



Whole genome sequencing of six dog breeds from continuous altitudes reveals adaption to high-altitude hypoxia

Xiao Gou, Zhen Wang, Ning Li, et al.

Genome Res. published online April 10, 2014

Access the most recent version at doi:[10.1101/gr.171876.113](https://doi.org/10.1101/gr.171876.113)

P<P Published online April 10, 2014 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Comprehensive immune receptor profiling.
Discover the **DriverMap™ AIR Assay** difference.

LEARN
MORE



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

Whole genome sequencing of six dog breeds from continuous altitudes reveals adaption to high-altitude hypoxia

Xiao Gou#1,2, Zhen Wang#3,4, Ning Li#2, Feng Qiu#4,5, Ze Xu 5, Dawei Yan 1, Shuli Yang 1, Jia Jia 4, Xiaoyan Kong 1, Zehui Wei 6, Shaoxiong Lu 1, Linsheng Lian 1, Changxin Wu 2, Xueyan Wang 1, Guozhi Li 1, Teng Ma 1, Qiang Jiang 1, Xue Zhao 1, Jiaqiang Yang 1, Baohong Liu 5, Dongkai Wei 5, Hong Li 3,4, Jianfa Yang 1, Yulin Yan 1, Guiying Zhao 1, Xingxing Dong 1, Mingli Li 1, Weidong Deng 1, Jing Leng 1, Chaochun Wei 4,8, Chuan Wang 7, Huaming Mao*1, Hao Zhang*2, Guohui Ding*3,4, Yixue Li*3,4,8,9

1. College of Animal Science and Technology, Yunnan Agricultural University, Heilongtan, Northern district, Kunming 650201, China
2. College of Animal Science and Technology, China Agricultural University, 17 East Qinghua Road, Beijing 100083, China
3. Key Laboratory of Systems Biology, Shanghai Institutes for biological Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China
4. Shanghai Center for Bioinformation Technology, Shanghai Industrial Technology Institute, 1278 Keyuan Road, Shanghai 201203, China
5. EG Information Technology Enterprise (EGI), Encode Genomics Biotechnology Co., Ltd., 100 Qinzhou Road, Shanghai 200235, China
6. College of Animal Science and Technology, Northwest Agricultural and Forestry University, No.3 Taicheng Road, Yangling 712100, China
7. National Center for Protein Science • Shanghai, National Facility for Protein Science in Shanghai, Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China

8. School of Life Science and Technology, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai 200240, China
9. School of Life Science and Technology, Shanghai Tongji University, 1239 Siping Road, Shanghai 200092, China

These authors contribute equally to this work.

* To whom correspondence should be addressed.

Email: maohm@vip.sina.com (H. Mao), zhanghao827@163.com (H. Zhang), gwding@sibs.ac.cn (G. Ding), yxli@sibs.ac.cn (Y. Li).

Running title: Sequencing of dogs reveals adaption to hypoxia

Key words: dog genome, Tibetan Mastiff, high-altitude hypoxia, selective sweep, *EPAS1*

Abstract

The hypoxic environment imposes severe selective pressure on species living at high altitude. To understand the genetic bases of adaption to high altitude in dogs, we performed whole-genome sequencing of 60 dogs including five breeds living at continuous altitudes along the Tibetan plateau from 800 to 5,100 m as well as one European breed. More than 150× sequencing coverage for each breed provides us a comprehensive assessment of the genetic polymorphisms of the dogs including Tibetan Mastiffs. Comparison of the breeds from different altitudes reveals strong signals of population differentiation at the locus of hypoxia-related genes including endothelial Per-Arnt-Sim (PAS) domain protein 1 (*EPAS1*) and beta hemoglobin cluster. Especially, four novel non-synonymous mutations specific to high-altitude dogs are identified at *EPAS1*, one of which occurred at a quite conserved site in the PAS domain. The association testing between *EPAS1* genotypes and blood-related phenotypes on additional high-altitude dogs reveals that the homozygous mutation is associated with the decreased blood flow resistance, which may help to improve hemorheologic fitness. Interestingly, *EPAS1* was also identified as a selective target in Tibetan highlanders, though no amino acid changes were found. Thus, our results not only indicate parallel evolution of humans and dogs in adaption to high-altitude hypoxia, but also provide a new opportunity to study the role of *EPAS1* in the adaptive processes.

Introduction

The mechanisms of organismal adaptation to high-altitude hypoxia are of great interest during recent years. Highland wild animals have a long life history at high altitude, and the whole genomes of yak (descendants of wild yak) (Qiu et al. 2012), Tibetan antelope (Ge et al. 2013), snow leopard (Cho et al. 2013) and wild boar (Li et al. 2013) have been sequenced. In contrast, the human settlement history on highland is rather short, which dates from about 25,000 years ago (Zhao et al. 2009). Whole-genome genotyping and re-sequencing have been performed for three typical highland populations including Tibetans (Beall et al. 2010; Bigham et al. 2010; Simonson et al. 2010; Yi et al. 2010; Peng et al. 2011; Xu et al. 2011), Andeans (Bigham et al. 2009; Bigham et al. 2010) and Ethiopians (Alkorta-Aranburu et al. 2012; Scheinfeldt et al. 2012).

The increased oxygen uptake and delivery are physiological hallmarks of high-altitude adaption. On one hand, the capacity of oxygen uptake is determined by hemoglobin concentration and oxygen affinity. For example, the Andean highlanders display a high level of hemoglobin concentration (Beall et al. 2002; Beall 2007). The high oxygen affinity of hemoglobin is found in many highland animals such as yak (Weber et al. 1988), alpaca (Piccinini et al. 1990), deer mice (Storz et al. 2007; Storz et al. 2009), bar-headed goose (Zhang et al. 1996; Liang et al. 2001) and Andean goose (Jessen et al. 1991). On the other hand, the rate of oxygen delivery is determined by blood flow. For example, although Tibetans maintain a nearly normal level of hemoglobin concentration and a low level of oxygen saturation, they display a high level of blood flow, resulting in the increase of oxygen delivery (Beall et al. 2001; Erzurum et al. 2007).

Whole-genome scans revealed that positive selection for human high-altitude adaption occurred in the hypoxia-inducible factor (HIF) pathway (Bigham et al. 2009; Beall et al. 2010; Bigham et al. 2010; Simonson et al. 2010; Yi et al. 2010; Peng et al. 2011; Xu et al. 2011; Alkorta-Aranburu et al. 2012; Scheinfeldt et al. 2012), which

regulates genes associated with blood physiology. In addition, metabolic pathways may also be involved in the adaptive process of yak (Qiu et al. 2012) and Tibetan antelope (Ge et al. 2013).

Although a lot of studies focused on wildlife and human highlanders, no research was performed on domesticated animals that migrated to the plateau with humans, which represent an adaption pattern on a short evolutionary time scale of thousands of years. For example, the Tibetan Mastiff is a native dog living in Tibetan plateau with an altitude of 3,000-6,000 m. It is also an ancient dog in the world (Li and Zhang 2012). However, the genetic and physiological mechanisms of its adaption to high-altitude environments remain elusive.

In this study, we sampled five dog breeds including the Tibetan Mastiff from continuous altitudes along the Ancient Tea Horse Road in southwestern China as well as one European breed. We performed whole-genome sequencing for the dogs and identified candidate genes for high-altitude adaption using selective sweep mapping. We also measured the hematologic and hemorheologic parameters of the dogs and tested the association between the candidate alleles and blood physiology.

Results

Genetic diversity and population structure

We sampled 60 dogs belonging to six breeds living at continuous altitudes (Table 1). To avoid potential kinships, each dog was sampled from one individual village, which is dispersed along the Ancient Tea Horse Road and far away from each other. The sex ratio was kept as 1:1 for each breed. We classified the breeds into three altitude levels, i.e., high (>3,000 m), middle (2,000-3,000 m) and low (<2,000 m). The high-altitude breeds include Tibetan Mastiffs (TM) and Diqing indigenous dogs (DQ). The middle-altitude breed was chosen as Lijiang indigenous dogs (LJ). The low-altitude breeds contain Yingjiang indigenous dogs (YJ), Kunming dogs (KM) and

German Shepherds (GS). Grouping multiple breeds living at similar altitudes could contribute to reducing the influence of breed-specific genetic background.

We performed whole-genome sequencing for each individual dog. The coverage depth after genome alignment is about 15× for each dog (Supplemental Figure 1, Supplemental Table 1 and Supplemental Table 2), resulting in more than 150× for each breed for variants calling (Supplemental Table 3-Supplemental Table 5). The levels of genome-wide genetic diversity, as measured by Watterson's θ and pairwise nucleotide diversity π , are relatively higher in the TM and indigenous breeds (DQ, LJ and YJ) than those in KM and GS (Table 1 and Supplemental Figure 2), consistent with the fact that the former are more ancient breeds than the latter working dogs that are domesticated recently (Vonholdt et al. 2010; Wang et al. 2013). In agreement with this, the minor allele frequency (MAF) distribution and Tajima's D indicate a larger proportion of low frequency minor alleles in the ancient breeds than in the modern breeds (Supplemental Figure 3). Additionally, linkage disequilibrium (LD) and haplotype analyses show that the ancient breeds have a faster decay of pairwise correlation coefficient (r^2) and lower integrated haplotype homozygosity (iHH) (Voight et al. 2006) than the modern breeds (Supplemental Figure 4).

To examine genetic relationships among the dogs, we conducted principal component analysis (PCA) based on whole-genome SNPs. In the PCA plot, the LJ, YJ, KM and GS breeds could be separated by the first and second eigenvectors, while TM and DQ are mixed (Figure 1A). However, when more eigenvectors are incorporated, most dogs in TM and DQ could still be separated (Figure 1A). These results suggest close but different genetic backgrounds between TM and DQ. To resolve their phylogenetic relationships, we constructed a neighbor-joining (NJ) tree based on the pairwise genetic distances (Figure 1B). The tree also splits TM and DQ into distinct branches except one dog in TM. Considering possible admixture among the breeds, we further performed population structure analysis with frappe (Tang et al. 2005), which estimates individual ancestry and admixture proportions assuming K ancestral populations (Figure 1C). When $K = 2$, we found a division between the ancient (TM,

DQ, LJ and YJ) and modern breeds (KM and GS). When $K = 3$, we found a division between the breeds living at high altitude (TM and DQ) and at middle/low altitude (LJ and YJ). When $K = 5$, TM and DQ are separated, though slight extent of ingression from DQ happened in the TM breed.

Signature of selection associated with high altitude

To detect signature of natural selection associated with high altitude, we searched the dog genome for regions with high coefficient of nucleotide differentiation F_{ST} among the breeds living at high (TM, DQ), middle (LJ) and low altitude (YJ, KM, GS) (Figure 2A). The approach of population differentiation has been widely used to detect selective sweeps based on whole genome SNPs, such as artificial selection during dog domestication (Axelsson et al. 2013) and diversification (Akey et al. 2010; Vonholdt et al. 2010). We scanned the autosomes with a non-overlapping 100 kb window and calculated the F_{ST} value for each window. As previous studies (Axelsson et al. 2013), we focused on the regions with extremely high Z-transformed F_{ST} values ($Z(F_{ST}) > 5$) in the genome-wide empirical distribution. In total, 28 unique autosomal regions containing 141 candidate genes were identified (Supplemental Table 6), and five genes of them including *EPAS1*, *MSRB3*, *HBB*, *CDK2* and *GNB1* belong to the Gene Ontology (GO) categories ‘response to oxygen levels’ and ‘response to oxidative stress’ (Figure 2A). To date the sweeps during the migration of dogs from low to high altitude, we also calculated the pairwise F_{ST} and $Z(F_{ST})$ between the three altitude levels (Supplemental Figure 5 and Supplemental Table 7).

The region with the strongest differentiation signal ($Z(F_{ST}) = 10.83$, Figure 2A) among the three altitude levels incorporates endothelial Per-Arnt-Sim (PAS) domain protein 1 (*EPAS1*), a gene encodes the hypoxia-inducible factor (HIF) 2A (Patel and Simon 2008). The region is also highly differentiated when the high-altitude breeds were compared with the low- and middle-altitude ones ($Z(F_{ST}) = 10.97$ and 8.02, respectively (Supplemental Figure 5), suggesting that it is involved in a selective sweep during the move of dogs from middle to high altitude (Supplemental Table 7). Interestingly, *EPAS1* was also identified as a selective target in Tibetan people (Beall

et al. 2010; Yi et al. 2010; Peng et al. 2011; Xu et al. 2011), which is associated with their hemoglobin concentration (Beall et al. 2010; Yi et al. 2010). Network analysis indicated that the other candidate hypoxia-response genes we identified would all be regulated by HIF signaling pathway (Supplemental Figure 6), suggesting an essential role of *EPAS1* in the adaption of high-altitude dogs.

The region with the second largest F_{ST} value ($Z(F_{ST}) = 9.75$, Figure 2A) among the three altitude levels is overlapped with methionine sulfoxide reductase B3 (*MSRB3*). Both the high- and middle-altitude breeds show high differentiation to the low-altitude ones in this region ($Z(F_{ST}) = 9.77$ and 6.75 , respectively, Supplemental Figure 5), implying that the selective sweep occurred on the migration of dogs from low to middle altitude (Supplemental Table 7). *MSRB3* has been identified as a gene associated with ear types in dogs (Boyko et al. 2010; Vaysse et al. 2011), but its association with altitude has not been reported previously. The beta hemoglobin (*HBB*) cluster region is also among the top three outliers ($Z(F_{ST}) = 8.51$, Figure 2A) and has a large differentiation between the high- and low-altitude breeds ($Z(F_{ST}) = 8.84$, Supplemental Figure 5), indicating a selective sweep on the dogs' settlement from low to high altitude (Supplemental Table 7). The increases in oxygen affinity of hemoglobin have been found in diverse animal species to mediate the adaptive response to high altitude hypoxia (Storz and Moriyama 2008).

In addition to the increase of population differentiation, selective sweeps could result in the decrease of genetic diversity and increase of LD around the selective target (Nielsen 2005). Therefore, we compared π and iHH of the six breeds around the *EPAS1* (Figure 2B) and *HBB* loci (Figure 2C), respectively. The π value of TM is decreased in the *EPAS1* region and is the lowest among the six breeds. The iHH value of TM is increased in the *EPAS1* region and is the highest among the six breeds (Figure 2B). Both of the results supported that a selective sweep occurred in the *EPAS1* region in TM. The DQ breed also shows an increased iHH value in the *EPAS1* region as TM. However, the π value is increased in DQ, which may be caused by genetic admixture (Figure 2B). The diversity of LJ in the *EPAS1* region is the highest

among the six breeds, suggesting extensive admixture and weak directional selection in the middle altitude dogs (Figure 2B). When we inspected the beta hemoglobin cluster region, we observed a decrease of π value and increase of iHH value in both TM and DQ (Figure 2C), confirming the F_{ST} screening that a selective sweep occurred in the loci of high-altitude dogs.

Mutations in the *EPAS1* gene and physiological association

We examined the mutations in the *EPAS1* gene in detail and found four non-synonymous mutations in our sample, which cause amino acid variants Gly305-Ser (G305S), Asp494-Glu (D494E), Val500-Met (V500M) and Pro750-Ser (P750S) in the translated protein (based on Ensembl annotation (Flicek et al. 2014), Figure 3A, Supplemental Table 8). The four mutant alleles are in complete linkage disequilibrium (pairwise $r^2 = 1$) in all breeds, which are part of a 8-kb LD block (Supplemental Figure 7). Haplotype tree (Figure 3B) and network (Supplemental Figure 7) of the LD block show that the haplotypes comprising the four mutant alleles all belong to the high-altitude clade, which have a long evolutionary distance from the haplotypes of the low-altitude clade. Also, the phylogenetic tree of the six breeds reconstructed from the LD block show separated clades between high and middle/low altitude (Supplemental Figure 8), which is quite distinct from the genome-wide tree (ancient-modern separation). Interestingly, the frequency of the four mutant alleles is increased with the elevated altitude. It keeps rare in low-altitude breeds (KM 0%, GS 0%, YJ 5%), moderate in middle-altitude breeds (LJ 30%) but reaches high frequency in high-altitude breeds (DQ 80%, TM 100%) (Figure 3C). To confirm this observation based on high-throughput sequencing, we performed genotyping of the four alleles on additional samples with Sanger sequencing and the result was similar (KM 0% ($n = 11$), GS 0% ($n = 10$), YJ 0% ($n = 20$), LJ 20% ($n = 32$), DQ 73% ($n = 30$), TM 94% ($n = 35$)) (Figure 3C). The high divergence of the haplotypes and alleles among the breeds indicates that the non-synonymous mutations may be responsible for the selective sweep at *EPAS1*.

Among the four variants, only one (G305S) occurred in a well defined protein domain (PAS domain, Figure 3A). This domain facilitates heterodimerization with the aryl hydrocarbon receptor nuclear translocator (ARNT), resulting in a functional transcriptional complex (Patel and Simon 2008). Taking the crystal structure of PAS domain in human EPAS1 as template, our homology modeling suggested that the mutation G305S occurred in a beta sheet, which may affect the thermodynamic stability of the domain (Supplemental Figure 9). To further evaluate the functional impact of the variants, we aligned the mutant EPAS1 protein with its ortholog proteins in diverse vertebrates (Figure 3A). The comparison reveals that G305S is also a quite conserved amino acid mutation, which is invariant among all the other animals we examined. In contrast, the remaining three variants could all be found in other species. Prediction of functional effects of the variants supports that only G305S is deleterious, while the others are tolerated (Supplemental Table 8). All of these results imply that G305S is the most likely causal mutation for the *EPAS1* sweep in high-altitude dogs.

To explore the physiology of the dogs' adaption to high-altitude hypoxia, we measured the hematologic and hemorheologic parameters for indigenous dogs living in a range from 500 to 3,000 m (Supplemental Table 9). There are evident tendencies of the blood-related traits with the increase of altitude. Especially, the adult hemoglobin concentrations among middle- and high-altitude dogs reach 170 g/L as compared with 130 g/L among low-altitude dogs, which is a clear advantage in the rise of arterial oxygen content. However, with the increase of hemoglobin concentration, the blood viscosity and vascular resistance are also increased (Supplemental Table 9), which bring about potential risks for cardiovascular systems (Somer and Meiselman 1993; Allen and Patterson 1995). As the association between SNPs at *EPAS1* and blood-related traits have been well established in humans (van Patot and Gassmann 2011), we conducted association testing for the variant G305S in DQ (Supplemental Table 10), the high-altitude breed where enough homozygotes ($n = 40$) and heterozygotes ($n = 29$) could be collected. Although no evident relationship

with hemoglobin concentration was found, we discovered a statistically significant association between the genotypes and vascular resistance (Figure 3D). The homozygotes with two mutant alleles show decreased vascular resistance as compared with the heterozygotes. The results suggest that the G305S variant may play important roles in the improvement of blood fluidity, which confers advantages for oxygen delivery and cardiovascular functions.

Discussion

In this study, we performed whole-genome sequencing of 60 dogs including five breeds from continuous altitudes along the Ancient Tea Horse Road in the Tibetan Plateau as well as one breed of German Shepherds. We cataloged millions of SNPs of each breed for evolutionary and genetic researches. Especially, it is the first study to characterize the genetic polymorphisms of Tibetan Mastiffs, which are known as “oriental deiform dog” in Tibet and one of the most ancient and ferocious dogs in the world (Li and Zhang 2012). Our population structure analyses revealed that Tibetan Mastiffs have a close but different genetic background with other indigenous dogs living at high altitude (Figure 1), though their origin and relationship with plateau wolves need further studies. We also found the existence of admixture among Tibetan Mastiffs and indigenous dogs (Figure 1), which raised the demands for genetic testing in the identification and breeding of Tibetan Mastiffs.

We conducted selective sweep mapping for the breeds from different altitudes and identified several candidate regions with high extent of differentiation on the genome scale (Figure 2A). Two of the genes contained in the top regions, *EPAS1* and *HBB*, were previously reported as targets of natural selection for hypoxia adaption in humans and other animals (Storz and Moriyama 2008; van Patot and Gassmann 2011). The two loci also displayed reduced genetic diversity and/or increased LD in high-altitude dogs especially Tibetan Mastiffs (Figure 2B, C). These results suggest that driven by the same hypoxic environment, natural selection could act on a similar group of genes in the adaptive process of different species. Interestingly, the parallel

evolution on the molecular level may extensively exist between humans and dogs in metabolic and neurological processes, probably due to their similar environments in the recent past (Wang et al. 2013).

We discovered and validated four novel non-synonymous mutations at *EPAS1* in high-altitude dogs, of which G405S occurred in a well defined domain and quite conserved site (Figure 3A). The mutant alleles and corresponding haplotypes have a large divergence between low- and high-altitude dogs (Figure 3B, C), implying that the amino acid variations may be responsible for the signal of selective sweep at the *EPAS1* locus.

The transcription factor encoded by *EPAS1* plays a key role in transcriptional response to hypoxia in various physiological and pathologic conditions, including up-regulating the expression of erythropoietin (*EPO*) (Patel and Simon 2008). Especially, mutations at *EPAS1* in humans are strongly associated with hematologic phenotypes (van Patot and Gassmann 2011). On one hand, clinical case studies found that several gain-of-function mutations occurred in close vicinity to the primary hydroxylation site (Pro531) at *EPAS1* cause erythrocytosis (Percy et al. 2008a; Percy et al. 2008b). On the other hand, the *EPAS1* polymorphisms unique to the native Tibetan people were found to be associated with their lower hemoglobin concentrations (Beall et al. 2010; Yi et al. 2010), suggesting a loss-of-function role of *EPAS1* in high-altitude adaption. As erythrocytosis is a common symptom of chronic mountain sickness which will lead to high blood viscosity and cardiovascular disorders, the decrease in hemoglobin level may provide a protective mechanism for Tibetan people (Beall et al. 2010; van Patot and Gassmann 2011). In fact, comparing with Andeans and other highland migrants with shorter settlement history, Tibetans do not show markedly elevated hemoglobin concentration (Beall 2007), probably due to their unique *EPAS1* genotypes (van Patot and Gassmann 2011). Instead, Tibetans display a higher blood flow for oxygen delivery (Erzurum et al. 2007), which could compensate the effect of lower hemoglobin concentration.

Nonetheless, it is worth noting that all *EPAS1* variations detected previously in the Tibetan population are at introns (Beall et al. 2010; Yi et al. 2010), inhibiting further study of its molecular mechanism in the adaptive process. Fortunately, the key amino acid mutation we identified at *EPAS1* in dogs, G405S, is predicted to be damaging, which is also likely to cause the loss of function of *EPAS1*. Moreover, we demonstrated that the variation is associated with lower blood flow resistance in high-altitude dogs (Figure 3D). Therefore, we propose that similar mechanisms may be adopted by both humans and dogs in adaption to high-altitude hypoxia. Although the physiological responses need further comparison between humans and dogs, these novel variations of *EPAS1* could shed new insights into the study of the gene function.

Methods

Sample preparation and sequencing

For each dog, the genomic DNA was extracted from 200 μ l peripheral venous blood with QIAamp DNA Blood Mini kit (Qiagen). The quality and integrity of DNA was controlled by A260/280 ratio and agarose gel electrophoresis. For sequencing library preparation, the genomic DNA was sheared to fragments of 300-500 bp, which were then end-repaired, 'A'-tailed and ligated to Illumina sequencing adapters. The ligated products with sizes of 400-500 bp were selected on 2% agarose gels and then amplified by LM-PCR. The libraries were sequenced on Illumina HiSeq 2000 with 2 $\times$ 100 bp paired-end mode, which was controlled by Illumina HiSeq Control Software.

Reads processing and variants calling

The raw reads were processed by two rounds of quality control (QC) (Supplemental Table 1). Firstly, low quality reads marked by Illumina sequencers in the FASTQ files (Casava v1.8) were filtered. Secondly, for each read, the low quality end with base quality scores lower than 20 were trimmed. The read was removed if its length was smaller than 35 after trimming. Only paired reads were preserved during QC. Reads after QC were mapped to the dog reference genome assembly canFam3 (<http://hgdownload.soe.ucsc.edu/goldenPath/canFam3/bigZips/canFam3.fa.gz>) using BWA-MEM (v0.7.4) (Li and Durbin 2009) with default parameters on a per-individual basis. We removed duplicated reads and calculated mapping statistics including the coverage of depth from the .bam files by SAMtools (v0.1.19) (Li et al. 2009) (Supplemental Table 2).

The SNPs and small indels were called using the SAMtools pipeline (Li 2011) on a per-breed (ten individuals) basis (Supplemental Table 3). Firstly, .mpileup files were generated by samtools mpileup with the parameters “-u -C50 -DS -q20”. Secondly, .vcf files containing raw variants were generated by bctools view with the parameters “-evcgN”. Finally, the raw variants were filtered by vcfutils.pl varFilter with the

minimum depth “-d 20” and maximum depth “-D 300”. The cutoffs were set according to the sequencing depth distribution per individual (Supplemental Figure 1). Other parameters for filtering were default, including *p*-value of Hardy–Weinberg equilibrium (0.0001). Given the position of a variant, the genotype with the maximum posterior probability of each individual was given in the .vcf file. The variants after filtering were annotated by ANNOVAR (v2013-06-21) (Wang et al. 2010) according to Ensembl (Flicek et al. 2014) gene annotation (<http://hgdownload.soe.ucsc.edu/goldenPath/canFam3/database/ensGene.txt.gz>) (Supplemental Table 4 and Supplemental Table 5).

Population genetics analysis

The .vcf files for the six breeds were merged by vcf-merge in VCFtools (v0.1.11) (Danecek et al. 2011). When a genotype was missing in one breed, it was set to the reference allele with the option “-R 0/0”. Only bi-allelic SNPs in autosomes were preserved for subsequent population structure and polymorphism analysis. The merged .vcf file was converted to tab-delimited text file by vcf-to-tab in VCFtools, and to PLINK format files (.ped and .map) by VCFtools and PLINK (v1.07) (Purcell et al. 2007) when appropriate. The principal component analysis (PCA) was performed by GCTA (v1.13) (Yang et al. 2011) (Figure 1A), which first generated the genetic relationship matrix (GRM) with the “--make-grm” option and then estimated the first four principal components with the “--pca 4” option. The pairwise genetic distances between individuals were measured by the number of net nucleotide substitutions (formula 12.67 in ref. (Nei and Kumar 2000)), which were calculated for each SNP sites and then summed up. Based on the distance matrix, the NJ tree was constructed and displayed by MEGA (v5.2) (Tamura et al. 2011) (Figure 1B). The population structure was inferred by frappe (v1.1) (Tang et al. 2005) with a maximum likelihood method (Figure 1C). The number of ancestral clusters *K* was ranged from 2 to 6, with 10000 iterations each run. The graphs of the population structure were display by distruct (v1.1) (Rosenberg 2004) on a per-individual basis.

Standard population genetic statistics, including Watterson's θ , pairwise nucleotide diversity π (Supplemental Figure 2) and Tajima's D (Supplemental Figure 3) were calculated for each breed by the Bio::PopGen::Statistics package in BioPerl (v1.6.1) (Stajich et al. 2002). The coefficient of nucleotide differentiation F_{ST} among the breeds was calculated by the Bio::PopGen::PopStats package in BioPerl (Supplemental Table 6). The statistics were estimated over a sliding window with fixed size along the genome. The Manhattan plot of genome-wide F_{ST} values was generated using the mhtplot function in the R package gap (Figure 2A and Supplemental Figure 5). The level of linkage disequilibrium (LD) measured by the correlation coefficient (r^2) between unphased markers (Supplemental Figure 4A) was calculated by Haploview (v4.2) (Barrett et al. 2005) with the option “-dprime -maxdistance 50 -minMAF 0.1 -hwcutoff 0.05”. We also phased the genotypes of all individuals into haplotypes by fastPHASE (v1.4.0) (Scheet and Stephens 2006) with the parameters “-T10 -K8 -u”. The integrated haplotype homozygosity (iHH) (Voight et al. 2006) for each breed was calculated by XP-EHH (Pickrell et al. 2009) based on the phased data (Supplemental Figure 4B). The iHH values were log-transformed and averaged over a sliding window with fixed size along the genome.

***EPAS1* mutation analysis**

The protein domains of EPAS1 were predicted by the Pfam web service (Punta et al. 2012) (Figure 3A). The ortholog proteins of EPAS1 in 17 representative vertebrates were retrieved from Ensembl (Flicek et al. 2014) (release 72) and only one to one ortholog were included (Figure 3A). The multiple sequence alignment was performed by ClustalX (v2.1) (Larkin et al. 2007) and displayed by Jalview (v2.8) (Waterhouse et al. 2009). The LD block containing the non-synonymous mutations was detected and visualized by Haploview (v4.2) (Barrett et al. 2005) (Supplemental Figure 7A). The haplotype NJ tree was constructed by MEGA (v5.2) (Tamura et al. 2011) with the default parameters (Figure 3B). The haplotype network was constructed and plotted by Network (v4.611, Fluxus Engineering) (Supplemental Figure 7B). It was calculated with the median-joining method (Bandelt et al. 1999) (ϵ

= 0), followed by reduction with the MP option (Polzin and Daneshmand 2003). The sequence of dog EPAS1 was searched against the PDB database (Rose et al. 2013) using BLAST-P (Altschul et al. 1997) to identify suitable templates for homology modeling (Supplemental Figure 9). I-TASSER (v2.1) (Roy et al. 2010) was used for structure modeling by threading the sequence onto the selected templates. The three-dimensional model structures were visualized with PyMOL (v1.6, Schrödinger, LLC). Structure-based energy calculation was performed by the Eris server (Yin et al. 2007) using flexible backbone method upon single amino acid substitutions. The functional consequences of amino acid mutations were predicted by SIFT (Kumar et al. 2009), which was invoked through Ensembl's (Flicek et al. 2014) Variant Effect Predictor (Supplemental Table 8).

Physiological measurement and association test

The foreleg venous blood was collected to measure the physiological parameters in situ. The hematologic and hemorheologic parameters were measured by BC-2800Vet Auto Hematology Analyzer (Mindray Co., Ltd, Shenzhen) and ZL1000 Auto Blood Rheology Analyzer (Zonci Co., Ltd, Beijing), respectively (Supplemental Table 9 and Supplemental Table 10). The non-synonymous mutations at *EPAS1* were genotyped by traditional Sanger sequence technology. The association testing was performed with each physiological parameter as the response variable, genotype as the predictive variable, and age (<1 year vs >1 year) as the covariate. We also modeled the gender as a covariate but no difference was found in this factor. The *F* test in the analysis of variance (ANOVA) was used to evaluate the significance of genotypes (Figure 3D).

Data access

All sequencing data from this study have been submitted to the NCBI Sequence Read Archive (SRA; <http://www.ncbi.nlm.nih.gov/sra>) under accession number SRP035294.

Acknowledgements

We acknowledge the anonymous reviewers for their constructive comments. We are indebted to many people whose names are not included in the author list, but who contributed to this project. This work was supported by State Key Basic Research Program (973) (2011CB910204, 2010CB529206, 2011CBA00801), Research Program of CAS (KSCX2-EW-R-04, KSCX2-YW-R-190, XXH12503-02-02-09), National Natural Science Foundation of China (31260544, U1036604, 31070752, 31301032), Chinese Ministry for Science and Technology Grant (2008BAI64B01), Chinese High-Tech R&D Program (863) (2013AA102503, 2009AA02Z304, 2012AA020404), China Postdoctoral Science Foundation (2013M531226), Shanghai Postdoctoral Scientific Program (13R21417300) and SA-SIBS Scholarship Program.

Disclosure declaration

The authors declare no conflicts of interest.

Figure legends

Figure 1. Genetic relationships and population structure of the 60 dogs based on all autosomal SNPs. (A) Principle component plot. The first (PC1) and second component (PC2) are shown in the main figure, while the third (PC3) and forth component (PC4) are shown in the insert figure. The percentages indicate the proportion of variance explained by each component. (B) Unrooted neighbor-joining tree. The evolutionary distance is measured by the number of net nucleotide substitutions between individuals. (C) Population structures with the number of ancestral clusters K from 2 to 6. Each color represents one ancestral cluster and each vertical bar represents one dog. The length of colored segments represents corresponding ancestry attributions.

Figure 2. Selective sweep analysis of the breeds from different altitude levels. (A) Manhattan plot of F_{ST} among the high- (TM, DQ), middle- (LJ) and low- (YJ, KM, GS) altitude breeds. The F_{ST} was calculated for each 100 kb autosomal window. The dashed line denotes a threshold of $Z(F_{ST}) = 5$. (B) Diversity π and iHH (integrated haplotype homozygosity) around the *EPAS1* loci. The region with $Z(F_{ST}) > 5$ is shaded. π and iHH were calculated for each 10 kb window and Z-transformed on the genome scale. The values were smoothed by lowess regression. (C) π and iHH around the beta hemoglobin loci.

Figure 3. *EPAS1* mutations in the coding region. (A) Structural and evolutionary analysis of the four amino acid variants. The protein coordinate is based on Ensembl ID ENSCAFP00000003819. The coordinate of NCBI RefSeq XP_531807.2 is slightly different but the variants remain the same. The upper panel shows the Pfam domains of the protein. The orthologous protein sequences from 17 vertebrates are aligned with the mutant residues shown in red. The neighbor-joining (NJ) tree derived from the multiple alignment is shown in the left panel. The red stars indicate two hydroxylation sites (Pro405 and Pro531), which is essential for oxygen sensing (Patel and Simon 2008). PAS, Per-Arnt-Sim; HIF,

hypoxia-inducible factor; CTAD, C terminal transactivation domain. (B) NJ tree of the 8-kb haplotypes comprising of the four non-synonymous mutations. The evolutionary distance is defined as the number of different nucleotides between haplotypes. The haplotypes from high- and low-altitude breeds were grouped into two distinct clades. (C) Percentages of homozygotes and heterozygotes in the six breeds. The mutant and reference alleles are represented by '+' and '-', respectively. Besides the 60 dogs sequenced with the Illumina technology, additional samples were genotyped with Sanger sequencing. The sample size is shown beside the bars. (D) Association between *EPAS1* genotypes (mutant allele: A; reference allele: G) and blood flow resistance in DQ. The blood flow resistance was measured at different shear rate. In the boxplot, the line within the box defines the median; the ends of the boxes define the 25th and 75th percentiles; and the error bars define the 10th and 90th percentiles. The ANOVA *F*-test was performed, taking the age as covariant.

Tables

Table 1. Sample, sequencing and polymorphism information for the six dog breeds.

Breed (abbreviation)	Sample size	Location	Altitude	Total coverage of depth ^a (×)	Watterson's θ ($\times 10^{-3}$)	Nucleotide diversity π ($\times 10^{-3}$)
Tibetan Mastiff (TM)	10	Cuomei, Tibet, China ($n = 4$)	5,100 m	154.22	1.22	1.35
		Yushu, Qinghai, China ($n = 4$)	4,200 m			
		Diqing, Yunnan, China ($n = 2$)	3,300 m			
Diqing indigenous dog (DQ)	10	Diqing, Yunnan, China	3,300 m	154.88	1.22	1.35
Lijiang indigenous dog (LJ)	10	Lijiang, Yunnan, China	2,400 m	154.72	1.27	1.35
Kunming dog (KM)	10	Kunming, Yunnan, China	1,800 m	159.23	0.99	1.15
German Shepherd (GS)	10	Kunming, Yunnan, China	1,800 m	159.62	0.61	0.75
Yingjiang indigenous dog (YJ)	10	Yingjiang, Yunnan, China	800 m	153.46	1.28	1.38

^a Only autosomes were considered for depth and polymorphism calculation.

Reference

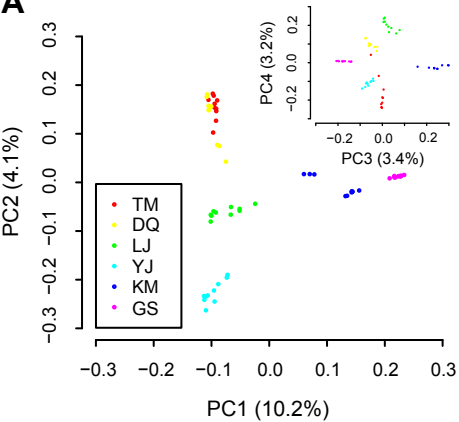
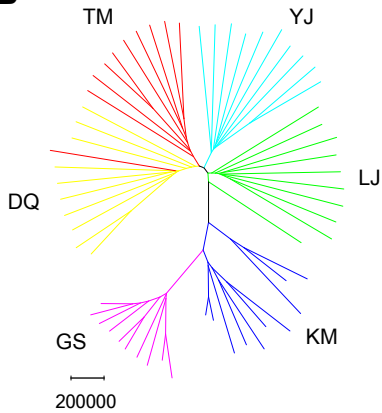
- Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW. 2010. Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci U S A* **107**(3): 1160-1165.
- Alkorta-Aranburu G, Beall CM, Witonsky DB, Gebremedhin A, Pritchard JK, Di Rienzo A. 2012. The genetic architecture of adaptations to high altitude in Ethiopia. *PLoS Genet* **8**(12): e1003110.
- Allen MT, Patterson SM. 1995. Hemoconcentration and stress: a review of physiological mechanisms and relevance for cardiovascular disease risk. *Biological psychology* **41**(1): 1-27.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17): 3389-3402.
- Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**(7441): 360-364.
- Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**(1): 37-48.
- Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**(2): 263-265.
- Beall CM. 2007. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc Natl Acad Sci U S A* **104** Suppl 1: 8655-8660.
- Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, McCormack M et al. 2010. Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. *Proc Natl Acad Sci U S A* **107**(25): 11459-11464.
- Beall CM, Decker MJ, Brittenham GM, Kushner I, Gebremedhin A, Strohl KP. 2002. An Ethiopian pattern of human adaptation to high-altitude hypoxia. *Proc Natl Acad Sci U S A* **99**(26): 17215-17218.
- Beall CM, Laskowski D, Strohl KP, Soria R, Villena M, Vargas E, Alarcon AM, Gonzales C, Erzurum SC. 2001. Pulmonary nitric oxide in mountain dwellers. *Nature* **414**(6862): 411-412.
- Bigham A, Bauchet M, Pinto D, Mao XY, Akey JM, Mei R, Scherer SW, Julian CG, Wilson MJ, Herraes DL et al. 2010. Identifying Signatures of Natural Selection in Tibetan and Andean Populations Using Dense Genome Scan Data. *Plos Genetics* **6**(9).
- Bigham AW, Mao X, Mei R, Brutsaert T, Wilson MJ, Julian CG, Parra EJ, Akey JM, Moore LG, Shriver MD. 2009. Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. *Hum Genomics* **4**(2): 79-90.
- Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, Zhao K, Brisbin A, Parker HG, vonHoldt BM et al. 2010. A simple genetic

- architecture underlies morphological variation in dogs. *PLoS Biol* **8**(8): e1000451.
- Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, Oh S, Kim HM, Jho S, Kim S et al. 2013. The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat Commun* **4**: 2433.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**(15): 2156-2158.
- Erzurum SC, Ghosh S, Janocha AJ, Xu W, Bauer S, Bryan NS, Tejero J, Hemann C, Hille R, Stuehr DJ et al. 2007. Higher blood flow and circulating NO products offset high-altitude hypoxia among Tibetans. *Proceedings of the National Academy of Sciences of the United States of America* **104**(45): 17593-17598.
- Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S et al. 2014. Ensembl 2014. *Nucleic Acids Res* **42**(Database issue): D749-755.
- Ge RL, Cai QL, Shen YY, San A, Ma L, Zhang Y, Yi X, Chen Y, Yang LF, Huang Y et al. 2013. Draft genome sequence of the Tibetan antelope. *Nature Communications* **4**.
- Jessen TH, Weber RE, Fermi G, Tame J, Braunitzer G. 1991. Adaptation of bird hemoglobins to high altitudes: demonstration of molecular mechanism by protein engineering. *Proc Natl Acad Sci U S A* **88**(15): 6519-6522.
- Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* **4**(7): 1073-1081.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21): 2947-2948.
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**(21): 2987-2993.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.
- Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung CK, Chen L, Ma J et al. 2013. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet*.
- Li Y, Zhang YP. 2012. High genetic diversity of Tibetan Mastiffs revealed by mtDNA sequences. *Chinese Sci Bull* **57**(13): 1483-1487.
- Liang Y, Hua Z, Liang X, Xu Q, Lu G. 2001. The crystal structure of bar-headed goose hemoglobin in deoxy form: the allosteric mechanism of a hemoglobin species with high oxygen affinity. *J Mol Biol* **313**(1): 123-137.

- Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press.
- Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet* **39**: 197-218.
- Patel SA, Simon MC. 2008. Biology of hypoxia-inducible factor-2alpha in development and disease. *Cell Death Differ* **15**(4): 628-634.
- Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, Tao X, Wu T, Ouzhuluobu, Basang et al. 2011. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Mol Biol Evol* **28**(2): 1075-1081.
- Percy MJ, Beer PA, Campbell G, Dekker AW, Green AR, Oscier D, Rainey MG, van Wijk R, Wood M, Lappin TR et al. 2008a. Novel exon 12 mutations in the HIF2A gene associated with erythrocytosis. *Blood* **111**(11): 5400-5402.
- Percy MJ, Furlow PW, Lucas GS, Li X, Lappin TR, McMullin MF, Lee FS. 2008b. A gain-of-function mutation in the HIF2A gene in familial erythrocytosis. *The New England journal of medicine* **358**(2): 162-168.
- Piccinini M, Kleinschmidt T, Jurgens KD, Braunitzer G. 1990. Primary structure and oxygen-binding properties of the hemoglobin from guanaco (Lama guanacoe, Tylopoda). *Biol Chem Hoppe Seyler* **371**(7): 641-648.
- Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW et al. 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* **19**(5): 826-837.
- Polzin T, Daneshmand SV. 2003. On Steiner trees and minimum spanning trees in hypergraphs. *Oper Res Lett* **31**(1): 12-20.
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J et al. 2012. The Pfam protein families database. *Nucleic Acids Res* **40**(Database issue): D290-301.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**(3): 559-575.
- Qiu Q, Zhang GJ, Ma T, Qian WB, Wang JY, Ye ZQ, Cao CC, Hu QJ, Kim J, Larkin DM et al. 2012. The yak genome and adaptation to life at high altitude. *Nature Genetics* **44**(8): 946-+.
- Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, Green RK, Goodsell DS, Prlic A, Quesada M et al. 2013. The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res* **41**(Database issue): D475-482.
- Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**(1): 137-138.
- Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**(4): 725-738.

- Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* **78**(4): 629-644.
- Scheinfeldt LB, Soi S, Thompson S, Ranciaro A, Woldemeskel D, Beggs W, Lambert C, Jarvis JP, Abate D, Belay G et al. 2012. Genetic adaptation to high altitude in the Ethiopian highlands. *Genome Biol* **13**(1): R1.
- Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB et al. 2010. Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**(5987): 72-75.
- Somer T, Meiselman HJ. 1993. Disorders of blood viscosity. *Annals of medicine* **25**(1): 31-39.
- Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H et al. 2002. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* **12**(10): 1611-1618.
- Storz JF, Moriyama H. 2008. Mechanisms of hemoglobin adaptation to high altitude hypoxia. *High Alt Med Biol* **9**(2): 148-157.
- Storz JF, Runck AM, Sabatino SJ, Kelly JK, Ferrand N, Moriyama H, Weber RE, Fago A. 2009. Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proceedings of the National Academy of Sciences of the United States of America* **106**(34): 14450-14455.
- Storz JF, Sabatino SJ, Hoffmann FG, Gering EJ, Moriyama H, Ferrand N, Monteiro B, Nachman MW. 2007. The molecular basis of high-altitude adaptation in deer mice. *PLoS Genet* **3**(3): e45.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**(10): 2731-2739.
- Tang H, Peng J, Wang P, Risch NJ. 2005. Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* **28**(4): 289-301.
- van Patot MCT, Gassmann M. 2011. Hypoxia: Adapting to High Altitude by Mutating EPAS-1, the Gene Encoding HIF-2 alpha. *High Altitude Medicine & Biology* **12**(2): 157-167.
- Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, Fall T, Seppala EH, Hansen MS, Lawley CT et al. 2011. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet* **7**(10): e1002316.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol* **4**(3): e72.
- Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt JD, Boyko AR, Earl DA, Auton A et al. 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**(7290): 898-902.

- Wang GD, Zhai WW, Yang HC, Fan RX, Cao X, Zhong L, Wang L, Liu F, Wu H, Cheng LG et al. 2013. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nature Communications* **4**.
- Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**(16): e164.
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**(9): 1189-1191.
- Weber RE, Lalthantluanga R, Braunitzer G. 1988. Functional characterization of fetal and adult yak hemoglobins: an oxygen binding cascade and its molecular basis. *Arch Biochem Biophys* **263**(1): 199-203.
- Xu S, Li S, Yang Y, Tan J, Lou H, Jin W, Yang L, Pan X, Wang J, Shen Y et al. 2011. A genome-wide search for signals of high-altitude adaptation in Tibetans. *Mol Biol Evol* **28**(2): 1003-1011.
- Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**(1): 76-82.
- Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**(5987): 75-78.
- Yin S, Ding F, Dokholyan NV. 2007. Eris: an automated estimator of protein stability. *Nature methods* **4**(6): 466-467.
- Zhang J, Hua Z, Tame JR, Lu G, Zhang R, Gu X. 1996. The crystal structure of a high oxygen affinity species of haemoglobin (bar-headed goose haemoglobin in the oxy form). *J Mol Biol* **255**(3): 484-493.
- Zhao M, Kong QP, Wang HW, Peng MS, Xie XD, Wang WZ, Jiayang, Duan JG, Cai MC, Zhao SN et al. 2009. Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc Natl Acad Sci U S A* **106**(50): 21230-21235.

A**B****C**