Population Variation Revealed High-Altitude Adaptation of Tibetan Mastiffs

Yan Li,1,2 Dong-Dong Wu,*,1,2 Adam R. Boyko,3 Guo-Dong Wang,1,2 Shi-Fang Wu,1 David M. Irwin,1,4,5 and Ya-Ping Zhang*,1,2,6

1State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
2Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China
3Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University
4Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada
5Banting and Best Diabetes Centre, University of Toronto, Toronto, Ontario, Canada
6Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, China

*Corresponding author: E-mail: wudongdong@mail.kiz.ac.cn; zhangyp@mail.kiz.ac.cn.
Associate editor: Anna Di Rienzo

Abstract

With the assistance of their human companions, dogs have dispersed into new environments during the expansion of human civilization. Tibetan Mastiff (TM), a native of the Tibetan Plateau, was derived from the domesticated Chinese native dog and, like Tibetans, has adapted to the extreme environment of high altitude. Here, we genotyped genome-wide single-nucleotide polymorphisms (SNPs) from 32 TMs and compared them with SNPs from 20 Chinese native dogs and 14 gray wolves (Canis lupus). We identified 16 genes with signals of positive selection in the TM, with 12 of these candidate genes associated with functions that have roles in adaptation to high-altitude adaptation, such as EPAS1, SIRT7, PLXNA4, and MAFG that have roles in responses to hypoxia. This study provides important information on the genetic diversity of the TM and potential mechanisms for adaptation to hypoxia.

Key words: high-altitude adaptation, Tibetan Mastiff, domestication.

With the assistance of their human companions, dogs have dispersed into new environments as human civilization expanded during the Paleolithic (Germonpre et al. 2009). Among these dogs was the Tibetan Mastiff (TM), which has adapted to the extremely high altitude of the Tibetan Plateau (typically 4,500 m) (Messerschmidt 1983). As a guardian of property, the TM has acquired enhancements to its metabolic capacity to maintain a powerful body and physiological adaptations for persistence at high elevations that compensate for hypoxia (Cui and Guo 2010). Research on multiple species has revealed the identity of various candidate genes that may contribute to high-altitude adaptation, including studies on humans (Bigham et al. 2009; Beall et al. 2010; Simonson et al. 2010; Yi et al. 2010; Peng et al. 2011; Xu et al. 2011; Scheinfeld et al. 2012; Xiang et al. 2013), deer mice (Storz et al. 2007; Chevrion et al. 2012), Tibetan antelope (Ge et al. 2013), some birds (Cai et al. 2013; Qu et al. 2013), and yak (Qiu et al. 2012). The TM is derived from the Chinese Native dog (CN), which resides at lower altitudes (Li and Zhang 2012). To address the issue of divergence, we take advantage of this close genetic similarity and use the CNs as a powerful control for dissecting the biological basis of high-altitude adaptation in TMs. Here, we genotyped more than 120,000 single-nucleotide polymorphisms (SNPs) across the genome in 32 TMs, and integrated data from 20 CNs and 14 gray wolves from our previous analysis (Li et al. 2013), to study the adaptive evolution of TMs.

Analyses of Population Structure and Variation

We first used principal component analysis (PCA) to examine the SNP data. The first two axes in an initial clustering analysis using PCA identified two distinct groups, one being the wild relatives (wolves) and the second cluster consisting of the domesticated dogs (fig. 1A). The third axis of the PCA separated the TMs from the CNs (fig. 1B). Phylogenetic trees constructed using the weighted neighbor-joining method (Bruno et al. 2000) with pairwise FST values among the different breed populations, including wolves, red wolves, and coyotes, supported a close relationship between CNs and TMs.
and other ancient dogs, indicating that the TMs and CNs are ancient breeds (supplementary data S1, Supplementary Material online). To verify the PCA-based inferences of the genetic structure, we next implemented a Bayesian model-based approach to assess the population relationship of our samples of TMs, CNs, and gray wolves using the program STRUCTURE (Pritchard, Stephens, Rosenberg, et al. 2000). Domesticated dogs are clearly distinguished from the wild gray wolves, assuming two ancestral populations (K = 2), a result that corroborates previous studies (vonHoldt et al. 2010, 2011). TMs and CNs are marginally separated when K = 3 (fig. 1C), suggesting a much closer relationship between these two populations.

Next, we analyzed the population variation. Observed heterozygosity was highest in the TMs (H_0 = 29.93%), with the CNs showing slightly lower levels (29.23%), with the wolf being lowest (24.4%). Similar inbreeding coefficients are observed in the CNs and TMs (0.18 and 0.16, respectively). Gray wolves captured from the wild presented a relatively high inbreeding coefficient (F = 0.32) and low level of heterozygotes (24.4%), which are consistent with previous reports on runs of homozygosity in wolves (Boyko et al. 2010). A possible explanation for this observation may be due to the extinction of wolves from eastern and southern Asia. Alternatively, SNP ascertainment may have overestimated the number of derived polymorphisms in dogs and
underestimated the variation within wolves. To better understand population variation, haplotype-based analyses were performed. Linkage disequilibrium (LD) decay analysis revealed that longer blocks exist in the TM compared with CNs (fig. 1D). In addition, average EHH values within 500-kb windows across the genome did not change much among different core regions with different number of core haplotypes in CNs (EHH range from 0.099 to 0.175), TMs (EHH range from 0.115 to 0.266), and wolves (EHH range from 0.078 to 0.251) (fig. 1E). However, this pattern differs from that seen in breed dog, where an apparent negative relationship was observed from 0.872 to 0.541 (Li et al. 2013).

Genome-Wide Mapping of Selected Genes in the TM Genome

Compared with the CNs in highland, TMs have lower hemoglobin levels in the blood (Wen and Yuan 1998) but have excellent endurance in their locomotory abilities in highland (Kunming Police Dog Base, Ministry of Public Security, Kunming, China, personal communication), which illustrates potential adaptations for the hypoxia found at high altitudes. To attempt to identify potential mechanisms for the adaptation in Tibet Plateau, we searched for loci that show evidence of selection specific to the TM using the statistic \( \hat{d}_i \) (Akay et al. 2010), which evaluates the population differentiation of TM from the other two populations. A total of 353 SNPs, which were within the top 1% of the distribution (\( \hat{d}_i > 5.117 \)), were identified and the 139 genes near these sites (supplementary table S1, Supplementary Material online). In addition, we calculated pairwise XP-EHH values, a haplotype-based parameter that is robust to ascertainment bias (Sabeti et al. 2006), for each SNP between the TMs and CNs and identified 115 genes that were near the 378 SNPs within the 99% percentile (XP-EHH > 0.806) (supplementary table S2, Supplementary Material online). We merged the SNP lists generated by these two approaches to identify a subset of 33 SNPs that showed the strongest signature of selection by displaying both high \( \hat{d}_i \) and XP-EHH values, and these SNPs mapped to 11 genes (fig. 2). Of these 11 candidate genes showing strong signatures of selection, 9 are illustrated in figure 3 as having plausible biological functions associated with high-altitude adaptation and should be targets for further functional studies. These genes are EPAS1, SIRT7, PLXNA4, MAFG, ENO3, DNAH9, KIF1C, KIF16B, and MR (fig. 3).

Hypoxia-inducible factors (HIFs) are a group of transcription factors that regulate cellular adaptation to hypoxia and accumulate under hypoxic conditions (fig. 3). EPAS1, also known as HIF2\( \alpha \), is one of these HIFs and has previously been implicated in hypoxia adaptation in Tibetans through its association with hemoglobin concentration (Beall et al. 2010; Yi et al. 2010; Peng et al. 2011; Xu et al. 2011). In TMs, the mean hemoglobin concentration is lower than that in CNs in highland (Wen and Yuan 1998), thus is similar to the pattern seen in Tibetans. This observation supports the possibility that convergent evolution in the mechanisms in dealing with hypoxia has occurred in dogs and humans. A second candidate selected gene in TMs, SIRT7, negatively regulate HIF-1\( \alpha \) and HIF-2\( \alpha \) (Hubbi et al. 2013) and has important functions in the homeostasis of cardiomyocytes (Vakhrusheva et al. 2008).

Angiogenesis, which is largely initiated by the HIF pathway (Rey and Semenza 2010), is a crucial response to hypoxia. Indeed, candidate genes that contribute to angiogenesis and/or vasculogenesis were universally detected in species living at high altitudes, although there was little overlap in specific genes among species. NOS3 was detected in the...
Energy production is critical for survival in the hypoxic conditions of high altitude. Genes involved in energy metabolism have frequently been found in studies of positive selection in high-altitude animals (Qiu et al. 2012; Cai et al. 2013; Ge et al. 2013; Qu et al. 2013). Here, several genes involved in energy metabolism also show signatures of selection in TMs. ENO3, one of the three enolase isoenzymes expressed in skeletal muscle cells of the adult, is a key enzyme involved in glycolysis to producing ATP. In our study, the gene ENO3 was identified as a candidate selected gene in the TM. Activation of ENO3 by cortisol may offset some of the effects of having reduced levels of aldosterone secretion at high altitudes (Maher et al. 1975; Heiniger et al. 2003) (fig. 3). Although the genetic basis for positive selection on ENO3 is unclear, changes in the level of expression of ENO3 or activity levels between MR and glucocorticoids/mineralocorticoids may have been selected in TMs.

A total 16 genes that are overlapped between genes having high di SNPs and genes having high XP-EHH values were retrieved, which includes the above 11 genes as well as 5 additional genes (fig. 2): NRF6, ENTHD2, SLC38A10, ESYT3, and RYR3. NRF6, ENTHD2, SLC38A10, ESYT3, and RYR3, a ryanodine receptor, functions to release Ca$^{2+}$, playing a critical role in pulmonary vasoconstriction. Microtubules, which when damaged repress HIF-1α translation, are critical for nuclear trafficking and the transcriptional activity of HIF-1 (Carbonaro et al. 2012). DNAH9, KIF1C, and KIF16B are motor proteins (like dynein and kinesin), move along microtubules, powered by the hydrolysis of ATP, and play essential roles in intracellular transport. 3) Cortisol is elevated in the blood of individuals living at high altitudes (Moncada et al. 1965; Sawhney et al. 1991; Gardner et al. 2001). Under hypoxic conditions, HIF-1-dependent gene expression can be upregulated by glucocorticoids via GR (glucocorticoid receptor) (Kodama et al. 2003). MR (mineralocorticoid receptor) can be activated equally by mineralocorticoids (e.g., aldosterone, deoxycorticosterone) and glucocorticoids. High-altitude hypoxia can downregulate the activity of 11β-HSD2 (corticosteroid 11-beta-dehydrogenase isozyme 2), an enzyme that converts cortisol to the inactive cortisone in humans, which subsequently would cause an enhanced activation of MR or GR by cortisol that should offset, at least in part, the effects of reduced levels of aldosterone secretion at high altitudes (Maher et al. 1975; Heiniger et al. 2003).

Tibetan antelope (Ge et al. 2013); SRF, TXNRD2, and WNT7B in the ground tit (Qu et al. 2013); ADAM17 in the yak (Qu et al. 2012); VEGF in Andeans (Bigham et al. 2009); VAV3 in Ethiopian highlanders (Scheinfeldt et al. 2012); and here we found PLXNA4, PLXNA4 has roles in enhancing VEGF (vascular endothelial growth factor) signaling to promote angiogenesis (Kigel et al. 2012) and may have been selected in TMs for both roles.

Both members of the globin gene family and genes associated with globin gene expression have been hypothesized as having roles in adaptation to high altitude (Storz and Moriyama 2008). Several genes from this category with selective signatures were identified in studies in Tibetans (e.g., HBB and NFE2) (Bigham et al. 2010; Yi et al. 2010) and Ethiopian highlanders (e.g., THRB and ARNT2) (Scheinfeldt et al. 2012). In our study, the gene MAFG, which functions as a partner for p45 NF-E2 in activating α- and β-globin gene expression and promotes erythroid differentiation (Blank et al. 1997), was identified as having a selective signature.
possibility that one gene in each cluster was adaptively selected, while the other genes were carried by the selective sweep. Which was selected and which was swept cannot be distinguished with our data. A second caveat is the ascertainment bias of the genome-wide SNP data. Indeed, a significant difference between the frequency spectrum of the actual genotyped SNPs and the calculated corrected frequency spectrum was found (data not shown); however, the long-range haplotype test is relatively robust to SNP ascertainment bias (Sabeti et al. 2006). Both population differentiation and long-range haplotype tests were used to detect signatures for positively selected sites, increasing the likelihood that we identified true targets. Obtaining whole-genome sequences from multiple TM individuals would be necessary for validating potential selective targets, which should also lead to the identification of additional key genes responsible for the adaptation of TMs to their home on the Tibetan plateau. In addition, these candidate genes need additional functional analysis to validate the potential roles of the selected alleles for high-altitude adaptation.

Materials and Methods

SNPs from 40 TM individuals were genotyped using the Affymetrix v2 Canine array based on the CanFam 2.0 assembly. SNPs in the CNs and the wild gray wolves were described in our previous study (Li et al. 2013). A platinum library file (DogSNPs520431P) and the MAGIC (Multidimensional Analysis for Genotype Intensity Clustering) calling algorithm (Boyko et al. 2010) were used to ensure the reliability and robustness of the SNP annotation and the accuracy of the genotyping, yielding a total of 48,445 high-quality genotyped SNPs. We discarded SNPs that had > 10% missing data or that failed the Hardy–Weinberg equilibrium test (P < 0.001) within each population using PLINK (Purcell et al. 2007), resulting in a retained data set of 35,266 SNPs for the downstream analyses. Further, to assess recent relationships among samples, following previous methods (vonHoldt et al. 2010), we used known parental relationships to calibrate identity-by-state (IBS) scores (0.84). Of the 40 original samples, 8 individuals were excluded due to their high pairwise genetic similarity with another one (IBS > 0.84, suggesting close relationship), leaving 32 individuals as unrelated samples used in the downstream analyses. PCA was performed using R (www.r-project.org, last accessed February 21, 2014), based on the genotype data of each individual at each allele. Population structure was evaluated using STRUCTURE software (Pritchard, Stephens, and Donnelly 2000). Heterozygosity, inbreeding coefficient, and decay of LD were calculated by PLINK (Purcell et al. 2007). The haplotype of each chromosome was inferred by fastPHASE (Scheet and Stephens 2006). EHH and REHH values were calculated for each core haplotype using SWEEP software (www.broadinstitute.org/mpg/sweep, last accessed February 21, 2014). Pairwise per-SNP genetic differentiation (FST) values between breeds and di were calculated as described by Akkey et al. (2002, 2010) to evaluate population differentiation between TM, CNs, and gray wolves. XP-EHH between TMs and CNs was calculated by a program from http://hgd.p.uchicago.edu/Software/ (last accessed February 21, 2014). Candidate genes were identified if they were located within 5,000 bp of a SNP having a selective signature (Li et al. 2013). For gene identification, we chose two approaches. First, we identified SNPs within the top 1% of di and XP-EHH values. These SNPs were mapped to 11 genes. Then, we identified SNPs within the top 1% of di values, which map to 139 genes, and SNPs with top 1% XP-EHH values, which map to 115 genes. The overlap between these two sets was 16 genes having SNPs with high di values and having SNPs with high XP-EHH values.

Supplementary Material

Supplementary data S1 and tables S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Drs Bridgett M. vonHoldt and Robert K. Wayne for their suggestions and comments on the manuscript. They thank Bing Liu from CAS Tibetan Mastiff Co., Ltd for his help with sampling the Tibetan Mastiff. This work was supported by grants from the National Natural Science Foundation of China (31321002, 91231108, 31132005), and the 973 program (2013CB835200).

References


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Mallar J, Sklar P, De Bakker PIW, Daly MJ. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81:559–575.


