1032, https://doi.org/10.1016/j.jnutbio.2010.01.005.

[34] S. Ghareghomi, S. Ahmadian, N. Zarghami, H. Kahroba, Fundamental insights into the interaction between telomerase/TERT and intracellular signaling pathways, Biochimie 181 (2021) 12–24, https://doi.org/10.1016/j.biochi.2020.11.015.

[35] C.M. Counter, et al., Telomerase activity is restored in human cells by ectopic expression of hTERT (hEST2), the catalytic subunit of telomerase, Oncogene 16 (1998) 1217–1222, https://doi.org/10.1038/sj.onc.1201882.

[36] K.H. Shin, et al., Introduction of human telomerase reverse transcriptase to normal human fibroblasts enhances DNA repair capacity, Clin. Cancer Res. 10 (2004) 2551–2560, https://doi.org/10.1158/1078-0432.ccr-0669-3.

[37] Y. Qu, et al., Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau, Nat. Commun. 4 (2013) 2071, https://doi.org/10.1038/ncomms3071.

[38] T.B. Cui, G. X., A Monograph of Tibetan Mastiff in China, Shanghai Scientific & Technical Publishers, 2009.

[39] R.M. Schweizer, R.K. Wayne, Illuminating the mysteries of wolf history, Mol. Ecol. 29 (2020) 1589–1591, https://doi.org/10.1111/mec.15438.

[40] Q. Qiao, et al., Transcriptome sequencing of Crucihimalaya himalaica (Brassicaceae) reveals how Arabidopsis close relative adapt to the Qinghai-Tibet plateau, Sci. Rep. 6 (2016) 21729, https://doi.org/10.1038/srep21729.

[41] K.D. Mansfield, et al., Mitochondrial dysfunction resulting from loss of cytochrome c impairs cellular oxygen sensing and hypoxic HIF-alpha activation, Cell Metab. 1 (2005) 393–399, https://doi.org/10.1016/j.cmet.2005.05.003.

[42] R.G. Boutilier, J. St-Pierre, Surviving hypoxia without really dying, Comp Biochem Physiol A Mol Integr Physiol 126 (2000) 481–490, https://doi.org/10.1016/s1095-6433(00)00234-8.

[43] R.L. Ge, et al., Draft genome sequence of the Tibetan antelope, Nat. Commun. 4 (2013) 1858, https://doi.org/10.1038/ncomms2860.

[44] J.F. Storz, A.V. Signore, Introgressive hybridization and hypoxia adaptation in high-altitude vertebrates, Front. Genet. 12 (2021), 696484, https://doi.org/10.3389/fgene.2021.696484.

[45] M.E. Pamenter, J.E. Hall, Y. Tanabe, T.S. Simonson, Cross-species insights into genomic adaptations to hypoxia, Front. Genet. 11 (2020) 743, https://doi.org/10.3389/fgene.2020.00743.

[46] M. Miglani, et al., Shorter telomere length, higher telomerase activity in association with tankyrase gene polymorphism contribute to high-altitude pulmonary edema, Hum. Mol. Genet. 29 (2020) 3094–3106, https://doi.org/10.1093/hmg/ddaa205.

[47] Z.A. Cheviron, G.C. Bachman, A.D. Connaty, G.B. McClelland, J.F. Storz, Regulatory changes contribute to the adaptive enhancement of thermogenic capacity in high-altitude deer mice, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 8635–8640, https://doi.org/10.1073/pnas.1120523109.

[48] G. McClelland, G. Zwingelstein, C.R. Taylor, J.M. Weber, Increased capacity for circulatory fatty acid transport in a highly aerobic mammal, Am. J. Phys. 266 (1994) R1280–R1286, https://doi.org/10.1152/ajpregu.1994.266.4.R1280.

[49] X. Yang, et al., HTQC: a fast quality control toolkit for Illumina sequencing data, BMC Bioinformatics 14 (2013) 33, https://doi.org/10.1186/1471-2105-14-33.

[50] N.I. Weisenfeld, V. Kumar, P. Shah, D.M. Church, D.B. Jaffe, Direct determination of diploid genome sequences, Genome Res. 27 (2017) 757–767, https://doi.org/10.1101/gr.214874.116.

[51] H. Li, R. Durbin, Fast and accurate short read alignment with burrows-wheeler transform, Bioinformatics 25 (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.

[52] F.A. Simao, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics 31 (2015) 3210–3212, https://doi.org/10.1093/bioinformatics/btv351.

[53] W. Bao, K.K. Kojima, O. Kohany, Repbase update, a database of repetitive elements in eukaryotic genomes, Mob. DNA 6 (2015) 11, https://doi.org/10.1186/s13100-015-0041-9.

[54] A.F.A. Smit, R. H. P. G. {HYPERLINK "http://www.repeatmasker.org/RepeatModeler/" \h}.

[55] M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel, Bioinformatics 19 (Suppl. 2) (2003), https://doi.org/10.1093/bioinformatics/btg1080 ii215-225.

[56] E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise, Genome Res. 14 (2004) 988–995, https://doi.org/10.1101/gr.1865504.
[57] B.J. Haas, et al., Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies, Nucleic Acids Res. 31 (2003) 5654–5666, https://doi.org/10.1093/nar/gkg770.
[58] E.P. Nawrocki, et al., Rfam 12.0: updates to the RNA families database, Nucleic Acids Res. 43 (2015) D130–D137, https://doi.org/10.1093/nar/gku1063.
[59] K. Lagesen, et al., RNAmmer: consistent and rapid annotation of ribosomal RNA genes, Nucleic Acids Res. 35 (2007) 3100–3108, https://doi.org/10.1093/nar/gkm160.
[60] R. Apweiler, et al., UniProt: the universal protein knowledgebase, Nucleic Acids Res. 32 (2004) D115–D119, https://doi.org/10.1093/nar/gkh131.
[61] R. Apweiler, et al., The InterPro database, an integrated documentation resource for protein families, domains and functional sites, Nucleic Acids Res. 29 (2001) 37–40, https://doi.org/10.1093/nar/29.1.37.
[62] M. Kanehisa, The KEGG database, in: Novartis Found Symp 247, 2002, pp. 91–101, discussion 101–103, 119–128, 244–152.
[63] M.A. Harris, et al., The gene ontology (GO) database and informatics resource, Nucleic Acids Res. 32 (2004) D258–D261, https://doi.org/10.1093/nar/gkh036.
[64] NR. NR. ftp://ftp.ncbi.nih.gov/blast/db/.
[65] L. Li, C.J. Stoeckert Jr., D.S. Roos, OrthoMCL: identification of ortholog groups for eukaryotic genomes, Genome Res. 13 (2003) 2178–2189, https://doi.org/10.1101/gr.1224503.
[66] D. Darriba, G.L. Taboada, R. Doallo, D. Posada, ProtTest 3: fast selection of best-fit models of protein evolution, Bioinformatics 27 (2011) 1164–1165, https://doi.org/10.1093/bioinformatics/btr088.
[67] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (2014) 1312–1313, https://doi.org/10.1093/bioinformatics/btu033.
[68] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797, https://doi.org/10.1093/nar/gkh340.
[69] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (2007) 1586–1591, https://doi.org/10.1093/molbev/msm088.

View publication stats