

MOLECULAR BIOLOGY & GENETICS

Special Topic: Convergent Adaptive Evolution

Convergent genomic signatures of high-altitude adaptation among domestic mammals

Dong-Dong Wu^{1,5,10,*}, Cui-Ping Yang^{3,5,†}, Ming-Shan Wang^{1,5,†}, Kun-Zhe Dong^{4,†}, Da-Wei Yan^{2,†}, Zi-Qian Hao^{11,†}, Song-Qing Fan⁷, Shu-Zhou Chu⁷, Qiu-Shuo Shen^{3,5}, Li-Ping Jiang^{3,5}, Yan Li⁶, Lin Zeng^{1,5}, He-Qun Liu^{1,5}, Hai-Bing Xie^{1,5}, Yun-Fei Ma^{1,5}, Xiao-Yan Kong², Shu-Li Yang², Xin-Xing Dong², Ali Esmailizadeh⁸, David M Irwin^{1,9}, Xiao Xiao³, Ming Li³, Yang Dong^{1,5}, Wen Wang^{1,5}, Peng Shi^{1,5}, Hai-Peng Li^{10,11}, Yue-Hui Ma^{4,*}, Xiao Gou^{2,*}, Yong-Bin Chen^{3,5,10,*} and Ya-Ping Zhang^{1,5,6,10,*}

ABSTRACT

Abundant and diverse domestic mammals living on the Tibetan Plateau provide useful materials for investigating adaptive evolution and genetic convergence. Here, we used 327 genomes from horses, sheep, goats, cattle, pigs and dogs living at both high and low altitudes, including 73 genomes generated for this study, to disentangle the genetic mechanisms underlying local adaptation of domestic mammals. Although molecular convergence is comparatively rare at the DNA sequence level, we found convergent signature of positive selection at the gene level, particularly the *EPAS1* gene in these Tibetan domestic mammals. We also reported a potential function in response to hypoxia for the gene *C10orf67*, which underwent positive selection in three of the domestic mammals. Our data provide an insight into adaptive evolution of high-altitude domestic mammals, and should facilitate the search for additional novel genes involved in the hypoxia response pathway.

Keywords: high altitude, Tibetan Plateau, genome, convergent evolution, domestic animals

INTRODUCTION

Phenotypic convergence generally describes independent acquisition of the same or similar traits among different species [1]. Strictly defined, molecular convergence involves substitution of the same alleles in different lineages with independent origins, which means that convergent evolution refers to independent changes at a particular site from different ancestral amino acids to the same derived amino acid [1]. However, loosely defined, convergent evolution could occur at the same genes but different sites to generate the same or similar phenotypic consequence in independent lineages [2,3].

High-altitude environments are considered physiologically challenging to endothermic animals.

Plateaus, which take up a third of the Earth's land, are believed to provide fertile ground for investigating mechanisms of adaptation to hypoxia, and, particularly, convergent evolution [4,5]. To date, numerous studies have revealed special convergent physiological and morphological adaptations to hypoxia in many vertebrate species from highland. For example, endotherms living in high plateau environments have enhanced cardio-pulmonary function, increased-size hearts and lungs, as well as decreased morbidity as a result of pulmonary diseases compared with closely related species from lowland environments [6]. Tibetans were found to exhibit characteristics contributing to improved hypoxic and hypercapnic ventilatory responsiveness

¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; ²Key Laboratory of Animal Nutrition and Feed Science of Yunnan Province, Yunnan Agricultural University, Kunming 650201, China; ³Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; ⁴Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China; ⁵Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China; ⁶State Key Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming 650091, China; ⁷Department of Pathology, the Second Xiangya Hospital, Central South University, Changsha 410011, China; ⁸Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, PB 76169-133, Iran; ⁹Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, M5S 1A8, Canada; ¹⁰Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China and ¹¹CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

*** Corresponding authors.**

E-mails: wudong-dong@mail.kiz.ac.cn; zhangyp@mail.kiz.ac.cn; ybchen@mail.kiz.ac.cn; gouxiaosa@163.com; mayuehui@caas.cn

[†]Equally contributed to this work.

Received 29 November 2019; Accepted 18 December 2019

including larger lung volumes, better pulmonary function and greater lung-diffusing capacity [7]. They also maintain higher levels of arterial oxygen saturation both at rest and during exercise, and show reduced loss of aerobic performance with increasing altitude [7]. Several aspects of the genetic basis for the adaptation to hypoxia have been elucidated in some vertebrate species [6,8–24]. In particular, these studies suggest that modifications of hemoglobin function are key to mediating an adaptive response to high-altitude hypoxia in a number of mammals, birds and amphibians [8,10,11,25,26]. Despite this, much remains to be learned about the molecular genetic basis underlying high-altitude adaptation in general. To achieve this goal, analyzing whether molecular convergent evolution occurs commonly for high-altitude adaptation among different species may provide important insights.

The Tibetan Plateau, with an average elevation of over 4500 m, is the largest and highest plateau in the world. The extremely harsh environment of this plateau includes low oxygen content, low air pressure, low temperatures and strong UV radiation. Indeed, the unique geographical and environmental features have shaped the biological diversity of the Tibetan Plateau, and provide fertile ground for investigating the genetic loci underlying adaptation to high altitude in humans, in wild animals such as the chiru and Tibetan wolf, as well as in domestic animals such as the yak, Tibetan chicken, Tibetan Mastiff (dog) and Tibetan pig [6,12–24]. Previous findings have shown that animals primarily domesticated at low altitudes (sheep, goats, cattle, chickens, donkeys, pigs, dogs and horses) have been brought to the Tibetan plateau in the past few thousand years by Tibetans, which likely facilitated the permanent human occupation of this plateau [27]. Although these Tibetan domestic animals have inhabited the plateau for less than 4000 years [27], they harbor special physiological traits that allow their survival in Tibet. Studying the genetic basis of high-altitude adaptation in these animals will not only promote our understanding of their inhabitation history, but also provide essential information about the molecular convergent evolution process.

In the present study, to reveal the genetic mechanisms underlying high-altitude adaptation of domestic mammals, we used genomes, with an additional 73 generated for this study, from horse, sheep, goat, cattle, pig and dog populations that live in the Tibetan Plateau and at low altitudes. We investigated positive selection operating on these Tibetan domestic mammals to detect the occurrence of convergent evolution for high-altitude adaptation.

RESULTS

In the present study, a total of 327 (134 individuals inhabiting the Tibetan Plateau including 19 goats, 24 horses, 20 sheep, 11 dogs, 19 cattle and 41 pigs, and 193 control individuals from lowland areas including 20 goats, 10 horses, 70 sheep, 34 dogs, 30 cattle and 29 pigs, Supplementary Table S1) whole genome sequences (including 73 new sequences) were analyzed to identify genome-wide SNPs. To explore potential adaptation to high altitude, differentiation at SNPs in each species was calculated using XP-EHH (cross population extended haplotype homozygosity) [28], Δ DAF (the difference of the derived allele frequencies) and F_{ST} [29,30] between the highland and lowland populations to evaluate their evolutionary properties.

Patterns of convergent evolution at the DNA sequence level

We first studied potential convergent evolution at the DNA sequence level. SNPs from each domestic mammal were coordinated to the human genome (hg19) using the LiftOver program (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>). Through this analysis, we found an average of 182 K SNPs occurring at the same human coordinated position for any pair of domestic mammals (Supplementary Table S2). We retrieved SNPs showing values of F_{ST} , XP-EHH and Δ DAF higher than the 99th percentile values of genome-wide SNPs, and found that the proportion of SNPs showing higher values is nearly 1% among these SNPs occurring at the same human coordinated position for any pair of domestic mammals (Supplementary Table S3). We further identified potential convergent sites, at which SNPs showed higher values of F_{ST} , XP-EHH or Δ DAF in both species for any pair comparison. This analysis identified only 3–45 sites showing potential convergent evolution between any pair of Tibetan domestic mammals (Supplementary Table S4). This analysis supports the previous finding that molecular convergence at the DNA sequence level is not common [31–33].

Shared signatures of positive selection in Tibetan domestic mammals at the gene level

We integrated the F_{ST} , XP-EHH and Δ DAF analyses (named *iFXD*) as previously described [34] to improve our power and to identify positively selected genes in the six Tibetan domestic mammals. Based on the outlier approach in the top 1% ranking, in total, 199, 187, 190, 181, 189 and

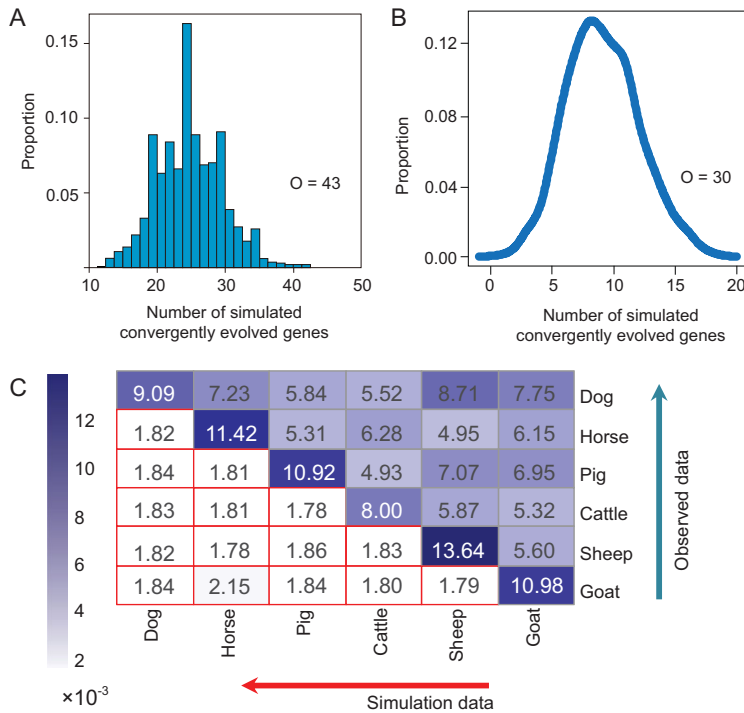


Figure 1. Frequent shared targets of positive selection in Tibetan domestic mammals. (A) Distribution of the numbers of simulated shared positively selected genes. Observed number of shared positively selected genes was 43 ($O = 43$). Namely, 43 protein coding genes evolved under positive selection in at least two Tibetan domestic mammals (in the top 1% ranking). (B) Distribution of the numbers of simulated shared positively selected genes based on one-to-one orthologous genes. We further investigated evidence of convergent signature of positive selection in 10 103 one-to-one orthologous genes. The number of genes observed to potentially evolve under positive selection (in the top 5% ranking) in at least three mammals is 30 ($O = 30$) (Supplementary Fig. S2). (C) Observed and simulated ratios of gene-gene interactions among positively selected genes within species and between positively selected genes of different species. Lower red circled boxes present the simulated ratios, and above blue boxes are observed ratios ($\times 10^{-3}$).

170 protein coding genes were identified to exhibit extreme divergence with evidence for positive selection in goats, horses, sheep, dogs, cattle and pigs, respectively (Supplementary Table S5). After inference of demographic history and population simulation, we found that the outlier approach is robust to demographic history and SNP density (Note S1, Supplementary Figs S1–S4 and Tables S6–S10). Among these, 43 protein coding genes were found to display the evidence for positive selection in at least two lineages (Fig. 1A, Supplementary Fig. S5). In contrast, only 29 protein coding genes were found to display the evidence for positive selection in the control (lowland population) in at least two lineages ($P = 0.036$, $\chi^2 = 4.42$). To further demonstrate that these were not false-positives, we conducted a permutation test by randomly sampling genes. In all 1000 random samples, the numbers of shared positively selected genes was less than 43 (Fig. 1A).

We further investigated evidence of shared positive selection in a set of one-to-one orthologous genes by integrating signals of selection among the six Tibetan domestic mammals. We retrieved 30 genes that had potentially evolved under positive selection (in the top 5% ranking) in at least three mammals (Supplementary Fig. S6). When we examined 1000 sets of randomly chosen genes, none of the datasets contained more than 20 genes under positive selection in at least three mammals (Fig. 1B). These data indicate that convergent signature of positive selection occurred among these different species.

We then took one further step to examine whether positive selection was occurring on particular single genes or on genes within certain pathways. For this, we conducted a gene enrichment analysis, and found that a gene-gene interaction network containing *EPAS1* was overrepresented in the candidate positively selected genes in each of the Tibetan mammals examined (Supplementary Fig. S7). *EPAS1* encodes a subunit of the HIF transcription factor and is a key gene for hypoxia adaptation in Tibetans [12,13,22]. In our analyses, this gene was found to have evolved under positive selection in five Tibetan mammals (top 1% ranking) and in Tibetan cattle (top 5% ranking), corroborating the previously reported convergence of genetic adaptation to high altitude in dogs and humans [12,13,16–18,22]. These data suggested that during adaptive evolution, positive selection drove changes in genes within a network involved in similar functional pathways rather than single genes.

Inspired by the above findings, we further investigated whether positively selected genes exhibited more frequent interactions with each other among different species. As expected, positively selected genes displayed significantly more gene-gene interactions with each other than randomly selected genes (Fig. 1C). In addition, positively selected genes of one species also interacted significantly with those of other species based on the gene-gene interaction data from humans (Fig. 1C).

Common rapid evolution of hypoxia response genes in Tibetan domestic mammals

To validate the contention that positive selection drove changes in genes within a network involved in similar functional pathways, we examined the evolution of genes involved in the hypoxia response pathway, which might play an important role in high-altitude adaptation. The hypoxia response genes were defined in the study [35], which used the AmiGO tool (<http://amigo.geneontology.org>) to retrieve all genes within the Gene Ontology

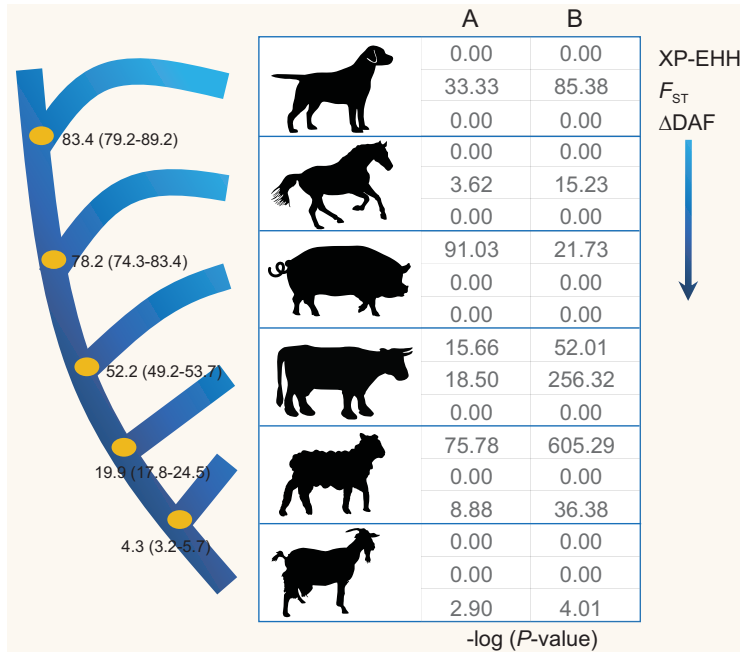


Figure 2. Rapid evolution of hypoxia response genes in six Tibetan domestic mammals. Significantly higher values of XP-EHH (cross population extended haplotype homozygosity), F_{ST} and ΔDAF (the difference of the derived allele frequencies) were observed for SNPs found in genes known to be involved in hypoxia response than in other genome-wide genes or randomly selected genes. Columns A (compared with genome-wide genes) and B (compared with random genes) represent the $-\log_{10}$ transformed P values calculated using Mann–Whitney U tests. The hypoxia response genes were retrieved from a previous study [35]. Numbers on the nodes of the phylogenetic tree are the divergence times from present (millions of years) and their confidence intervals. Dates are from a previous study [67].

biological process term ‘response to hypoxia’ plus all descendent terms (GO:0001666 ‘response to hypoxia,’ GO:0071456 ‘cellular response to hypoxia’ and GO:0070483 ‘detection of hypoxia’). F_{ST} , XP-EHH and ΔDAF analyses got different results, which is likely caused by the different statistical principles used by these methods, as described in a previous study [36]. However, as seen in the values of F_{ST} , XP-EHH and ΔDAF , SNPs in a set of hypoxia response genes displayed significantly higher levels of differentiation compared with SNPs in the remaining genome-wide set of genes ($P < 0.01$, by Mann–Whitney U test, Fig. 2). To further validate the significance and exclude the confounding factor of different numbers of genes for comparison, we randomly selected a group of genes with the same number with hypoxia response genes for each species. As expected, we got similar significant results ($P < 0.01$, Fig. 2). In addition, SNPs within genes in the gene ontology (GO) category associated with hypoxia also harbored significantly higher values for F_{ST} , XP-EHH and ΔDAF (Supplementary Fig. S8). These lines of evidence indicated

common rapid evolution of hypoxia response genes, allowing adaptation to hypoxia in these Tibetan domestic mammals.

Convergent signature of positive selection on EPAS1 for attenuating hemoglobin concentration in Tibetan humans and Tibetan horses

The above finding suggested that, although signatures of positive selection were significantly shared among different species, the sites targeted by selection in genes under convergent evolution are not identical. We reasoned that positive selection on the same genes might generate convergent functions or phenotypes in different species even though the sites targeted by selection were different. To validate this, we examined *EPAS1*, which is under positive selection in Tibetan humans and Tibetan domestic mammals, to determine whether changes at different sites in different species generated similar functional consequences. In Tibetan humans, adaptive variants of *EPAS1* favor adaptation to high altitude by attenuating maladaptive physiological alterations such as elevated hemoglobin concentration [37]. We choose a SNP (chr15: 52623471), which harbors the highest *iFXD* value in horse *EPAS1* and tested its association with hematologic traits in the Tibetan horse. As expected, we saw that the *EPAS1* SNP showed significant correlations with some measured physiological traits (Supplementary Fig. S9, $P < 0.01$ after false discovery rate (FDR) correction). Consistent with Tibetan humans, the derived allele of *EPAS1* SNP (chr15: 52623471) in the Tibetan horse was associated with lower hemoglobin concentration (HGB) (Supplementary Fig. S9), and its frequency in the Tibetan and lowland horses was 0.79 and 0.28, respectively. One may argue that the association may be attributable to demographic history, like frequent introgression from lowland horse. Under demographic history, SNPs across the genome showing differentiation between high and low populations would correlate with these blood phenotypes. Therefore, we further chose a SNP which changes the amino acid of *EPAS1* (R144C), and harbors the largest *iFXD* value among all of the non-synonymous SNPs examined. However, it did not show association with any of the blood-related traits. We proposed that the correlation of *EPAS1* SNP (chr15: 52623471) with blood traits is likely attributable to adaptive evolution, although we cannot absolutely exclude the possibility of demographic history.

Increased hemoglobin concentration and erythrocytosis after exposure to high altitude are counterproductive for individuals from lowlands, as these

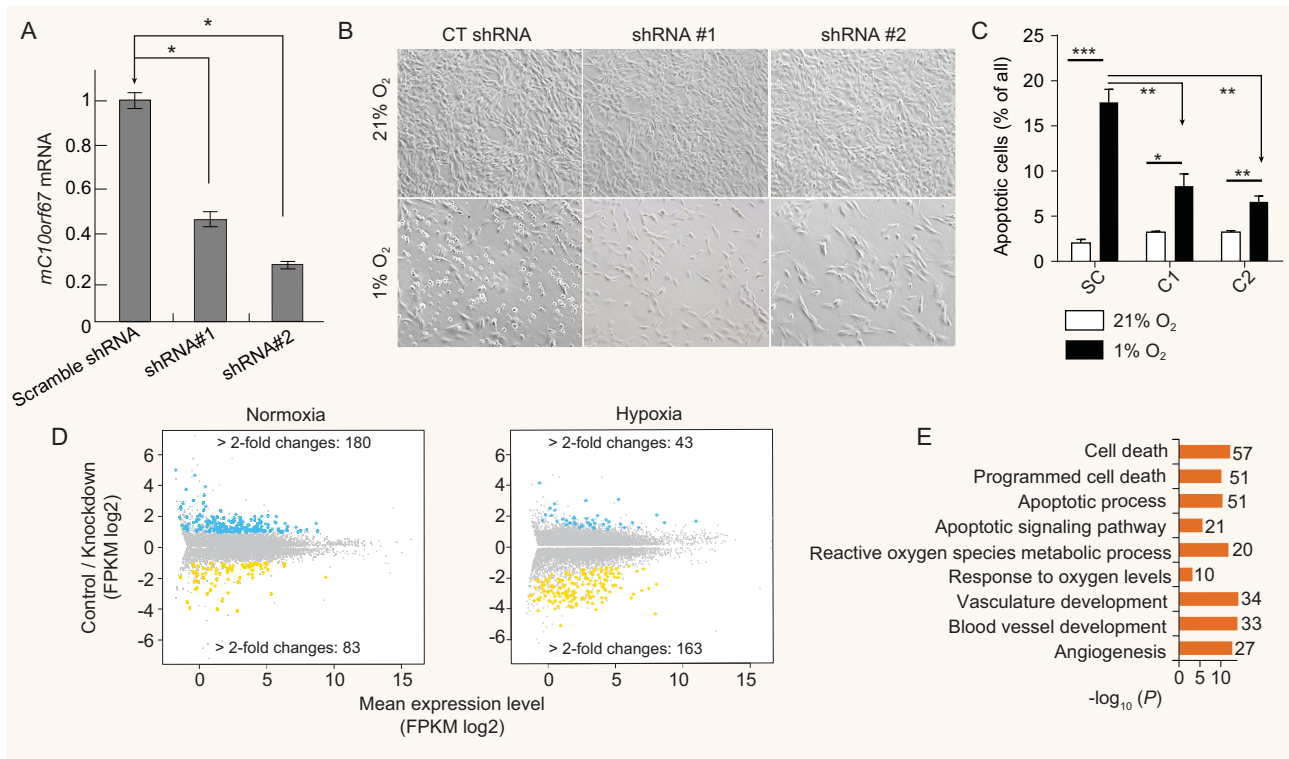


Figure 3. Functional analysis of *C10orf67*. (A) Relative mRNA expression levels of *mC10orf67* in NIH-3T3 cells stably expressing two independent shRNAs compared with scramble-control shRNA. * $P < 0.05$. (B) Suppression of apoptosis by depleting *C10orf67* expression in NIH-3T3 cells under 1% O₂ hypoxic conditions. (C) After 72 h treatment under hypoxic conditions, NIH-3T3 cells were stained with annexin V/PI, and the percentage of apoptotic cells was assessed by flow cytometry. The scramble-control shRNA stable cell line served as a control. SC, C1 and C2 indicate cells treated by control shRNA, shRNA #1 and shRNA#2, respectively. * $P < 0.05$, ** $P < 0.001$. (D) Comparison of gene expression with scramble-control and *C10orf67* knockdown by RNA-seq under normoxic (left panel) and hypoxic (right panel) conditions. A summary dot plot is shown, with blue dots representing down-regulated and yellow dots representing up-regulated genes ($n = 3$ samples for each, at least 2-fold difference and P -corrected < 0.05). (E) List of GO (gene ontology) terms related to hypoxia and apoptosis that were significantly enriched among the differentially expressed genes between the scramble-control and *C10orf67* knockdown cells under normoxic conditions by RNA-seq. The x-axis shows the $-\log_{10}$ transformed FDR corrected P values. The number of enriched genes in each GO is also presented.

changes result in increased blood viscosity that reduces uterine artery blood flow and a reduced rate of O₂ delivery to the uteroplacental circulation [4], which ultimately lead to elevated risk of stillbirth, preterm birth and reduced birth weight. In addition, individuals from low altitudes exposed to high altitudes often develop high hemoglobin concentrations associated with disadvantageous physiological outcomes including hemodynamic hyperviscosity, cardiac dysfunction and chronic mountain sickness [37]. Convergent evolution of *EPAS1* potentially occurred in Tibetan domestic mammals and Tibetan humans to attenuate maladaptive physiological alterations such as elevated hemoglobin concentration and erythrocytosis.

Function of a positively selected gene *C10orf67* in the response to hypoxia

As genes known to be involved in the response to hypoxia had evolved under positive selection in

Tibetan domestic mammals, we reasoned that novel genes involved in the hypoxia response pathway might be identified from our list of positively selected gene candidates. To test this, we examined the role of *C10orf67*, a gene with uncertain function but displayed evidence of positive selection in three Tibetan mammals (Supplementary Fig. S5, S6), in cellular response to hypoxia. Hypoxia will induce apoptosis, a form of programmed cell death. It is necessary to detect the apoptosis under hypoxic conditions to reveal the role of *C10orf67* in the response to hypoxia. We examined the rate of apoptosis in NIH-3T3 cells stably expressing two independent shRNAs targeting *C10orf67* under normoxic (21% O₂) and hypoxic (1% O₂) conditions (Fig. 3A, B). In contrast to the increased rate of apoptosis observed in control cells transfected with scrambled shRNA under hypoxia compared with normoxia, *C10orf67*-deficient cells displayed a dramatic reduction in the rate of apoptosis under hypoxic condition (Fig. 3B, C). We then used whole transcriptome sequencing

(RNA-seq) to characterize the regulatory network involving *C10orf67*. We compared the results obtained from cells depleted of *C10orf67* to those from cells transfected with scrambled shRNA under both normoxic and hypoxic conditions and observed 263 and 206 differentially expressed (differential expression was defined by having at least a 2-fold change) genes, respectively ($P < 0.01$, Fig. 3D). Intriguingly, many of the identified genes are involved in the response to changes in oxygen levels, angiogenesis, cell death and apoptotic processes (Fig. 3E, Supplementary Table S11), indicating a potential role for *C10orf67* in the response to hypoxia or during tumorigenesis.

Positive selection drove the population differentiation of non-synonymous SNPs in Tibetan sheep, dogs and cattle

Local positive selection tends to increase population differentiation, and preferentially targets functional SNPs resulting in amino acid change or cis-regulatory events, while under the assumption of neutrality, population differentiation is a result of demographic history, all loci are affected similarly [38]. In the current study, population differentiation, which was evaluated by F_{ST} [29], was more pronounced at non-synonymous SNPs than any other types of SNPs (e.g. SNPs located upstream/downstream of genes, intronic SNPs, synonymous SNPs) in sheep, dogs, cattle and horses (Supplementary Figs. S10, S11 and Table S12). Although local adaptation tends to increase population differentiation, many studies have also argued that background selection reduces the within-population diversity, and should lead to high F_{ST} [39–41]. We further calculated the derived allele frequencies of these high F_{ST} non-synonymous SNPs, and found that they harbor significantly higher derived allele frequencies in Tibetan domestic mammals, like sheep, dog and cattle (Supplementary Fig. S12). Therefore, positive selection likely has driven population differentiation of functional SNPs between Tibetan sheep, dog and cattle and other lowland populations, and likely contributes to local adaptation. For example, highly differentiated non-synonymous SNPs occurred in *EPAS1* of Tibetan dogs and cattle (Supplementary Table S13).

We further examined non-synonymous SNPs harboring high levels of population differentiation between Tibetan domestic mammals and lowland mammals ($F_{ST} > 0.5$) within these positively selected genes in each species (Supplementary Table S13). We found that these genes were mainly involved in apoptosis, metabolism and functions of the nervous system, supporting the widely accepted tight connection between hypoxia adaptation and

these pathways [42–45]. However, the functional consequences of these amino acid changing mutations remain to be determined. Functional validation of the association between these mutations and high-altitude adaptation will provide essential information about potential biomarkers for breeding at high altitude.

Other genes under positive selection in different domestic mammals

In the Tibetan sheep, we found that *MITF* (melanogenesis associated transcription factor) evolved under positive selection. *MITF* is crucial for pigmentation by regulating the differentiation and development of melanocytes [46]. The factor driving positive selection on *MITF* and whether it is related to high-altitude adaptation is unclear, but is likely to be associated with phenotypic evolution in Tibetan sheep. Indeed, while Tibetan sheep often have a white coat, 90% of the total population have variegated wool covering their heads and limbs, and solid-white or solid-black sheep are rare [47]. The positive selection observed for *MITF* is therefore likely associated with the change in coat pigmentation found in the Tibetan sheep. This variegated wool covering the heads, and particularly eyes, of the Tibetan sheep is believed to help them resist the strong UV radiation in highland areas [47].

In the Tibetan goat, the gene *DSG3* harbors the most significant signal of positive selection. A sliding window analysis (window size: 50 kb, step size: 25 kb) of the $iFXD$ values across the whole genome also revealed a region with significant values that spanned the *DSG* gene cluster (including *DSG2*, *DSG3*, and *DSG4* genes) (Supplementary Fig. S13). *DSG* encode desmosome proteins involved in the intercellular junctions of epithelia and cardiac muscle cells, and play important roles in development of the heart, as shown by mutations within desmosomes that cause arrhythmogenic right ventricular cardiomyopathy in humans [48]. Endothermic animals living on high plateaus typically have strong cardio-pulmonary function with larger hearts. The important role of *DSG* genes in heart development raised the possibility that positive selection on these genes might have yielded phenotypic changes in the Tibetan goat that were helpful for high-altitude adaptation. However, the caveat is that the underlying factor driving positive selection on gene *DSG* is not only high-altitude adaptation. For example, *DSG* genes also have important functions in the immune system [49], which is also the driving factor for rapid evolution of genes.

RYR2 (ryanodine receptor 2) functions as a component of calcium channels that supply Ca^{2+} for cardiac muscle excitation–contraction coupling.

Previous studies have reported evidence of positive selection on *RYR2* in the Tibetan chicken [50] and Tibetan wolf [24], presumably for a role in adaptation to hypoxia. Here, we also found evidence of positive selection on *RYR2* in the Tibetan goat. Hypoxia should induce the release of reactive oxygen species and cause an influx of extracellular Ca^{2+} mediated by ion channels as well as the intracellular release of Ca^{2+} from stores by the ryanodine receptor. Increased intracellular Ca^{2+} will then cause contraction of cells, an important response to hypoxia [51–53].

In Tibetan cattle, we observed positive selection on the gene *HMGA2*. Variants within *HMGA2* have been reported to be associated with body height in humans and domestic animals [54–57]. In our results, a non-synonymous SNP (A64P) within *HMGA2* displayed a high level of population differentiation between Tibetan cattle and other cattle (Supplementary Table S13). In addition, *ADH7* also evolved under positive selection in the Tibetan cattle, and harbored three non-synonymous SNPs showing a high level of differentiation between Tibetan cattle and other cattle. *ADH7* encodes a class IV alcohol dehydrogenase that is very active as a retinol dehydrogenase. The gene had been reported to evolve under positive natural selection in Tibetan humans, and was associated with their weight and body mass index [58]. In sum, we believe that positive selection on *ADH7* might be a potential convergent signature, and be associated with the stature of Tibetan cattle.

DISCUSSION

Shared targets of positive selection for local adaptation of domestic mammals

In the present study, we used large-scale genomic data to acquire insights into the adaptive evolution of Tibetan domestic mammals. Our results were indicative that targets of positive selection were commonly shared among different species. This may suggest convergent evolution occurring among different domestic mammals. However, we did not find frequent molecular convergence at the DNA level. Therefore, convergent signature of positive selection could also induce the rapid divergence of genes among these species.

One caveat of study on convergent evolution is false-positive [31–33,59]. Signal of convergent evolution may be observed frequently between control groups, such as lowland populations in this study. Therefore, more populations from lowland will be needed in future to identify positively selected genes in the lowland populations, and this would be helpful for studying patterns of convergent evolution in

the control population. Another caveat of our study is that the factors driving positive evolution are unclear. The parsimony factor is high-altitude adaptation. However, we also note that some positively selected genes are involved in immune system, which have been found to evolve rapidly in many organisms as a result of arm race hypothesis. Therefore, common rapid evolution of immune genes might also be a factor for shared signature of positive selection.

In addition, we also report the likely role of a positively selected gene *C10orf67* in the response to hypoxia. We reasoned that studying genetic loci underlying high-altitude adaptation should facilitate identification of novel genes in unique pathways such as hypoxic response.

Why is positive selection on *EPAS1* a common event in all these mammals?

Using large-scale genomic data, we reported the putative contribution of multiple positively selected genes, particularly *EPAS1*, and positive selection on amino acid changing SNPs, to high-altitude adaptation of Tibetan domestic mammals. Although we have acquired insights into the adaptive evolution of these mammals, it remains unclear which variants of *EPAS1* facilitate high-altitude adaptation of Tibetan domestic mammals. Therefore, more experiments dissecting the functional consequences of these variants are needed.

It is assumed that phenotypic traits shared by Tibetan domestic mammals are likely attributed to high-altitude adaptation changes associated with positive selection on certain genes, and *EPAS1* is likely one of those. Elevated hemoglobin concentration and erythrocytosis occur commonly during the initial short-term acclimatization responses to hypoxia. While these physiological changes affect their breeding and health [4,37], those eventually adapted must have undergone a series of physiological processes to maintain homeostasis. Therefore, one of the reasons for frequent positive selection on *EPAS1* might be to attenuate the rapidly induced common elevated hemoglobin concentration and erythrocytosis [4,37]. Given that *EPAS1* encodes one of the subunits of the hypoxia-inducible factors (HIFs) and has pleiotropic functions, it is of great interest to determine the functional consequences of selection on *EPAS1* and associated physiological processes in future studies.

Potential application of adaptive mutations in domestic mammal breeding at high altitude

The extreme environment of the Tibetan plateau limits livestock husbandry, which dominates the

Tibetan economy. To solve the problem that many of the domestic animals transported from lowland areas cannot live or reproduce well at high altitude, we sought to identify variants in genes underlying hypoxia adaptation in domestic animals so that commercial breeds of domestic mammals from lowland areas carrying advantageous variants can be generated and introduced to different plateaus. It is intriguing to confirm that *EPAS1* evolved under positive selection for hypoxia adaptation, and we believe that certain functional variants within *EPAS1* could be used to help improve livestock breeding. In addition, many mutations related to amino acid changes showed great population differentiation between highlanders and lowlanders, suggesting their potential as candidate biomarkers for economic purposes, although further studies are needed to validate their biological roles in high-altitude adaptation.

METHODS

Genome sequences and SNPs calling

In this study, blood samples from 19 goats, 20 sheep and 34 horses were collected from individuals inhabiting different regions of the Tibetan Plateau and surrounding lowland regions in China. DNA was extracted from the blood samples using a standard phenol–chloroform extraction protocol. Using the Illumina standard protocol, genomic libraries with insert size ~ 400 bp were constructed, and sequenced by HiSeq2000 or HiSeq2500. Additional genomes were from other studies. A final total of 134 genomes from Tibetan domestic mammals (19 Tibetan goats, 24 Tibetan horses, 20 Tibetan sheep, 11 Tibetan Mastiffs, 19 Tibetan cattle and 41 Tibetan pigs) and 193 genomes of domestic mammals from lowlands were used for population genomic analysis (Table S1).

Reference genome sequences of cattle, horse, pig, dog and sheep were downloaded from ENSEMBL (version 72), and the reference genome sequence of the goat was downloaded from <http://goat.kiz.ac.cn/> (CGD) [60]. Low-quality bases in sequence reads were filtered using Btrim, and qualified reads were aligned onto the references using BWA-MEM with default settings except for ‘-t 8 -M -R’ options [61]. A series of post-processes was carried out with the aligned bam format files, including sorting, duplicate marking and local realignment using the relevant tools in the Picards (picard-tools-1.56) and Genome Analysis Toolkit (GATK, GenomeAnalysisTK-2.6-4-g3e5ff60) packages [62]. SNPs were called and filtered using the UnifiedGenotyper and VariantFiltration com-

mands in GATK. Loci with RMS mapping quality < 25 and genotype quality < 40 , for which reads with zero mapping quality constituted $> 10\%$ of all reads at this site, were removed. Loci with more than two alleles and within clusters (more than three SNPs in a 10 base pair (bp) window) were removed. All SNPs were annotated using the ANNOVAR program (www.openbioinformatics.org/annovar/), which annotated different types including non-synonymous, synonymous, splicing, 5' UTR, 3' UTR, intronic, upstream, downstream and intergenic SNPs. Variants that were missing with $< 50\%$ were imputed and haplotypes were inferred using BEAGLE (<http://faculty.washington.edu/browning/beagle/beagle.html>) [63].

Genes associated with hypoxia

Sequences of 154 human genes involved in the hypoxia response pathway were downloaded from a previous study [35]. One-to-one orthologous genes between humans and domestic mammals, including dog, pig, horse, cattle and sheep, were downloaded from Ensembl using BioMart (<http://www.ensembl.org/>). One-to-one orthologous genes between humans and goat were retrieved by a reciprocal best-hit BLASTP search.

Calculation of F_{ST} , XP-EHH, ΔDAF and $iFXD$

We used F_{ST} [29,30], XP-EHH [28] and ΔDAF to evaluate the differentiation of each SNP between the highland and lowland populations in each mammal. The F_{ST} of each SNP was calculated as previously described [29]. F_{ST} with negative values having no biological explanation were arbitrarily set to 0. The XPEHH program (<http://hgdp.uchicago.edu/Software/>) was used to calculate the XP-EHH value for each SNP. Here, we did not consider variation of the recombination rate because of a lack of data for these genomes. However, the EHH method, which does not consider variation in the recombination rate, has been proven to possess the power to detect signatures of positive selection, and XP-EHH has been successfully used to detect positively selected genes in several species including dogs [17], chickens [50] and horses [54] that lack recombination rate data. ΔDAF was calculated for each SNP as the DAF in the domestic animals from the highlands minus the DAF in domestic animals from the lowlands. The ancestral allele was inferred according to the allele of the outgroup, including the European pig for Tibetan pig, yak for Tibetan cattle, golden jackal for Tibetan dog, Przewalski's horse for

Tibetan horse, sheep for Tibetan goat and *Ovis dalli* for Tibetan sheep.

We calculated *iFXD* for each SNP by integrating F_{ST} , XP-EHH and ΔDAF :

$$iFXD = \prod_{i=1}^n \frac{Psi}{1 - Psi},$$

where *i* incorporates the three different methods and *Ps* represents the probability that the SNP is under positive selection.

Identification of candidate genes under positive selection in the Tibetan domestic mammals based on *iFXD* values

The *iFXD* value of each protein coding gene was calculated by averaging the *iFXD* values of all SNPs within a gene. The demographic history (a genome-wide force that affects patterns of variation at all loci in a genome in a similar manner, whereas directional selection acts on specific loci [36]) of domestic mammals in Tibet might be much more complex because of genetic admixture between mammals from highland and lowland areas. Inferring and simulating the demographic history of these Tibetan domestic mammals therefore becomes technologically difficult. However, the outlier approach based on empirical data is free from any assumption of demographic history, and is considered robust to avoid the potential confounding effect of population demographic history. Here, we employed an outlier approach based on genome-wide empirical data to retrieve the top 1% of genes showing high-level *iFXD* values.

Shared positive selection of genes among different species

One-to-one orthologous genes between dogs and pigs, horses, cattle and sheep were downloaded from Ensembl (<http://www.ensembl.org/>) using BioMart. One-to-one orthologous genes between dogs and goats were retrieved by a reciprocal best-hit BLASTP search. Overall, 10 103 one-to-one orthologous genes were identified among dogs, pigs, horses, cattle, sheep and goats. The value showing convergent signature of positive selection for each gene among the six mammals was identified as:

$$CE = \prod_{i=1}^6 \frac{Psi}{1 - Psi},$$

where *i* represents the different mammals and *Ps* represents the probability of genes being under positive selection in the mammal. In total, 30 candidate genes

were identified to harbor a high *iFXD* value (95% percentile rank) in at least three species.

Data simulation

Simulation was performed by randomly choosing the same number of genes as the number of positively selected genes in each species. The number of shared genes was counted among six simulated data. One-thousand random samples were performed in total.

Gene-gene interaction data

Protein-protein interactions were compiled from two resources, InWeb (InWeb3_HC_NonRed) [64] and BioGRID (BIOGRID-ALL-3.3.124.tab) [65]. We used only non-redundant interactions, and defined all interactions as undirected edges in a binary network. The ratio of gene-gene interaction of a group of genes, or between two groups of genes, was defined as the number of gene-gene interactions divided by the total number of gene pairs. Namely, within a group of genes, $R = I/(N*(N-1))$, and between two groups of genes, $R = I/(N1*N2)$, where *R* is the ratio, *I* is the number of gene-gene interactions, and *N*, *N1* and *N2* are the numbers of genes. Simulation was performed by randomly choosing genes with the same number as positively selected genes in each species. One-thousand simulations were performed, and the simulated ratios of gene-gene interactions were averaged.

Association of variant at gene *EPAS1* with physiological parameters in Tibetan horses

As described in our previous study [16], the fore-leg venous blood was collected to measure physiological parameters in situ. Twenty-one 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. Mutations in *EPAS1* were genotyped by Sanger sequencing technology. The *F* test in the analysis of variance was used to evaluate the significance of genotypes. Association was performed in horses at the same elevation.

Experiments to study the function of *C10orf67*

To study the function of *C10orf67*, independent shRNAs targeting different regions of the gene were constructed using the pLKO.1 vector to knockdown *C10orf67* expression in NIH-3T3 cells.

The 21 bp scrambled shRNA sequences were GCACTACCAGAGCTAACTCAG and ACCCTCTGAGTTAACATTTAA; the *mC10orf67* shRNA#1 sequence was GGGCAGAGAATGAAAGGTTAA; and the *mC10orf67* shRNA#2 sequence was AGCTTTCATGTCCTAAAGAAT. The sequences of all constructs were verified by Sanger sequencing. Lentiviruses were generated according to the manufacturer's protocol. In brief, supernatants containing different lentiviruses generated from HEK-293T cells were collected at 48 and 72 h post-transfection, respectively; cells were infected successively twice by 48 h and 72 h viruses in the presence of 4 $\mu\text{g}/\text{mL}$ polybrene. Stable cell lines were selected using puromycin. For the cell proliferation assay, tested cell lines were plated into 24-well plates and cell numbers were counted each day. The numbers of apoptotic cells were determined by flow cytometry using FITC-conjugated annexin V (BD Pharmingen) staining. The pipeline of Tophat, Cufflinks and Cuffdiff was employed to find the differentially expressed genes between *C10orf67* knockdown and scrambled-shRNA control NIH-3T3 cells under normoxia and hypoxia. Gene enrichment analysis was performed by g: profiler [66].

Gene enrichment analysis

We employed g: profiler [66] (<http://biit.cs.ut.ee/gprofiler>) for enrichment analysis with Benjamini-Hochberg FDR as the multiple correction. We looked for enriched modules in the BioGRID protein-protein interaction (PPI) network using the g: profiler program. For each species, only one significant gene-gene interaction network was found ($P < 0.01$), and the network contains gene *EPAS1*.

DATA AVAILABILITY

The genome raw data are available from the SRA (www.ncbi.nlm.nih.gov/sra) under bioproject number PRJNA281979 and GSA (<https://bigd.big.ac.cn/>) under bioproject number PRJCA001963.

SUPPLEMENTARY DATA

Supplementary data are available at [NSR](https://www.nature.com/nr) online.

FUNDING

This work was supported by the National Natural Science Foundation of China (91731304, 31621062) and Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13000000). YBC was supported by grants from the National Natural Science Foundation of China (81322030, 31271579) and the Yunnan Province High-level Talents Introduced Program and Yunnan Applied Basic Research Projects (2014FA038). XG was supported by Chinese 863 program

(2013AA102S03), and the National Natural Science Foundation of China (31260542, 31260543, 31160449, 31460581, 31560617). YHM was supported by the National Natural Science Foundation of China (31272403). DDW was supported by Qinghai Department of Science and Technology Major Project 'Sanjiangyuan National Park Animal Genome Project'.

Conflict of interest statement. None declared.

REFERENCES

- Storz JF. Causes of molecular convergence and parallelism in protein evolution. *Nat Rev Genet* 2016; **17**: 239–50.
- Tishkoff SA, Reed FA and Ranciaro A *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 2007; **39**: 31–40.
- Sackton TB, Grayson P and Cloutier A *et al.* Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* 2019; **364**: 74–8.
- Storz JF, Scott GR and Cheviron ZA. Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. *J Exp Biol* 2010; **213**: 4125–36.
- Cheviron ZA and Brumfield RT. Genomic insights into adaptation to high-altitude environments. *Heredity* 2012; **108**: 354–61.
- Qiu Q, Zhang G and Ma T *et al.* The yak genome and adaptation to life at high altitude. *Nat Genet* 2012; **44**: 946–9.
- Wu T and Kayser B. High altitude adaptation in Tibetans. *High Alt Med Biol* 2006; **7**: 193–208.
- Storz JF, Runck AM and Sabatino SJ *et al.* Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proc Natl Acad Sci USA* 2009; **106**: 14450–5.
- Storz JF, Sabatino SJ and Hoffmann FG *et al.* The molecular basis of high-altitude adaptation in deer mice. *PLoS Genet* 2007; **3**: e45.
- Tufts DM, Natarajan C and Revsbech IG *et al.* Epistasis constrains mutational pathways of hemoglobin adaptation in high-altitude pikas. *Mol Biol Evol* 2015; **32**: 287–98.
- Natarajan C, Projecto-Garcia J and Moriyama H *et al.* Convergent evolution of hemoglobin function in high-altitude Andean waterfowl involves limited parallelism at the molecular sequence level. *PLoS Genet* 2015; **11**: e1005681.
- Simonson TS, Yang Y and Huff CD *et al.* Genetic evidence for high-altitude adaptation in Tibet. *Science* 2010; **329**: 72–5.
- Yi X, Liang Y and Huerta-Sanchez E *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 2010; **329**: 75–8.
- Ge R-L, Cai Q and Shen Y-Y *et al.* Draft genome sequence of the Tibetan antelope. *Nat Commun* 2013; **4**: 1858.
- Qu Y, Zhao H and Han N *et al.* Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat Commun* 2013; **4**: 2071.
- Gou X, Wang Z and Li N *et al.* Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res* 2014; **24**: 1308–15.
- Li Y, Wu D-D and Boyko AR *et al.* Population variation revealed high-altitude adaptation of Tibetan Mastiffs. *Mol Biol Evol* 2014; **31**: 1200–5.

18. Wang G-D, Fan R-X and Zhai W *et al.* Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the Tibetan plateau. *Genome Biol Evol* 2014; **6**: 2122–8.
19. Lorenzo FR, Huff C and Myllymaki M *et al.* A genetic mechanism for Tibetan high-altitude adaptation. *Nat Genet* 2014; **46**: 951–6.
20. Foll M, Gaggiotti OE and Daub JT *et al.* Widespread signals of convergent adaptation to high altitude in Asia and America. *Am J Hum Genet* 2014; **95**: 394–407.
21. Li M, Tian S and Jin L *et al.* Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet* 2013; **45**: 1431–8.
22. Beall CM, Cavalleri GL and Deng L *et al.* Natural selection on EPAS1 (HIF2 α) associated with low hemoglobin concentration in Tibetan highlanders. *Proc Natl Acad Sci USA* 2010; **107**: 11459–64.
23. Wilson WR and Hay MP. Targeting hypoxia in cancer therapy. *Nat Rev Cancer* 2011; **11**: 393–410.
24. Zhang W, Fan Z and Han E *et al.* Hypoxia adaptations in the grey wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. *PLoS Genet* 2014; **10**: e1004466.
25. Storz JF. Hemoglobin function and physiological adaptation to hypoxia in high-altitude mammals. *J Mammal* 2007; **88**: 24–31.
26. Storz JF and Moriyama H. Mechanisms of hemoglobin adaptation to high altitude hypoxia. *High Alt Med Biol* 2008; **9**: 148–57.
27. Chen FH, Dong GH and Zhang DJ *et al.* Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P. *Science* 2015; **347**: 248–50.
28. Sabeti PC, Varilly P and Fry B *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* 2007; **449**: 913–8.
29. Akey JM, Zhang G and Zhang K *et al.* Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 2002; **12**: 1805–14.
30. Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 1965; **19**: 395–420.
31. Bigham A, Bauchet M and Pinto D *et al.* Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet* 2015; **6**: e1001116.
32. Foote AD, Liu Y and Thomas GWC *et al.* Convergent evolution of the genomes of marine mammals. *Nat Genet* 2015; **47**: 272–5.
33. Zou Z and Zhang J. No genome-wide protein sequence convergence for echolocation. *Mol Biol Evol* 2015; **32**: 1237–41.
34. Grossman SR, Andersen KG and Shlyakhter I *et al.* Identifying recent adaptations in large-scale genomic data. *Cell* 2013; **152**: 703–13.
35. Huerta-Sanchez E, DeGiorgio M and Pagani L *et al.* Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Mol Biol Evol* 2013; **30**: 1877–88.
36. Sabeti PC, Schaffner SF and Fry B *et al.* Positive natural selection in the human lineage. *Science* 2006; **312**: 1614–20.
37. Bigham AW and Lee FS. Human high-altitude adaptation: forward genetics meets the HIF pathway. *Genes Dev* 2014; **28**: 2189–204.
38. Barreiro LB, Laval G and Quach H. Natural selection has driven population differentiation in modern humans. *Nat Genet* 2008; **40**: 340–5.
39. Cutter AD and Payseur BA. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet* 2013; **14**: 262–74.
40. Cruickshank TE and Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 2014; **23**: 3133–57.
41. Hoban S, Kelley JL and Lotterhos KE *et al.* Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am Nat* 2016; **188**: 379–97.
42. Greijer AE and van der Wall E. The role of hypoxia inducible factor 1 (HIF-1) in hypoxia induced apoptosis. *J Clin Pathol* 2004; **57**: 1009–14.
43. Eales KL, Hollinshead KER and Tennant DA. Hypoxia and metabolic adaptation of cancer cells. *Oncogenesis* 2016; **5**: e190.
44. Masson N and Ratcliffe PJ. Hypoxia signaling pathways in cancer metabolism: the importance of co-selecting interconnected physiological pathways. *Cancer Metab* 2014; **2**: 3.
45. LaManna JC. Hypoxia in the central nervous system. *Essays Biochem* 2007; **43**: 139–51.
46. Levy C, Khaled M and Fisher DE. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med* 2006; **12**: 406–14.
47. Han JL, Yang M and Yue YJ *et al.* Analysis of agouti signaling protein (ASIP) gene polymorphisms and association with coat color in Tibetan sheep (*Ovis aries*). *Genet Mol Res* 2005; **14**: 1200–9.
48. Delmar M and McKenna WJ. The Cardiac desmosome and arrhythmogenic cardiomyopathies. *Circ Res* 2010; **107**: 700–14.
49. Amagai M, Nishikawa T and Nousari HC *et al.* Antibodies against desmoglein 3 (pemphigus vulgaris antigen) are present in sera from patients with paraneoplastic pemphigus and cause acantholysis *in vivo* in neonatal mice. *J Clin Invest* 1998; **102**: 775–82.
50. Wang M-S and Li Y Peng MS *et al.* Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Mol Biol Evol* 2015; **32**: 1880–9.
51. Hui AS, Bauer AL and Striet JB *et al.* Calcium signaling stimulates translation of HIF-alpha during hypoxia. *FASEB J* 2006; **20**: 466–75.
52. Shimoda LA and Udem C. Interactions between calcium and reactive oxygen species in pulmonary arterial smooth muscle responses to hypoxia. *Respir Physiol Neurobiol* 2010; **174**: 221–9.
53. Wang Y-X and Zheng Y-M. ROS-dependent signaling mechanisms for hypoxic Ca²⁺ responses in pulmonary artery myocytes. *Antioxid Redox Signal* 2010; **12**: 611–23.
54. Kader A, Li Y and Dong K *et al.* Population variation reveals independent selection toward small body size in Chinese debao pony. *Genome Biol Evol* 2016; **8**: 42–50.
55. Weedon MN, Lettre G and Freathy RM *et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet* 2007; **39**: 1245–50.
56. Weedon MN, Lango H and Lindgren CM *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 2008; **40**: 575–83.
57. Makvandi-Nejad S, Hoffman GE and Allen JJ *et al.* Four loci explain 83% of size variation in the horse. *PLoS One* 2012; **7**: e39929.
58. Yang J, Jin Z-B and Chen J *et al.* Genetic signatures of high-altitude adaptation in Tibetans. *Proc Natl Acad Sci USA* 2017; **114**: 4189–94.
59. Xu S, He Z and Guo Z *et al.* Genome-wide convergence during evolution of mangroves from woody plants. *Mol Biol Evol* 2017; **34**: 1008–15.
60. Dong Y, Xie M and Jiang Y *et al.* Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat Biotechnol* 2013; **31**: 135–41.
61. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013; [arXiv: 1303.3997](https://arxiv.org/abs/1303.3997).
62. McKenna A, Hanna M and Banks E *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–303.

63. Browning SR and Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 2007; **81**: 1084–97.
64. Rossin EJ, Lage K and Raychaudhuri S *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet* 2011; **7**: e1001273.
65. Stark C, Breitkreutz B-J and Reguly T *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 2006; **34**: D535–9.
66. Reimand J, Arak T and Vilo J. g:Profiler—a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res* 2011; **39**: W307–15.
67. Jiang Y, Xie M and Chen W *et al.* The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 2014; **344**: 1168–73.