

SCIENTIFIC REPORTS



OPEN

Genetic adaptations of the plateau zokor in high-elevation burrows

Yong Shao^{1,7}, Jin-Xiu Li², Ri-Li Ge³, Li Zhong², David M. Irwin^{1,4,5}, Robert W. Murphy^{1,6} & Ya-Ping Zhang^{1,2}

Received: 19 August 2015

Accepted: 27 October 2015

Published: 25 November 2015

The plateau zokor (*Myospalax baileyi*) spends its entire life underground in sealed burrows. Confronting limited oxygen and high carbon dioxide concentrations, and complete darkness, they epitomize a successful physiological adaptation. Here, we employ transcriptome sequencing to explore the genetic underpinnings of their adaptations to this unique habitat. Compared to *Rattus norvegicus*, genes belonging to GO categories related to energy metabolism (e.g. mitochondrion and fatty acid beta-oxidation) underwent accelerated evolution in the plateau zokor. Furthermore, the numbers of positively selected genes were significantly enriched in the gene categories involved in ATPase activity, blood vessel development and respiratory gaseous exchange, functional categories that are relevant to adaptation to high altitudes. Among the 787 genes with evidence of parallel evolution, and thus identified as candidate genes, several GO categories (e.g. response to hypoxia, oxygen homeostasis and erythrocyte homeostasis) are significantly enriched, are two genes, *EPAS1* and *AJUBA*, involved in the response to hypoxia, where the parallel evolved sites are at positions that are highly conserved in sequence alignments from multiple species. Thus, accelerated evolution of GO categories, positive selection and parallel evolution at the molecular level provide evidences to parse the genetic adaptations of the plateau zokor for living in high-elevation burrows.

The plateau zokor, *Myospalax baileyi*, is small subterranean rodent that inhabits the Qinghai-Tibet Plateau at elevations ranging from 2000 to 4200 m¹. Zokors spend 85–90% of their lives in underground nests, which are usually greater than 2 m deep for females and approximately 1.5 m deep for males². The depths of the burrows means that the zokor experiences even lower levels of oxygen than those found on the surface of the Qinghai-Tibet Plateau and higher levels of carbon dioxide³.

The plateau zokor exemplifies a successful adaptation to an extreme environmental condition. Compared to pikas (*Ochotona curzoniae*) and Sprague-Dawley rats (*Rattus norvegicus*), the plateau zokor exhibit remarkably increased red blood corpuscle counts and contents of hemoglobin and myoglobin^{3–6}. In contrast, they have markedly lower pulse rates⁶, decreased lactate dehydrogenase activity, and lower 2,3-bisphosphoglycerate-content in their blood compared with pikas, mice and rats that live in the same area³. The oxygen pressure in the arterial blood of the plateau zokor is about 1.5 times higher than that of pikas and rats, and about 0.36 and 0.26 times, respectively, greater in venous blood³. Partial pressure for carbon dioxide in arterial and venous blood of the plateau zokor is 1.5-fold and 2.0-fold higher than in rats and pikas, respectively³. The difference in oxygen saturation of arterial to venous blood was about 2 and 4.5-fold higher in plateau zokor than in pikas and rats, respectively³. However, oxygen saturation

¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming 650223, China. ²Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming 650091, China. ³Key Laboratory for High Altitude Medicine of Ministry of Chinese Education and Research Center for High Altitude Medicine, Qinghai University, Xining 810001, China. ⁴Department of Laboratory Medicine and Pathobiology, University of Toronto, Ontario, M5S 1A8, Canada. ⁵Banting and Best Diabetes Centre, University of Toronto, Ontario, M5S 1A8, Canada. ⁶Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, 100 Queen's Park, Toronto, Ont., M5S 2C6, Canada. ⁷Kunming College of Life Science, University of the Chinese Academy of Sciences, Kunming 650204, China. Correspondence and requests for materials should be addressed to Y.-P.Z. (email: zhangyp@mail.kiz.ac.cn)

Tissue	Raw Reads Number	Clean Reads Number	Contigs Number	Contigs N50 (bp)
Brain	60,294,120	57,366,679	66674	2371
Kidney	44,738,308	43,607,559	51306	2275
Liver	55,713,348	53,250,755	43874	1617
Muscle	67,915,990	62,924,369	99059	884
Retina	43,793,426	38,956,940	90362	812
Mixture	272,455,192	256,106,302	208451	2433

Table 1. The summary of the transcriptome data for a plateau zokor. ^abp = base pair.

of the plateau zokor is nearly 5.7 and 9.3 times lower in venous blood than that of pikas and rats, respectively³. These observations suggest that the plateau zokor has a strong capability to acquire oxygen from their hypoxic-hypercapnic environment.

To elucidate adaptations, and their genetic bases, of animals to extreme environments is a primary mission of modern evolutionary biology⁷. Recently developed bioinformatics help us comparatively analyze massive amounts of data generated by next-generation sequencing technologies at differential scales (e.g. genomics, transcriptomics and proteomics) across diverse species^{8–10}. High altitude environments, especially low temperature, low oxygen density and intense UV radiation, result in harsh physiological challenges for animals^{11,12}. A series of studies suggested that genes involved in the hypoxia-inducing factor (HIF) signaling pathway play key roles in the plateau adaptation of mammals^{13–16}. Here, we examined the plateau zokor, a small endothermic mammal that experiences dual stress from high altitude hypoxia and sealed burrow hypoxia, and thus may become an excellent model for studying adaptations to low oxygen density and hypercapnic concentrations. Revealing the genetic bases of these adaptations in the plateau zokor therefore should be helpful to develop a comprehensive understanding of plateau adaptations of whole endotherms.

The naked mole rat, *Heterocephalus glaber*, is another strictly subterranean rodent¹⁷. Like the plateau zokor, they also live in full darkness, at low oxygen and high carbon dioxide concentrations and possess extremely similar habitats and physiological characteristics¹⁸. Thus, the plateau zokor and the naked mole rat represent a pair of functionally convergent homoplasies (evolutionary analogues) in adapting to the hypoxic environment of sealed burrows¹⁹. Utilization of the draft genome of the naked mole-rat¹⁸ together with the transcriptomes of multiple tissues from the plateau zokor should help us deeply explore genetic adaptations allowing the survival of these rodent species to harsh subterranean environments.

Results

De novo assembly and annotation of transcriptomes. We attained about 272.45 million (25.01 Gb sequence data) raw reads for five tissues (brain, liver, kidney, skeletal muscle and retina) of a plateau zokor. About 256.10 million (19.67 Gb) high-quality reads remained after quality control. The transcriptomes of the five tissues (brain, liver, kidney, skeletal muscle and retina) were pooled to generate an improved *de novo* assembly. The contig N50 parameters of the *de novo* assembly of the pooled data were superior to the single tissue transcriptome *de novo* assemblies, resulting in the identification of 208,451 non-redundant transcripts with a contig N50 length of 2433 base pairs (Table 1). The transcripts were aligned to the NCBI Non-Redundant Protein Database (NR database) using Blastx program (E-value $1e^{-5}$), allowing the annotation of 73,116 transcripts, based on 40,528 NCBI non-redundant proteins, including 34,523 (~85.2 %) that mapped to 19,736 NCBI genes (Supplementary Dataset 1). NCBI gene2go database (<ftp://ftp.ncbi.nih.gov/gene/DATA>) was then utilized to retrieve the GO ids for each of the annotated genes. In summary, 56 GO terms, mainly covering biological GO categories at three ontologies levels (23 GO terms in Biological Process, 15 GO terms in Molecular Function and 18 GO terms in Cellular Component), were identified using WEGO²⁰, a web tool for plotting GO annotations. Our analyses of GO terms generally represented the main biological GO classification and ensured the integrity of the downstream functional analyses of the candidate genes.

Predicted orthologous genes. We identified 8,217 genes that were one-to-one orthologous gene pairs between the plateau zokor (*M. baileyi*), and each of rat (*R. norvegicus*), kangaroo rat (*Dipodomys ordii*), guinea pig (*Cavia porcellus*), naked mole-rat (*H. glaber*) and human (*Homo sapiens*). After removing low quality and short sequences, 8,020 high-confidence one-to-one genes were retained for the downstream analyses. Our final dataset, therefore, included six species and 8020 genes.

Phylogenetic tree and divergence time. By integrating the 8020 single-copy orthologs among the six species (*M. baileyi*, *R. norvegicus*, *D. ordii*, *C. porcellus*, *H. glaber* and *H. sapiens*) and one-to-one orthologous gene pairs from *H. sapiens* and *Oryctolagus cuniculus*, downloaded from ENSEMBL 75 BIOMART database, a total of 7011 high-confidence single-copy orthologous gene pairs were obtained from the seven species (*M. baileyi*, *R. norvegicus*, *D. ordii*, *C. porcellus*, *H. glaber*, *O. cuniculus* and *H. sapiens*).

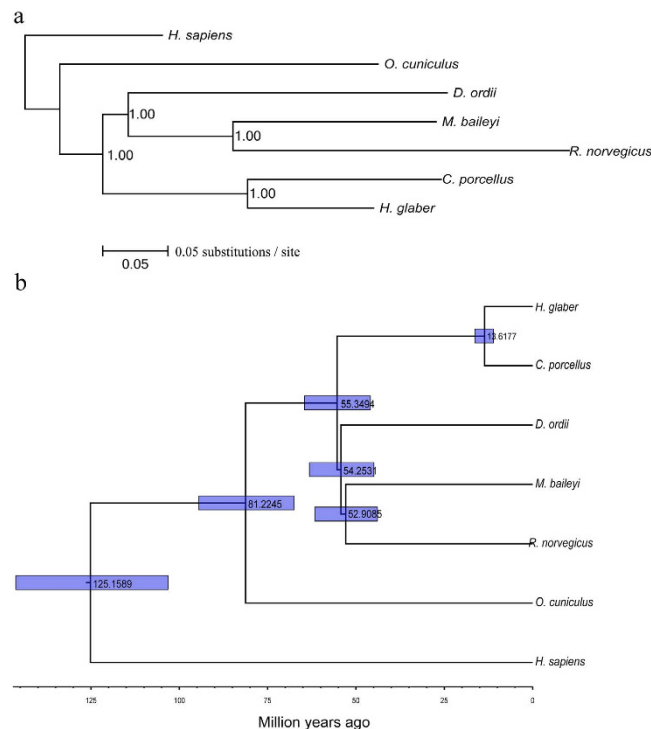


Figure 1. (a) Phylogenetic placement of the plateau zokor derived from an analysis of 4-fold degenerate sites from concatenated orthologous sequence by MrBayes algorithm²¹. Substitutions/site and Bayesian posterior probabilities are shown. (b) The evaluation of the divergence times across seven mammalian species. The Bayesian relaxed molecular clock method was implemented using the BEAST v1.7.5 software²². We utilized an indirect estimate¹⁸ of the divergence time as the calibration point for the split of the ancestor of rabbit from the ancestor of mouse, rat and naked mole-rat. All nodal ages were indicated by medians and 95% HPD intervals (blue bars).

After multiple sequence alignment (MSA) and quality trimming, 7011 one-to-one orthologous genes were concatenated into a single supergene to extract four-fold degenerate sites for construction of a phylogenetic tree using the MrBayes_v3.0 software²¹. High Bayesian posterior probabilities in this constructed phylogenetic tree have well resolved the phylogenetic relationships between the plateau zokor and other rodent species (Fig. 1a). Based on this reliable phylogenetic tree, we estimated the divergence time of each node using BEAST v1.7.5²². As expected, the plateau zokor groups with the rat with an ancestor approximately ~52 million years ago, whereas the ancestor of the plateau zokor, rat and kangaroo rat diverged from the ancestor of the guinea pig and naked mole-rat approximately ~55 million years ago (Fig. 1b). Our estimated divergence time are similar to those of a previous report²³.

Evolutionary rate in plateau zokor. We utilized the free ratio model (M1) in PAML4²⁴ to calculate the selective constraints acting on each of the orthologous genes using our accepted tree topology (Fig. 2a). Because outlier-genes with larger d_N/d_S may generate deviations in evaluating the overall selective constraint in species, our dataset was filtered to remove genes with $d_N/d_S > 4$, yielding a set of 7117 orthologs shared by the plateau zokor and rat lineages. As saturation of substitutions influences the estimates of d_N/d_S , we used previously described methods²⁵ to test for saturation and plotted the ratio of transitions and transversions to d_S ; these relationships were linear up to the d_S of 1, thus orthologues with $d_S > 1$ were excluded from further analyses (Fig. 3). After the saturation test, 7077 orthologous genes were remained, with these highly reliable genes used to evaluate the overall selective constraints in these two related rodent species. The low overall d_N/d_S mean value in the plateau zokor (0.0904) and the rat (0.1360) demonstrates that the majority of genes experienced purifying selection on both lineages. Compared to the rat, the plateau zokor has a significantly lower mean d_N/d_S value at the genomic level (Wilcoxon rank sum test, $P < 2.2E-16$) (Fig. 2b), suggesting that genes evolved a lower rate in the plateau zokor compared to rat since their split from a common ancestor. The mean d_N/d_S value (0.1360) for the rat calculated in this study is very similar to that of a previous study (0.137)²⁶, suggesting that our analysis is reliable.

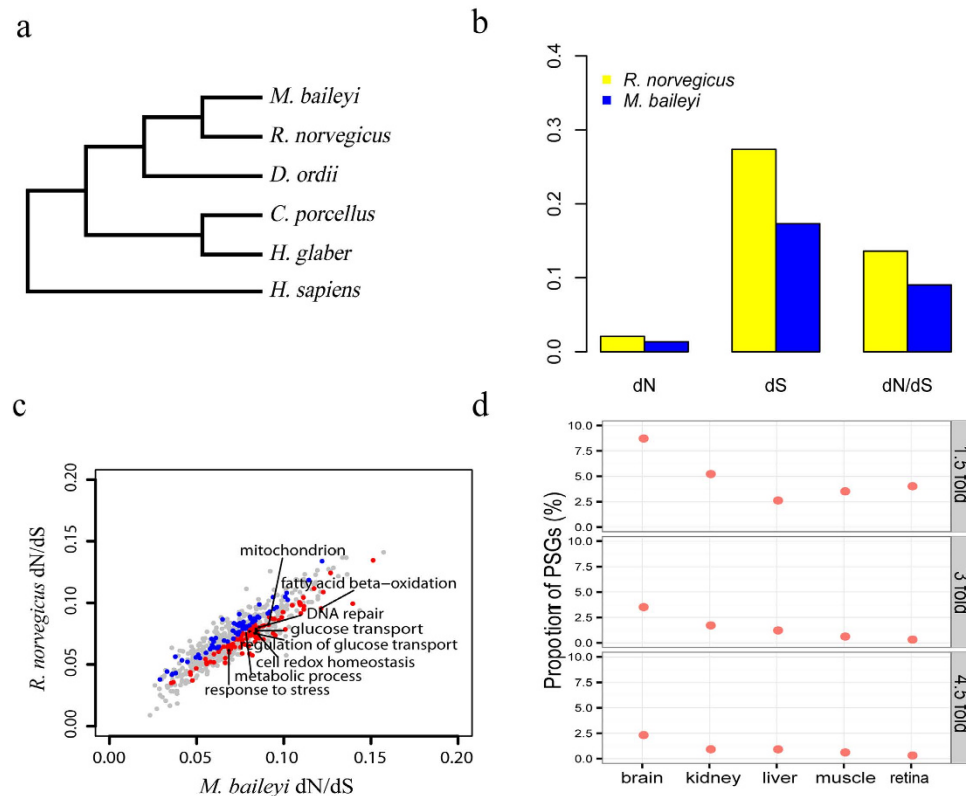


Figure 2. (a) Tree topology of six mammal species. This tree topology originated from our phylogenetic analyses among seven mammal species (*M. baileyi*, *R. norvegicus*, *D. ordii*, *C. porcellus*, *H. glaber*, *O. cuniculus* and *H. sapiens*). (b) Comparison of the evolutionary rates between the plateau zokor and rat lineages. Mean d_N , d_S , and d_N/d_S ratio were originated in 7077 reliable orthologous gene pairs of both lineages are presented. (c) Comparisons of the d_N/d_S ratios between the plateau zokor and rat lineages by GO functional categories. Blue and red dots represented categories with an elevated evolutionary rate along the rat and plateau zokor lineages, respectively. X-axis stood for the mean d_N/d_S ratio of each GO functional category in the plateau zokor lineage and Y-axis stood for the mean d_N/d_S ratio of each GO functional category in the rat lineage. (d) Percentage of tissue-specific highest expressed positively selected genes in each tissue compared to the total number of positively selected genes in the plateau zokor lineage. The 1.5, 3 and 4.5 fold indicate the expression ($\log_2(\text{fpkm}+1)$) of positively selected genes in a single tissue is greater than 1.5, 3 and 4.5 fold of any other tissue, respectively. The y-axis shows the percentage of the number of tissue-specific highest expressed positively selected genes compared to the number of total positively selected genes in the plateau zokor lineage. The PSGs in the figure are the positively selected genes.

Accelerated evolution of genes in specific GO categories. Average d_N/d_S values for genes in each GO category were calculated based on our reconstructed tree topology (Fig. 2a) using the free ratio model (M1) implemented in PAML4²⁴. Among the Go categories, 95 harbor significantly higher average d_N/d_S values in the plateau zokor lineage compared to the rat ($P < 0.05$, binomial test), with these categories including those involved in energy metabolism including glucose transport (GO:0015758, $P = 6.58E-15$), regulation of glucose transport (GO:0010827, $P = 3.09E-11$), mitochondrion (GO:0005739, $P = 1.34E-06$), fatty acid beta-oxidation (GO:0006635, $P = 0.0004$) (Fig. 2c and Supplementary Dataset 2). In contrast, the rat possessed only 65 categories exhibiting significantly higher d_N/d_S values compared to the plateau zokor, which included those for collagen (GO:0005581, $P = 3.22E-58$) and protein polyubiquitination (GO:0000209, $P = 1.50E-06$).

Positive selection. The improved branch-site model in PAML4²⁴ was used to detect positively selected genes (PSGs) among the 8,020 orthologous genes along the plateau zokor branch using our reconstructed tree topology (Fig. 2a). A total of 346 PSGs (4.3%) were identified with signals of positive selection (Supplementary Dataset 3). G:Profiler²⁷ enrichment analyses showed that the candidate PSGs were significantly enriched into categories that included response to stress ($P = 4.28E-16$), respiratory gaseous exchange ($P = 9.31E-05$), blood vessel development ($P = 1.54E-03$) and ATPase activity ($P = 1.32E-04$)

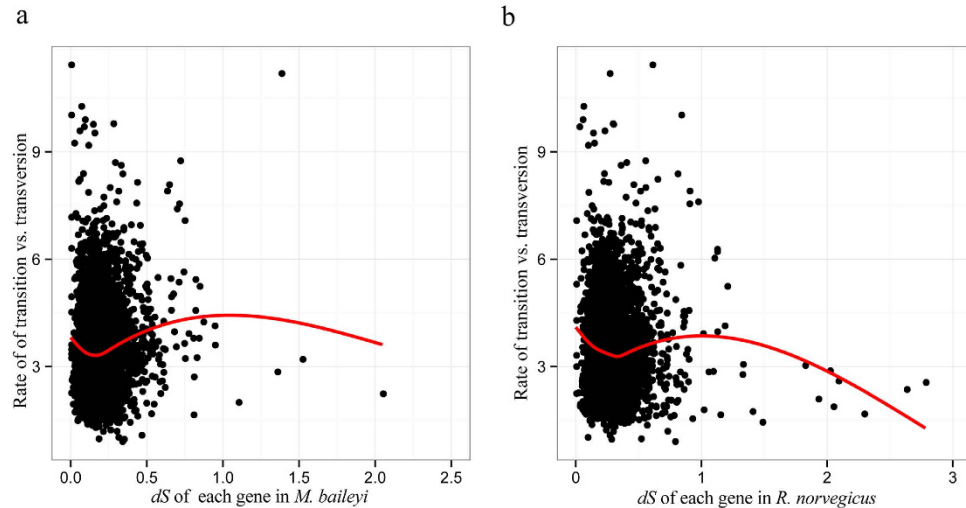


Figure 3. Saturation test of the one-to-one orthologous gene pairs. Each ratio (transition/transversion) of gene was calculated by the free-ratio model of PAML4²⁴. (a) Plot of the ratio (transition/transversion) vs. d_S in the plateau zokor lineage. (b) Plot of the ratio (transition/transversion) vs. d_S in rat lineage.

(Supplementary Dataset 4). These categories appear to be biologically relevant to adaptation to high altitudes. According to previous studies, angiogenesis is largely initiated by the HIF pathway and is a crucial response to hypoxia^{28,29}. Here, 12 PSGs (*ANGPTL3*, *PRCP*, *SEC24B*, *EPHA1*, *MCAM*, *KAT6A*, *MYH9*, *CYSLTR2*, *MYLK*, *SPINT1*, *PPAP2B* and *STRA6*) involved in blood vessel development were identified in the plateau zokor and they may represent adaptive responses to hypoxia. Since ATPase genes have a role in providing energy in conditions of low PO_2 , and show evidence of adaptation in the Tibetan antelope (*Pantholops hodgsonii*)³⁰, another sympatric plateau species, PSGs associated with ATPase activity also may be adaptive to supply energy in the extremely hypoxic sealed burrow environment at high altitudes for the plateau zokor.

In particular, 5 PSGs (*SFTP2A*, *CSF2RB*, *MAPK8IP3*, *PASK* and *MTG2*) involved in respiratory gaseous exchange were over-represented. This category was not found to over-enriched in PSGs in studies of other plateau species, such as Tibetan antelope (*P. hodgsonii*)³⁰, Tibetan boar (*Sus scrofa*)³¹ and yak (*Bos grunniens*)¹⁶, thus these genes may be specific to the plateau zokor and involved in the adaptive evolution of the respiratory system response to hypoxic-hypercapnic environment. In addition, tissue-specific analyses of the expression of PSGs showed the tendency that the brain hosted a higher proportion of PSGs with highest expression than any other tissue (kidney, retina, muscle and liver) (Fig. 2d).

Convergent/parallel evolution. Adaptive evolution at the molecular level also can be studied by detecting convergent/parallel evolution at the amino acid sequence level³². Since both the plateau zokor and naked mole-rat live in sealed burrows and possess extremely similar phenotypic physical characteristics, we detected convergent/parallel evolved genes to reveal possible adaptive mechanisms on these two lineages. In summary, no convergently evolved genes were identified between the plateau zokor and naked mole-rat, however 787 candidate parallel evolved genes, including 700 having a single parallel amino acid change, 72 with two parallel amino acid substitutions and 15 having multiple (>2) parallel changes were identified on the two branches (Supplementary Dataset 5).

Our GO enrichment analyses of these candidate PEGs showed that they were significantly over-represented in GO categories such as response to hypoxia ($P = 3.27E-04$), oxygen homeostasis ($P = 5.85E-03$), erythrocyte homeostasis ($P = 3.25E-02$), angiogenesis ($P = 5.63E-08$), blood vessel development ($P = 1.69E-10$), respiratory tube development ($P = 8.75E-03$) and ATPase activity ($P = 6.79E-07$) (Supplementary Dataset 6). Some of these GO categories are similar to the GO categories identified in the PSG analysis, suggesting that positively selected and parallel evolved genes exist in the same pathways (e.g. blood vessel development, respiratory tube development and ATPase activity) hinting that these genes related to energy metabolism and the development of the respiratory tube may contribute vital roles to adapting to the extremely high-elevation hypoxic burrow environment. However, in contrast, some significant GO categories (e.g. erythrocyte homeostasis, response to hypoxia and oxygen homeostasis) only were specifically enriched in the PEGs list and these specific GO categories possibly directly contribute to the burrow adaptation at high altitude because of their functional importance. Among these categories, 12 genes (*EPAS1*, *SLC29A1*, *MYO5B*, *AJUBA*, *TGFBR3*, *VASN*, *ACAA2*, *PINK1*, *SIRT1*, *SIRT2*, *SOD2* and *NARFL*) are involved in response to hypoxia; 5 genes (*EPAS1*, *DNASE2*, *ADAR*, *TGFBR3* and *SMAP1*) are involved in erythrocyte homeostasis and two genes (*SOD2* and *NARFL*) are involved in oxygen homeostasis. Since the response to hypoxia GO category is generally associated with

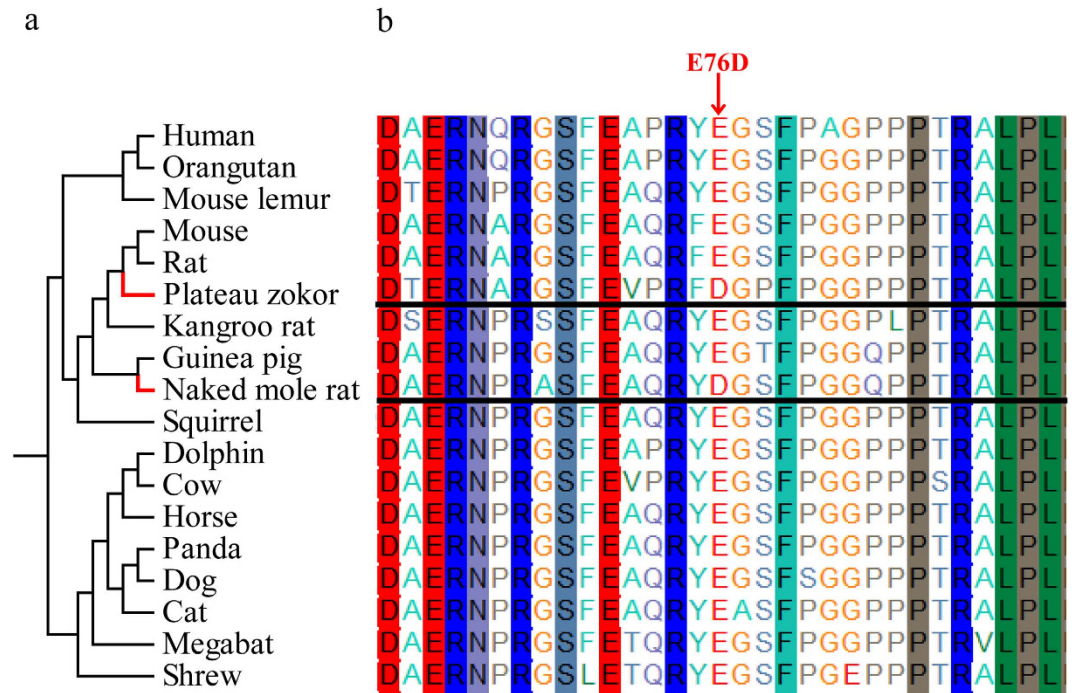


Figure 4. (a) Topology of the species tree. In this species tree, the two parallel evolving branches (plateau zokor and naked mole rat) were highlighted in red. (b) Multiple sequence alignments (MSAs) of the predicted coding sequences of *AJUBA* plateau zokor and other species. The MSAs were generated by PRANK^{55–57} with the aligned coding sequence translated into protein sequence using MEGA5⁶³. In the protein sequences, the parallel evolved site (E76D) in the plateau zokor is indicated by the red arrow.

high-elevation adaptation^{15,33}, this suggests that their parallel evolved sites may not have appeared by random, but instead possess important functions, although the functions of these sites remain unclear. The gene *AJUBA* is an example, where an examination of homologous coding sequences in diverse mammals showed that the parallel evolved site (Glu76Asp) in the plateau zokor (is homologous to 86Glu site in human) occurs at an otherwise extremely conserved site in other species (Fig. 4).

Protein-Protein Interaction (PPI) Network. As mapping of candidate genes to protein-protein interaction (PPI) networks is essential to understand their biological functions, we combined the PSGs and PEGs, resulting in a total of 1,133 candidate genes, and mapped them to the PPI network database (InnateDB)³⁴ to further explore their functional roles. Sub-networks of more than 5 nodes were retained and our analyses resulted in the identification of 935 seed proteins (queries) (82.5%) mapped to sub-networks consisting of 5,954 nodes (proteins) and 15,871 edges (protein-protein interactions) (Fig. 5). Of the 23 genes with the highest degree of interaction (> 100), 14 (*FN1*, *ITGA4*, *SIRT1*, *SHC1*, *ACTB*, *IQCB1*, *XRCC5*, *HTT*, *CASP8*, *XIAP*, *EPS15*, *PSMD1*, *EPAS1* and *SIRT2*) underwent parallel evolution, with 3 of these (*RAF1*, *EEF2* and *MYH9*) also experiencing positive selection. Our results suggest that PSGs and PEGs may play central roles in maintaining these sub-networks.

Discussion

The plateau zokor is a specialized species endemic to the Qinghai-Tibet Plateau³⁵, which may have evolved a series of physiological strategies to counter the effects of hypoxia³ and can be regarded as an alternative evolved pattern to study adaptation to high elevation. Here, we report the transcriptomic data for a plateau zokor with a total of 20,8451 non-redundant transcripts from pooled multiple-tissue data with a Contig N50 attained length of 2,433 base pairs. The length of our sequences allowed the high-confident identification of orthologous genes, which could then be used to perform evolutionary analyses for plateau adaptation. Our analyses meaningfully contribute to the study of the genetic bases of adaptive evolution to high elevation in plateau mammalian species.

After diverged from their ancestor, compared to the rat lineage, overall the plateau zokor displayed a slower mean evolutionary rate in genomic level, although it possesses genes with more rapid evolution in GO categories related to energy metabolism. This result supports the hypothesis that highland adaptation of endothermic species mainly involves several features, including expanded gene families, elevated evolutionary rates and positive selection on genes involved in hypoxia and energy metabolism^{16,30,36}. Significantly accelerated evolution of genes in energy metabolism also support the conclusion

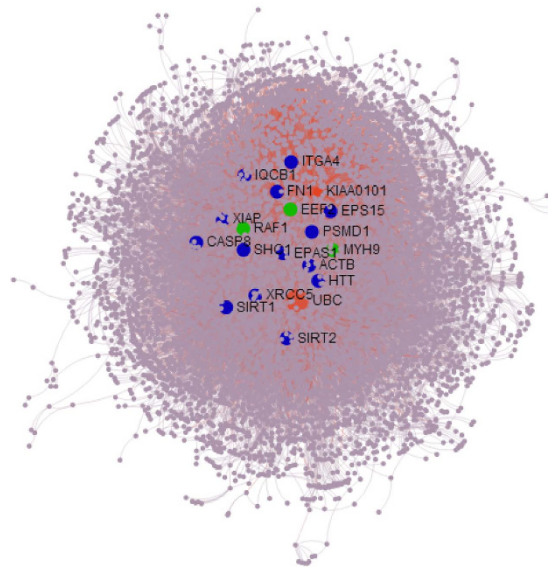


Figure 5. The PPI (Protein-Protein Interaction) network of the candidate genes. Candidate genes with more than 100 interactions ($\text{degree} \geq 100$) are named and indicated by green dots for PSGs and blue dots for PEGs. PSGs are positively selected genes and PEGs are parallel evolved genes.

that the plateau zokor obtains most of its energy by aerobic oxidation instead of anaerobic glycolysis³⁷. Adaptive evolution at the molecular level was well reflected in our screens for positive selection that identified genes involved adaptation³⁸. In the plateau zokor lineage, genes involved in energy metabolism, such as ATPase activity experienced positive selection and thus we suggest that this category of genes may be relevant to adaptation to high altitudes. PSGs associated with blood vessel development were also over-represented and may contribute their function to hypoxic responses as angiogenesis is usually considered to be a target of the HIF pathway^{28,29}. Several genes are of particular interest due to their functional implications. For example, *ANGPTL3* stimulates endothelial cell adhesion and migration via integrin and induces blood vessel formation *in vivo*³⁹. This gene experienced positive selection suggesting an important function in angiogenesis. Another gene, *PRCP* (serine protease prolylcarboxypeptidase), might influence systemic blood pressure and vascular anticoagulation and contributes to cell proliferation and angiogenesis⁴⁰, promoting the health or impair the repairs of blood vessels.

Our study also identified a potentially new GO category for plateau adaptation, respiratory gaseous exchange, as it was significantly enriched in our PSGs list. This category was not found in previous analyses for positive selection in other plateau species (yak (*B. grunniens*), Tibetan boar (*Sus scrofa*), Tibetan antelope (*P. hodgsonii*) and ground tit (*Parus humilis*))^{16,30,31,36}, suggesting that it might be specific to the plateau zokor. For example, the candidate genes *SFTPA2* and *CSF2RB* are involved in surfactant structure and the regulation of surfactant homeostasis in the lung^{41,42}. These genes may explain at molecular level why the plateau zokor possess a large capacity for lung gas diffusion. The number of PSGs with tissue-specific highest expression in brain is higher than in other tissues (e.g. liver, kidney, muscle and retina), suggesting that PSGs prefer to be highly expressed in the brain, since the brain might contribute more than other tissues to the response to hypoxia. We know for humans that the brain has high energy requirements, and consumes nearly 20% of the oxygen and 25% of the glucose of the entire body⁴³. Thus, in the plateau zokor, the brain may similarly consume a high proportion of the energy and oxygen.

Adaptive evolution at the molecular level can also be studied by detecting convergent/parallel evolution at the amino acid sequence level³². Species living in similar ecological environments could be shaped by convergent evolution to form physiological or morphological similarities⁴⁴. The naked mole-rat also lives in a strictly subterranean habitat that involves full darkness and low oxygen and high carbon dioxide concentrations^{18,45,46}. Our GO functional analyses showed parallel evolved genes were significantly enriched in the following GO categories: response to hypoxia, ATPase activity and angiogenesis. These GO categories illustrate that in order to survive their shared harsh environment, the plateau zokor and naked mole-rat may experience convergent functional improvements in the supply of optimum oxygen levels by influencing the function of genes related to the hypoxia response, energy mechanism and angiogenesis. For example, the gene *EPAS1* encodes HIF-2 α , which primarily regulates the production of erythropoietin (EPO). In turn, EPO is a key hormone that stimulates and regulates the production of erythrocytes⁴⁷. *EPAS1* is widely reported as a candidate adapted gene for living at high elevations^{13,15,48}. Additionally, analysis of Protein-Protein interaction (PPI) networks also demonstrates that *EPAS1* possesses strong interactions with other proteins. Taken together, *EPAS1* in the plateau zokor and naked mole-rat lineages exhibited parallel changes of amino acid sequences that are likely not due to random

chance, but are an adaptive mechanism in response to decreased oxygen in their sealed environments. Another gene, *AJUBA*, is also involved in response to hypoxia and is a key regulator of the hypoxic response regulating the cellular and physiological changes to oxygen levels by controlling the degradation, and transcriptional activity of hypoxia-inducible transcriptional activity, of hypoxia-inducible transcription factors (HIFs)⁴⁹. In particular, this gene hosts a highly conserved parallel evolved site (E76D) (identified in comparison of homologous sequences among multiple species) that suggests that it is also associated with highland adaptation, although the function of this particular amino acid substitution is unknown. Future functional experiments will be necessary to validate its importance. Thus, the plateau zokor exhibits multiple strategies to adapt to harsh highland environments. Although we did not directly examine the role of differentially expressed genes, due to the limitations of our sampling strategy, changes in gene expression likely also play important roles in the phenotypic evolution of this species⁵⁰.

Materials and Methods

Methods were carried out in accordance with approved guidelines.

Sample collection and library preparation. The care and treatment of the plateau zokor comply with the guidelines for the National Care and Use of Animals approved by the National Animal Research Authority (P. R. China). Experimental protocols involving live animals were approved by the Ethics and Experimental Animal Committee of Kunming Institute of Zoology, Chinese Academy of Science, China (Approval ID: SYDW-2015012).

A plateau zokor individual was collected in Qinghai Province and after euthanasia, the brain, liver, kidney, skeletal muscle and retina tissue were quickly biopsied, placed in liquid nitrogen, and stored at -80°C upon return to the laboratory. Total RNA from these tissues was extracted using TRIzol reagent (Invitrogen Corp., Carlsbad, CA). RNA purifications were performed using an RNeasy Mini Kit (Qiagen, Chatsworth, CA). Library constructions from the brain, liver and kidney of a plateau zokor (no replicates) followed the Illumina Genome Analyzer II RNA sample preparation kit (GA IIX, Illumina, Inc.) and Library preparation for skeletal muscle and retina (no replicates) were according to the Illumina HiSeq2000 RNA sample preparation kit (Illumina, San Diego, CA). All original data were deposited in the NCBI Sequence Read Archive database (Accession Number: SRP057676; Alias: PRJNA282349).

De novo assembly and annotation of transcriptomes in plateau zokor. Sequencing adaptors used for cDNA library construction were trimmed using Cutadapt⁵¹ (version_1.2.1), which removed adapter sequences from the high-throughput sequencing reads. We employed Btrim64⁵² (version_0.1.0) to delete regions with average quality scores of less than 20 and impose a minimal length equal or greater than 20 bp. High quality pooled paired end reads from multiple tissues were *de novo* assembled using Trinity⁵³ (version_2013_08_14) by default parameters. Transcripts of pooled transcriptome were aligned to the NCBI Non-redundant protein database (NR) using BLASTX program (E-value 1E^{-5}) to produce annotation results. NCBI gene2accession and gene2go databases (<ftp://ftp.ncbi.nih.gov/gene/DATA>) were then used, respectively, to retrieve gene ID and GO ID for each annotated gene. WEGO software²⁰ (<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>) was used to perform functional annotation analyses at three gene ontology levels (Biological Process, Molecular Function and Cellular Component).

Prediction of open reading frames. Coding sequences of the plateau zokor were predicted based on the assumption that the longest open reading frame in the longest transcript per gene had a greatest chance of being a protein-coding region. GETORF in EMBOSS⁵⁴ (version_6.4.0) was applied to obtain the nucleic sequences between the start and stop codons and limit the minimum size of a fragment to 120 bp. The transeq program in EMBOSS⁵⁴ (version_6.4.0) was then used to obtain translated protein sequences.

Identification of one-to-one orthologous genes. We used the longest protein sequences per gene to perform a best reciprocal hit (BRH) (E-value 1E^{-5}) methods to identify one-to-one orthologous genes among six species: plateau zokor (*M. baileyi*), rat (*R. norvegicus*), kangaroo rat (*D. ordii*), guinea pig (*C. porcellus*), naked mole-rat (*H. glaber*), and human (*H. sapiens*). Protein sequences of the naked mole rat were downloaded from the naked mole rat database (<http://mr.genomics.org.cn/page/species/index.jsp>). Genomic protein sequences of the rat, kangaroo rat, guinea pig and human were downloaded from ENSEMBL 75 (<http://www.ensembl.org/>). Predicted orthologous genes were submitted to multiple alignments using PRANK^{55–57} (Parameters: $-f = \text{fasta}$ $-F = \text{codon}$ $-\text{noxml}$ $-\text{notree}$ $-\text{nopost}$). We aligned the longest ORFs for the longest transcript-pairs across the six species. Gblocks^{58,59} (version_0.91b; Parameters: $-\text{t} = \text{c}$ $-\text{b}3 = 1$ $-\text{b}4 = 6$ $-\text{b}5 = \text{n}$) was employed to reduce the rate of false positive predictions by filtering out sequencing errors, incorrect alignments and no-orthologous regions based on codons. After trimming, we removed alignment lengths shorter than 10 bp to obtain one-to-one orthologous genes.

Phylogenetic tree and divergence time. To establish phylogenetic relationship of the plateau zokor to other rodent mammals, one-to-one orthologous gene pairs between the human (*H. sapiens*) and rabbit (*O. cuniculus*) were downloaded from ENSEMBL 75 BIOMART database and added to dataset of

one-to-one orthologous described above, generating one-to-one orthologous gene pairs for seven species. After multiple sequence alignments and trimming by the programs PRANK^{55–57} and Gblocks^{58,59}, respectively, these single-copy orthologous genes were concatenated into one supergene and 4-fold degenerate sites were extracted and used to construct a phylogenetic tree. Modeltest_0.1⁶⁰ was used to select the best substitute model and Mrbayes_v3.0²¹ was utilized to reconstruct the phylogenetic tree. Chain length was set to 10,000,000, with the first 1,000 samples treated as burned in, and the other parameters were set as defaults. Nodal ages within rodent mammals were evaluated, along with their 95% confidence intervals (CIs), from our 4-fold degenerate site data from the supergene sequence using the Bayesian relaxed molecular clock method implemented in BEAST v1.7.5²². Chain length was set to 10,000,000, with the first 1,000 samples treated as burned in. We utilized an indirect estimate¹⁸ of the divergence time as a calibration point with the split of the ancestor of rabbit from the ancestor of mouse, rat and naked mole-rat to estimate the internal nodal ages.

Evolutionary rate and accelerated evolution of GO categories. Free ratio model (M1) (Parameters: model = 1, NSSites = 0, fix_omega = 0, omega = 1) in PAML4²⁴ was used to calculate selective constraint of each orthologous gene using our constructed tree topology. Genes with $d_N/d_S > 4$ were first filtered and then a saturation test was performed to produce a final dataset to evaluate the overall selective constraints in the tested species. GO term information was downloaded from ENSEMBL (<http://www.ensembl.org>), and only those GO categories with more than 20 orthologs were included in our analyses. Lineage-specific mean d_N/d_S values were estimated by concatenating alignments from all orthologs for each GO category. Relatively accelerated evolutionary GO terms were identified using a binomial test according to a previous study²⁶.

Positive selection analyses. Branch-site model (Parameters: Null hypothesis: model = 2, NSSites = 2, fix_omega = 1, omega = 1; Alternative hypothesis: model = 2, NSSites = 2, fix_omega = 0, omega = 1) in PAML4²⁴ was used to detect positive selection in the one-to-one orthologous genes using our tree topology as the guide tree. Compared to other estimation models, the Branch-site model has the advantage of detecting positive selection that affects only a few sites on a pre-specified (foreground) branch of the species tree. The likelihood ratio test (LRT) was used to detect positive selection on the foreground branch. PSGs were inferred only if their *P* values were less than 0.05. After identifying positively selected genes, the Bayes empirical Bayes (BEB) method was implemented to calculate posterior probabilities and to record positively selected sites. *P*-values of all PSGs also were normalized by FDR using Benjamini & Hochberg approach⁶¹. G:Profiler software²⁷ was utilized to perform functional enrichment analyses for the PSGs by using ‘all known genes’ of human as the statistical background domain size. Additionally, we deduced that the variable trend of tissue-specific highest expressed PSGs in plateau lineage. Here, our analyses controlled the strict gradient thresholds with folds in expression level (>1.5fold, >3fold and >4.5fold) to ensure this trend more accurate. In brief, if $\log_2(\text{fpkm} + 1)$ of the PSGs in one tissue respectively was greater 1.5, 3 and 4 fold of the expressions of all other four tissues, then the PSG was regarded to be the tissue-specific highest expressed gene.

Convergent/parallel evolution. Convergent/Parallel evolved changes were identified using the PAML4²⁴ package to reconstruct the most likely ancestral states, and then a PERL program was used to calculate the number of parallel amino acid replacements for a specified pair of branches. If the posterior probability of the reconstructed ancestral amino acid site was more than 90%, then they were retained and their states were deemed as reliable. In addition, G:Profiler software²⁷ was used to perform functional enrichment analyses for these PEGs utilizing ‘all known genes’ of human as the statistical background domain size.

Protein-Protein Interaction (PPI) Network. We mapped the combined PSGs and PEGs candidate gene lists to the protein-protein interaction (PPI) network database (InnateDB)³⁴. Sub-network topologies were obtained by NETWORKANALYST software⁶² (<http://www.networkanalyst.ca/NetworkAnalyst>). Sub-networks with node counts ≤ 5 were filtered and the statistics of total nodes, edges and seed proteins were obtained from the mapping overview.

References

1. Fan, N. C. & Shi, Y. Z. A revision of the zokors of subgenus *Eospalax*. *Acta Theriol Sinica*. **2**, 180–199 (1982).
2. Zhou, W. Y. & Dou, F. M. Studies on activity and home range of plateau zokor. *Acta Theriol Sinica*. **10**, 31–39 (1990).
3. Wei, D. B., Wei, L., Zhang, J. M. & Yu, H. Y. Blood-gas properties of plateau zokor (*Myospalax baileyi*). *Comp Biochem Physiol A Mol Integr Physiol*. **145**, 372–375 (2006).
4. Wei, D. B. & Ma, J. B. Comparison of the content of myoglobin and lactate dehydrogenase in cardiac and skeleton muscle of plateau zorkor and mouse. *J. Qinghai Univ.* **19**, 20–21 (2001).
5. Wei, D. B. & Wei, L. The mensuration results of the number of red cell and the content of hemoglobin and myoglobin in plateau zokor. *J. Qinghai Univ.* **19**, 1–2 (2001).
6. Zeng, J., Wang, Z. & Shi, Z. Metabolic characteristics and some physiological parameters of mole rat (*Myospalax baileyi*) in alpine area. *Acta Biol Plat Sin.* **3**, 163–171 (1984).
7. Smith, N. G. C. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature*. **415**, 1022–1024 (2002).
8. Li, R. *et al.* The sequence and *de novo* assembly of the giant panda genome. *Nature*. **463**, 311–317 (2010).
9. Gerstein, M. B. *et al.* Comparative analysis of the transcriptome across distant species. *Nature*. **512**, 445–448 (2014).

10. Looso, M. *et al.* A *de novo* assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissue regeneration. *Genome Biol.* **14**, R16 (2013).
11. Scheinfeldt, L. B. & Tishkoff, S. A. Living the high life: high-altitude adaptation. *Genome Biol.* **11**, 133 (2010).
12. Cheviron, Z. A. & Brumfield, R. T. Genomic insights into adaptation to high-altitude environments. *Heredity.* **108**, 354–361 (2012).
13. Beall, C. M. *et al.* Natural selection on *EPAS1* (HIF2 α) associated with low hemoglobin concentration in Tibetan highlanders. *Proc Natl Acad Sci USA* **107**, 11459–11464 (2010).
14. Yi, X. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science.* **329**, 75–78 (2010).
15. Simonson, T. S. *et al.* Genetic evidence for high-altitude adaptation in Tibet. *Science.* **329**, 72–75 (2010).
16. Qiu, Q. *et al.* The yak genome and adaptation to life at high altitude. *Nat Genet.* **44**, 946–949 (2012).
17. Fang, X. *et al.* Adaptations to a subterranean environment and longevity revealed by the analysis of mole rat genomes. *Cell reports.* **8**, 1354–1364 (2014).
18. Kim, E. B. *et al.* Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature.* **479**, 223–227 (2011).
19. Gordon, M. S. & Notar, J. C. Can systems biology help to separate evolutionary analogies (convergent homoplasies) from homologies? *Prog Biophys Mol Biol.* **117**, 19–29 (2015).
20. Ye, J. *et al.* WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, W293–297 (2006).
21. Huelsenbeck, J. P. & Ronquist, F. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics.* **17**, 754–755 (2001).
22. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
23. Montgelard, C., Forty, E., Arnal, V. & Matthee, C. A. Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. *BMC Evol Biol.* **8**, 321 (2008).
24. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* **24**, 1586–1591 (2007).
25. Axelsson, E. *et al.* Natural selection in avian protein-coding genes expressed in brain. *Mol Ecol.* **17**, 3008–3017 (2008).
26. Consortium, C. S. A. A. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature.* **437**, 69–87 (2005).
27. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–200 (2007).
28. Rey, S. & Semenza, G. L. Hypoxia-inducible factor-1-dependent mechanisms of vascularization and vascular remodelling. *Cardiovasc Res* **86**, 236–242 (2010).
29. Li, Y. *et al.* Population variation revealed high-altitude adaptation of Tibetan mastiffs. *Mol Biol Evol.* **31**, 1200–1205 (2014).
30. Ge, R. L. *et al.* Draft genome sequence of the Tibetan antelope. *Nat Commun.* **4**, 1858 (2013).
31. Li, M. *et al.* Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet.* **45**, 1431–1438 (2013).
32. Zhang, J. & Kumar, S. Detection of convergent and parallel evolution at the amino acid sequence level. *Mol Biol Evol.* **14**, 527–536 (1997).
33. Yang, W., Qi, Y., Bi, K. & Fu, J. Toward understanding the genetic basis of adaptation to high-elevation life in poikilothermic species: a comparative transcriptomic analysis of two ranid frogs, *Rana chensinensis* and *R. kukunoris*. *BMC Genomics.* **13**, 588 (2012).
34. Breuer, K. *et al.* InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res.* **41**, D1228–1233 (2013).
35. Ma, B. Y., Wei, L., Sun, S. Z., Wang, D. W. & Wei, D. B. The plateau zokors' learning and memory ability is related to the high expression levels of foxP2 in the brain. *Sheng Li Xue Bao.* **66**, 135–144 (2014).
36. Qu, Y. *et al.* Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat Commun.* **4**, 2071 (2013).
37. Sun, S. Z., Wei, L., Wei, D. B., Wang, D. W. & Ma, B. Y. Differences of glycolysis in skeletal muscle and lactate metabolism in liver between plateau zokor (*Myospalax baileyi*) and plateau pika (*Ochotona curzoniae*). *Sheng Li Xue Bao.* **65**, 276–284 (2013).
38. Bakewell, M. A., Shi, P. & Zhang, J. More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc Natl Acad Sci USA* **104**, 7489–7494 (2007).
39. Camenisch, G. *et al.* *ANGPTL3* stimulates endothelial cell adhesion and migration via integrin α v β 3 and induces blood vessel formation *in vivo*. *J Biol Chem.* **277**, 17281–17290 (2002).
40. Adams, G. N. *et al.* Prolylcarboxypeptidase promotes angiogenesis and vascular repair. *Blood.* **122**, 1522–1531 (2013).
41. Silveyra, P., Wang, G. & Floros, J. Human SP-A1 (SFTPA1) variant-specific 3' UTRs and poly(A) tail differentially affect the *in vitro* translation of a reporter gene. *Am J Physiol Lung Cell Mol Physiol.* **299**, L523–534 (2010).
42. Suzuki, T. *et al.* Hereditary pulmonary alveolar proteinosis caused by recessive CSF2RB mutations. *Eur Respir J* **37**, 201–204 (2011).
43. Belanger, M., Allaman, I. & Magistretti, P. J. Brain energy metabolism: focus on astrocyte-neuron metabolic cooperation. *Cell Metab.* **14**, 724–738 (2011).
44. Stern, D. L. The genetic causes of convergent evolution. *Nat Rev Genet.* **14**, 751–764 (2013).
45. Edrey, Y. H., Park, T. J., Kang, H., Biney, A. & Buffenstein, R. Endocrine function and neurobiology of the longest-living rodent, the naked mole-rat. *Exp Gerontol.* **46**, 116–123 (2011).
46. Larson, J. & Park, T. J. Extreme hypoxia tolerance of naked mole-rat brain. *Neuroreport.* **20**, 1634–1637 (2009).
47. Lorenzo, F. R. *et al.* A genetic mechanism for Tibetan high-altitude adaptation. *Nat Genet.* **46**, 951–956 (2014).
48. Wang, G. D. *et al.* Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the tibetan plateau. *Genome Biol Evol.* **6**, 2122–2128 (2014).
49. Bridge, K. S. & Sharp, T. V. Regulators of the hypoxic response: a growing family. *Future Oncol.* **8**, 491–493 (2012).
50. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet.* **8**, 206–216 (2007).
51. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *J EMBnet.* **17**, 10–12 (2011).
52. Kong, Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics.* **98**, 152–153 (2011).
53. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* **29**, 644–652 (2011).
54. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
55. Loytynoja, A. & Goldman, N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics.* **11**, 579 (2010).
56. Loytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci USA* **102**, 10557–10562 (2005).

57. Loytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*. **320**, 1632–1635 (2008).
58. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* **17**, 540–552 (2000).
59. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* **56**, 564–577 (2007).
60. Posada, D. & Crandall, K. A. Modeltest: testing the model of DNA substitution. *Bioinformatics.* **14**, 817–818 (1998).
61. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Roy Stat Soc, Ser B.* **57**, 289–300 (1995).
62. Xia, J., Benner, M. J. & Hancock, R. E. NetworkAnalyst—integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res.* **42**, W167–174 (2014).
63. Tamura, K. *et al.* Mega5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* **28**, 2731–2739 (2011).

Acknowledgements

We thank Pro. Yong-Yi Shen for his help in data analysis. This work was supported by grants from the National Natural Science Foundation of China, and the Animal Branch of the Germplasm Bank of Wild Species of Chinese Academy of Sciences (the Large Research Infrastructure Funding).

Author Contributions

Y.P.Z. designed the experiments. Y.S. analyzed the data and wrote the paper. J.X.L. and L.Z. performed the separation, purification and sequencing of RNA. R.L.G. provided the data samples. D.M.I. and R.W.M. revised this paper. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Shao, Y. *et al.* Genetic adaptations of the plateau zokor in high-elevation burrows. *Sci. Rep.* **5**, 17262; doi: 10.1038/srep17262 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>