



Predicted structural change in erythropoietin of plateau zokors — Adaptation to high altitude

Zhenlong Wang^a, Yanming Zhang^{b,*}

^a Department of Bioengineering, Zhengzhou University, Zhengzhou, 450001, Henan, PR China

^b Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Plateau Institute of Biology, Chinese Academy of Sciences, Xining 810001, PR China

ARTICLE INFO

Article history:

Accepted 21 February 2012

Available online 8 March 2012

Keywords:

Qinghai–Tibet Plateau
Plateau zokor
Erythropoietin
cDNA cloning

ABSTRACT

Erythropoietin (EPO) is a glycoprotein hormone, expressed mainly in fetus liver and adult kidneys. EPO plays an important role in enhancing red blood cell formation in bone marrow under hypoxia. Plateau zokor (*Myospalax baileyi*), an subterranean burrowing endemic rodent inhabiting areas of 2 800–4 200 m above sea level on Qinghai–Tibet Plateau, is a typical high hypoxia tolerant mammal with high ratio of oxygen utilization in adaptation to the harsh plateau environment. To investigate the possible mechanisms of adaptation of plateau zokor EPO to high altitude, the complete cDNA and amino acid sequences of plateau zokor EPO have been described. Phylogenetic tree of Epo showed the convergence of the *Spalax* and *Myospalax*, indicating that, the convergent evolution was driven by similar hypoxic ecological niches. Our results showed that some common sites under positive selection in zokor (116M and 144A) and *Spalax* (102R, 116M, 144A and 152P) are the important sites for Epo biological activity. This study thus reports a gene level observation which may be involved in adaptation to underground life at high altitude.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Erythropoietin (EPO) is a glycoprotein hormone, expressed mainly in fetus liver and adult kidneys (Jacobson et al., 1957; Zanjani et al., 1977). The EPO gene is highly conserved among mammals (Shoemaker and Mitscock, 1986). EPO is produced by DNA dependent mRNA synthesis (Erslev, 1974; Goldberg et al., 1988; Schooley and Mahlmann, 1972) and is controlled by an oxygen detection system that responds to changes in venous rather than arterial PO₂ (Ebert and Bunn, 1999; Kurtz et al., 1988).

As a subterranean rodent endemic to the Qinghai–Tibet Plateau, plateau zokor (*Myospalax baileyi*) is tolerant of hypoxia and has a markedly high oxygen utilization ratio to cope with the plateau environment, making them a good model for research into adaptation to hypoxia (Zhang and Liu, 2003). Zokors spend their entire life underground (Norris et al., 2004; Wang et al., 1979; Zhang et al., 2003), exposed to fluctuating O₂ and CO₂ levels (Wei et al., 2006). Zokors have evolved physiological strategies underlying their respiratory and cardiovascular systems to cope with hypoxia (Qi et al., 2008; Wang et al., 2008; Wei et al., 2006). Compared to *Rattus*, plateau zokor has significantly higher microvessel density of cardiac muscle (Qi et al., 2008), myocardial performance (Qi et al., 2008) and capillary and mitochondrial density (Qi

et al., 2008). It has a higher erythrocyte count, increased lung diffusion capacity and different structure of hemoglobin and myoglobin (Wei et al., 2006). These characteristics indicate that they have adapted to high altitude at the physiological level. However, there is little genetic information on the properties of EPO in plateau zokors. In this study, the plateau zokor EPO cDNA was cloned and sequenced, and the predicted amino acid sequences were compared with those of other vertebrates in order to investigate the genetic basis of adaptation to high altitude.

2. Materials and methods

2.1. Study sites and animal sampling

The zokors were captured from 4 localities across southeast of Qinghai–Tibet Plateau during April to June of 2006 (Table 1). Sampling was performed between 9:00 am and 4:00 pm. All animals were live-trapped. The captured zokors were anesthetized with chloral hydrate (5%) first before killed by cervical dislocation. The kidneys were removed and immediately preserved in liquid nitrogen until later procedure. All procedures involved in animal handling were in accordance with the China Practice for the Care and Use of Laboratory animals and were approved by China Zoological Society.

2.2. Preparation of total RNA and cDNA synthesis

Total RNA was extracted from a 100-mg portion of frozen kidney tissue using TRIzol® Reagent (Invitrogen Corp, USA), diluted in 100 µl

Abbreviations: EPO, Erythropoietin; ORF, open reading frame; CDS, coding sequence; CK-2, casein kinase II phosphorylation site; CAS, Cell attachment sequence; SNP, single nucleotide polymorphism.

* Corresponding author.

E-mail address: zhangym@nwpib.ac.cn (Y. Zhang).

Table 1

Geographic and climatological data for 5 samples of *M. baileyi* in four sampling locations of Qinghai–Tibet Plateau, China.

Populations				Ecogeographical variables			
No.	Location	Latitude	Longitude	N	Altitude (m)	Tm (°C)	Rn (mm)
1	Deang, Dari	33°23.158'	100°12.578'	2	4256	−1.3	552.4
2	Manzhang, Dari 1	30°16.449'	100°27.235'	1	3963	−1.3	552.4
3	Manzhang, Dari 2	30°16.796'	100°27.056'	1	3977	−1.3	552.4
4	Aba, Sichuan	33°27.179'	101°55.255'	1	3474	3.3	712.0

Note: Symbols of variables are as follows: Tm = mean annual temperature (°C); Rn = mean annual rainfall (in mm). Climatic data were obtained from local weather bureau

RNase-free water and analyzed by agarose gel electrophoresis, only sharp bands of 18S and 28S rRNA indicating high quality preparations were used in later analysis. RNA concentration was determined by spectrophotometry using a NanoDrop® ND-1000 Spectrophotometer. RNA samples were treated with RNase-free DNase I (TaKaRa, Dalian, China), then stored at −80 °C.

Five micrograms of total RNA was taken for first-strand cDNA synthesis using a SuperScript™ II RT Kit (Invitrogen Corp) in a 20- μ l volume containing 5 \times First-Strand Buffer, Oligo(dT)18 (20 μ g/ml), 10 Mm dNTPs, 0.1 M DTT, RNaseOUT (40 units/ μ l) and Reverse Transcriptase (40 units/ μ l). The reaction was carried out at 65 °C for 5 min, 42 °C for 2 min. The reverse transcriptase was incubated at 42 °C for 50 min and inactivated at 70 °C for 15 min. Aliquots of undiluted cDNA were used for PCR and real-time PCR.

2.3. Cloning of plateau zokor erythropoietin gene

An open reading frame (ORF) is a portion of gene sequence that can potentially encode a protein. To obtain complete ORFs of plateau zokor erythropoietin, the forward (EpoF 20–39: GAG ATG GGG GTG CCC GAA CG, Accession No. NM_017001) and reverse (EpoR 583–602: GTC ACC TGT CCC CTC TCC TG) primers were designed according to the alignment of highly conserved coding sequence regions of the *Epo* gene in humans (Accession No. NM_000799), mice (Accession No. NM_007942), rats (Accession No. NM_017001) and *Spalax* sp. (Accession No. AJ715792, AJ715793, AJ715794, AJ715795) with Primer Premier (version 5.0, Premier Biosoft International, Palo Alto, USA). The product of the *Epo* gene transcript was amplified by PCR using *LA Taq* DNA polymerase (TaKaRa, Dalian, China) in a mixture of 200 μ M dNTP, 0.3 μ M primers (Beijing Genomics Institute, Beijing, China) and 2 \times GC Buffer I (Mg²⁺ plus) with 5 μ g undiluted cDNA template. The PCR reactions were started with 10 min at 95 °C and followed by 35 cycles consisting of 45 s at 94 °C, 30 s at 55 °C, and 60 s at 72 °C. The target PCR product of expected size was purified using a Axyprep DNA Gel Extraction Kit (Axygen Biosciences, California, USA) and cloned into pMD19-T vector (TaKaRa, Dalian, China). PCR products and clones were then sequenced. The full-length *Epo* coding sequence (CDS) of plateau zokor was determined according to the alignment regions with the above species. The entire sequence was submitted to the GenBank database (Accession No. 1021725).

2.4. Sequence analysis

The nucleotide and deduced amino acid sequences were compared with the sequences in the GenBank database using the BLAST program (<http://www.ncbi.nlm.nih.gov>). The signal peptide was predicted using the SignalP tool (<http://www.cbs.dtu.dk/services/SignalP>). Multiple alignments were done using the program CLUSTALX 1.81 (Thompson et al., 1997). The functional amino acid motifs were predicted using the MotifScan program in the PROSITE database of protein families and domains (<http://www.expasy.org/prosite>). The

secondary sequence structure was predicted using the consensus methods of Sspro, Sspro8 (Pollastri and Baldi, 2002), ACCpro, CONpro (Baldi and Pollastri, 2003), CMAPpro, and CCMAPprothe (Pollastri and Baldi, 2002) on the SCRATCH server (<http://www.igb.uci.edu/tools/scratch/>).

Tertiary structures were modeled using both automated and alignment modes of homology modeling provided by the SWISS-Adaptive MOELD Server (<http://swissmodel.expasy.org>) with the reference template of Homo sapiens Epo (PDB ID code: 1BUY) (Cheetham et al., 1998). For visualization and manipulation of the 3D molecule, we used the spdbv 3.7 tool (<http://swissmodel.expasy.org/spdbv/>) (Guex and Peitsch, 1997).

2.5. Evolutionary analysis

Phylogenetic trees were constructed using three different tree-making algorithms, neighbor-joining (NJ), maximum likelihood (ML), and maximum parsimony (MP), in version 3.66 of the PHYLIP software package using both nucleotide and amino acid sequences, respectively (Felsenstein, 2006). The stability among the clades of the phylogenetic tree was assessed by taking 100000 replicates of the dataset and performing analyses using the following programs: SEQBOOT, DNADIST, FITCH, DNAML, DNAPARS, PRODIST, PROTPARS, PROML, and CONSENSE from the PHYLIP software package. Human *Epo* was used as outgroup for all trees. Relative rate tests were performed using the program RRTree version 1.1 (<http://pbil.univ-lyon1.fr/software/rrtree.html>) (Robinson and Huchon, 2000). The ModelTest 3.7 (Posada and Crandall, 1998) and PAUP* 4.0b10 (Swofford, 2000) software were used to determine the best-fit model of molecular evolution and to compute the parameters of base frequencies, transition/transversion rate ratios (Ti/Tv), and gamma distribution shape parameters for the construction of phylogenetic trees and analyses of codon maximum likelihood.

2.6. Selective pressure analysis

Analyses were performed using the CODEML program from PAML version 3.15 (Yang, 1997). For a given tree and codon model, CODEML finds the set of parameter values (i.e., the likelihood score). Nested models were compared using a likelihood ratio test (LRT) (Yang et al., 2000). The LRT statistic was calculated using $2\Delta\ln L$ to compare the nested models to determine the best model. The significance of the LRT statistic was determined using a χ^2 distribution. Because high divergence can reduce the power of detecting the positive selection under models of variable ω ratios among sites (Yang and Nielsen, 2002), we excluded the sequence of fish and rodents, leaving other sequences in the dataset.

To analyze the possibility that positive selection acts on zokor and *Spalax Epo*, we used the maximum likelihood codon model from the CODEML program in the PAML package (Yang, 1997). The topology of the ML tree mentioned above was modified for all CODEML analyses. We treated branch B as the foreground branch and all other branches in the phylogeny as background branches. The branch-specific models allow for variable ω ratios among branches but invariable ω ratios in sites in the tree and can be implemented for the study of changes in selective pressures in specific lineages (Yang and Nielsen, 2002). The null model assumed the same ω ratio for all lineages in the tree (one ratio model) and the two-ratio models assigned two ω ratios, for the foreground (ω_1) and background branches (ω_0) respectively. The site-specific model allows the ω ratio to vary among sites but fix one ω ratio in all lineages (Nielsen and Yang, 1998). Three pairs of models, M1_a (nearly neutral) vs. M2_a (positive selection), M7 (beta) vs. M8 (beta and ω), and M0 (one-ratio) vs. M3 (discrete), were carried out in site-specific models (Wong et al., 2004). The branch-site models (models A and B) allow the ω ratio to vary both among sites and among lineages and were used to detect positive

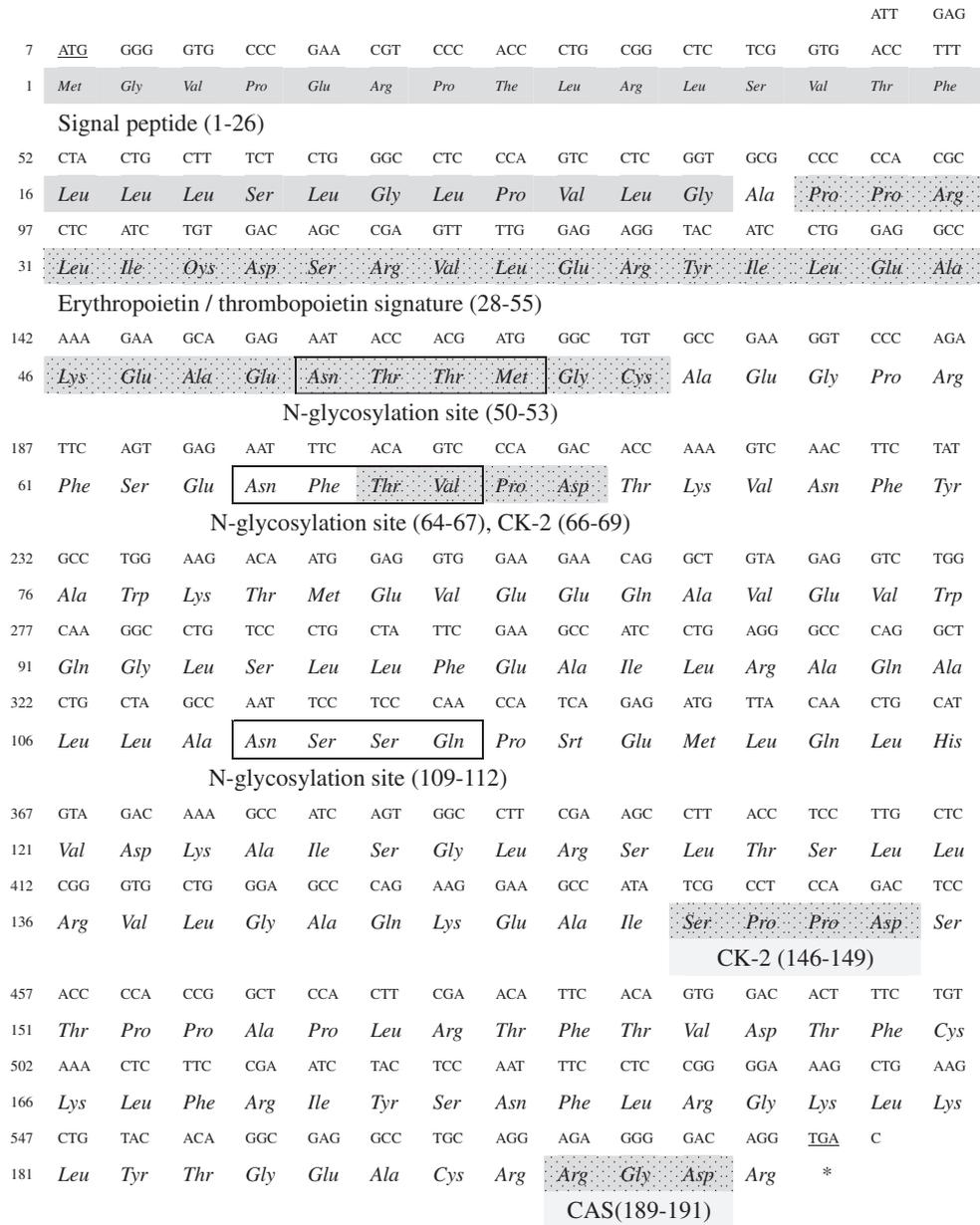


Fig. 1. Plateau zokor *EPO* mRNA and deduced amino acid sequence. The predicted motifs and signal peptide sequence are shaded with a gray background with name abbreviations under them. Boxed amino acid sequences indicate the motif site of *N*-glycosylation site. The start codon ATG and the stop codon TGA are underlined. The asterisk in amino acid sequence indicates the stop codon. The GenBank accession number of plateau zokor *EPO* gene is 1021725. The abbreviation of motifs is presented as follows: CK-2, casein kinase II phosphorylation site; CAS, Cell attachment sequence; SNP: 36M- A/G; 168M- C/A.

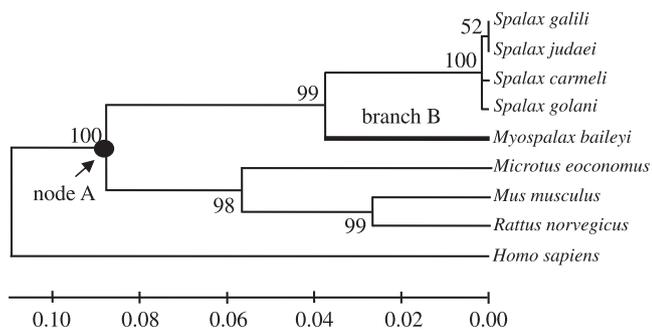


Fig. 2. Maximum likelihood tree of the *Epo* gene based on nucleotide sequences of *Epo* gene Bootstrap (100 000 replicates; seed = 80624). Node A indicates reconstructed ancestral sequences. Branch B indicates the zokor branch.

selection that affects only a few sites along a few lineages (Yang and Nielsen, 2002). In model A, ω_0 was assigned $0 < \omega_0 < 1$, and ω_1 was fixed at 1; hence, positive selection was permitted only in the foreground branch (Zhang et al., 2005). In model B, ω_0 and ω_1 are free

Table 3
Likelihood ratio test statistics ($2\Delta\ell$) for zokor *Epo* evolution analysis.

	$2\Delta\ell$	df	p-value
LRT of ω at branch B (Fig. 2)			
One ratio vs. two ratio	0.469	1	0.049
LRTs of variable ω values among sites			
M1a vs. M2a	0.000	2	1.000
M7 vs. M8	0.000	2	1.000
One ratio vs. M3	1.498	2	0.047
LRTs of variable ω values along branch B (Fig. 2)			
M1a vs. model A	0.000	2	1.000
M3 vs. model B	0.010	2	0.099

Table 2
Parameter estimates for evolutionary analysis of zokor *Epo*.

Model code	Estimate of parameters	ℓ	positively selected sites
M0: one-ratio	$\omega = 0.170$	-1348.916	None
Branch-specific model			
Two-ratio	$\omega_0 = 0.144, \omega_1 = 0.190$	-1348.681	
Site-specific models			
M1a: nearly neutral (K=2)	$p_0 = 0.941, (p_1 = 0.059)$	-1348.291	Not allowed
M2a: positive selection (K=3)	$p_0 = 0.941, p_1 = 0.028 (p_2 = 0.031) \omega_2 = 1.000$	-1348.291	116M ($p = 0.538$) 144A ($p = 0.553$)
M3: discrete (K=3)	$p_0 = 0.586, p_1 = 0.392 (p_2 = 0.022) \omega_1 = 0.047, \omega_2 = 0.371, \omega_3 = 0.371$	-1348.167	None
M7: beta	$p = 0.730, q = 3.233$	-1348.176	None
M8: beta and ω	$p_0 = 1.000, p = 0.730, q = 3.234 (p_1 = 0.000), \omega = 1.000$	-1348.176	116M ($p = 0.559$) 144A ($p = 0.583$)
Branch-site models			
Model A	$p_0 = 0.731, p_1 = 0.046, (p_2 + p_3 = 0.223), \omega_0 = 0.134, \omega_1 = 1.000, \omega_3 = 2.738$	-1348.291	None
Model B	$p_0 = 0.040, p_1 = 0.028 (p_2 + p_3 = 0.932), \omega_0 = 0.047 \omega_1 = 0.378, \omega_3 = 0.000$	-1348.162	None

and, thus, some sites may evolve by positive selection across the entire phylogeny, whereas other sites may evolve by positive selection in just the foreground branch. Model A is compared with M1a (nearly neutral) and model B is compared with M3 (discrete). Positive selection is indicated when a freely estimated ω parameter is greater than 1 and the LRT reaches a statistically significant level. We applied ML reconstruction of the ancestral sequence using the models of Goldman and Yang (1994) and of Yang et al. (1994). The Bayes theorem was used to identify candidate positive selection sites (Yang et al., 2005).

3. Results

3.1. Cloning *EPO* cDNA from plateau zokor

An *Epo* genomic fragment of 579 bp of *M. baileyi* was amplified and sequenced. The *Epo* mRNA initiated with an ATG codon and terminated with a TGA stop codon. Sequence is published in GenBank (GenBank accession no. 1021725). Two synonymous substitutions were found in zokors at position 36M- A/G and 168M- C/A. The zokor *Epo* is composed of 192 amino acids and encodes an apparent signal peptide sequence of 26 amino acids with signal cleavage site between Gly-26 and Ala-27 (Fig. 1). Thus, the mature secreted protein had a predicted molecular weight of 18.655 kDa and a pI of 7.914.

3.2. Evolutionary analysis

3.2.1. Phylogenetic tree construction

The best-fit model of molecular evolution of *Epo* sequence obtained from ModelTest 3.7 (Posada and Crandall, 1998) based on the likelihood ratio test was the HKY+G model. Settings for this

model were as follows: base frequencies (A=0.2148, C=0.3107, G=0.2656 and T=0.2089); transition / transversion ratio ($T_1/T_2 = 2.27$); and a shape parameter of the gamma distribution of 0.560. Parameters obtained from this analysis were used for the construction of the phylogenetic trees (Saitou and Nei, 1987). All phylogenetic trees constructed by NJ (neighbor joining), MP (maximum parsimony) and ML (maximum likelihood) methods produced similar topologies. Only the ML trees constructed from the nucleotide sequences is presented. The ancestral sequence at node A was reconstructed. The ML tree of *Epo* subdivided the eight rodents into two groups: subterranean rodents *M. baileyi* and *Spalax* allospecies, and non-subterranean rodents, rat, mouse and *Microtus oeconomus* (Fig. 2).

3.2.2. Selective pressure analysis

Likelihood values and parameters, as well as likelihood ratio test statistics of zokor *Epo* gene, are shown in Tables 2 and 3. One-ratio model produced the results of $\ell = -1348.916$ and $\omega = 0.170$. In the branch-specific likelihood analysis, the log likelihood value for zokor and *Spalax* branch ($\ell = -1348.681, \omega_1 = 0.190$) was significantly different from that for all other branches ($\ell = -1348.681, \omega_0 = 0.144$) ($2\Delta\ell = 0.469, df = 1, p = 0.049$). In site-specific likelihood models, M2a model suggested that 2.802% of sites were under positive selection with $\omega_B = 1.000$ and identified 116M and 144A under positive selection at the 53.8% and 55.3% probability ($p = 0.031$), respectively. M8 (beta and ω) produced similar results and also identified the same two sites under positive selection: 116M (55.9%) and 144A (58.3%). The differences between M7 and M8 were not statistically significant ($2\Delta\ell = 0.00005, df = 2, p = 1.000$). In the branch-site models, model A and model B did not detect the existence of positive selection sites. Thus, the LRT statistic of the M1a-M2a comparison was not of statistical significance. The

Table 4
Parameter estimates for evolutionary analysis of *Spalax* *Epo*.

Model code	Estimate of parameters	ℓ	Positively selected sites
M0: one-ratio	$\omega = 0.213$	-1422.290	None
Branch-specific model			
Two-ratio	$\omega_0 = 0.188, \omega_1 = 0.311$	-1422.002	
Site-specific models			
M1a: nearly neutral (K=2)	$p_0 = 0.864, (p_1 = 0.136)$	-1418.977	Not allowed
M2a: positive selection (K=3)	$p_0 = 0.921, p_1 < 0.001, (p_2 = 0.079) \omega_2 = 1.000$	-1418.900	102R(0.530), 116M(0.553), 144A(0.581), 152P(0.708)
M3: discrete (K=3)	$p_0 = 0.312, p_1 = 0.609 (p_2 = 0.079) \omega_1 = 0.145 \omega_2 = 0.145, \omega_3 = 1.505$	-1418.900	81E(0.594), 102R(0.809), 116M(0.864), 144A(0.850), 152P(0.757), 153P(0.963), 154A(0.727)
M7: beta	$p = 0.385, q = 1.222$	-1419.259	None
M8: beta and ω	$p_0 = 0.924, p = 17.058, q = 99.000 (p_1 = 0.076), \omega = 1.528$	-1418.903	102R(0.543), 116M(0.545), 144A(0.590), 153P(0.704)
Branch-site models			
Model A	$p_0 = 0.793, p_1 = 0.119, (p_2 + p_3 = 0.089), \omega_0 = 0.107, \omega_1 = 1.000, \omega_3 = 1.000$	-1418.838	None
Model B	$p_0 = 0.683, p_1 = 0.066, (p_2 + p_3 = 0.252), \omega_0 = 0.120 \omega_1 = 1.524, \omega_3 = 0.316$	-1418.733	81E(0.519), 102R(0.738), 116M(0.579), 144A(0.786), 152P(0.636), 153P(0.910), 154A (0.572)

Table 5
Likelihood ratio test statistics ($2\Delta\mathcal{L}$) for *Spalax* Epo evolution analysis.

	$2\Delta\mathcal{L}$	df	p-value
LRT of ω at branch B (Fig. 2)			
One ratio vs. two ratio	0.577	1	0.045
LRTs of variable ω values among sites			
M1a vs. M2a	0.155	2	0.093
M7 vs. M8	0.713	2	0.070
One ratio vs. M3	6.781	2	0.034
LRTs of variable ω values along branch B (Fig. 2)			
M1a vs. model A	0.279	2	0.087
M3 vs. model B	0.332	2	0.085

discrete model (M3) with K=3 site classes did not detect the existence of positive selection sites, although the M3 model was significantly better than one-ratio model ($2\Delta\mathcal{L} = 8.407$ df=2, $p = 0.015$).

Likelihood values and parameters, as well as likelihood ratio test statistics of *Spalax Epo* gene, are shown in Tables 4 and 5. In the branch-specific likelihood analysis, the log likelihood value for zokor and *Spalax* branch ($\mathcal{L} = -1422.002$, $\omega_1 = 0.311$) was significantly different from that for all other branches ($\mathcal{L} = -1348.681$, $\omega_0 = 0.188$) ($2\Delta\mathcal{L} = 0.577$, df=1, $p = 0.045$). In site-specific likelihood models, M2a model identified 4 sites under positive selection: 102R (53.0%) 116M (55.3%), 144A (58.1%) and 152P (70.8%). The discrete model (M3) with K=3 site classes showed 7 sites under positive selection: 81 E(59.4%), 102R(80.9%), 116M(86.4%), 144A(85.0%), 152 P(75.7%), 153 P(96.3%), 154 A(72.7%). The M3 model was significantly better than one-ratio model ($2\Delta\mathcal{L} = 6.781$, df=2, $p = 0.034$). M8 (beta and ω) produced similar results as M2a and identified 4 sites under positive selection. The differences between M7 and M8 were not statistically significant ($2\Delta\mathcal{L} = 0.713$, df=2, $p = 0.070$). In the branch-site models, model B produced similar results as M3 and detected 7

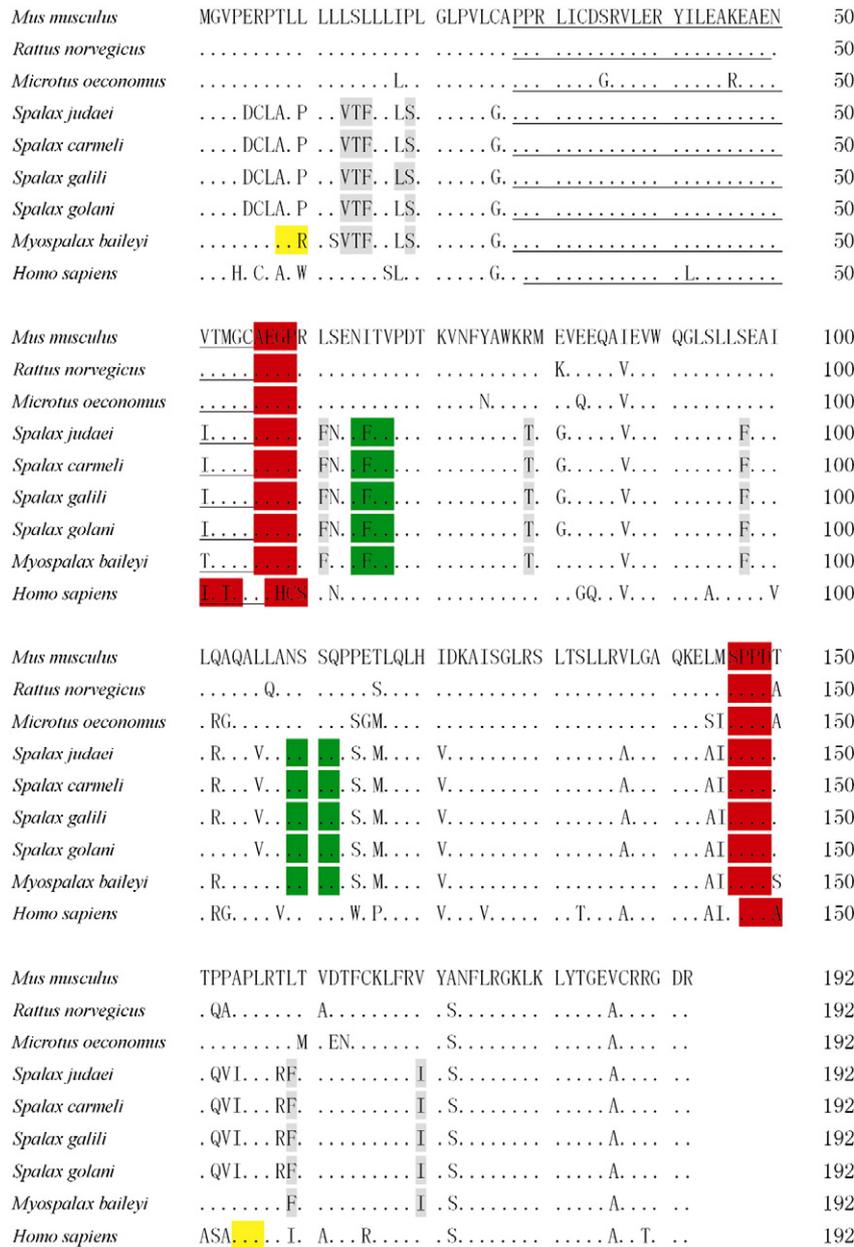


Fig. 3. Multiple alignments of EPO amino acids sequences. Residues identical to *Mus musculus* EPO are presented as dots (.). The predicted motifs are shaded by different colors [Protein kinase C phosphorylation site (PKC) is yellow; Casein kinase II phosphorylation site (CK2) is red; N-glycosylation site is green]. Underlined amino acid sequences indicate the motif of the Erythropoietin/thrombopoietin signature site. Convergent evolution sites between *M. baileyi* and *Spalax* are gray. The numbers at the right are the total numbers of amino acids.

sites under positive selection. Thus, the LRT statistic of the M1a-M2a comparison was not statistically significant.

3.2.3. Secondary and tertiary structure analysis

To be consistent with the evolutionary analysis, predicted mature protein (166 amino acids) were included in the analysis. The consensus methods of secondary structure prediction suggested that zokor Epo, like that of all other lineages, were composed of 4 helices with two conservative CYS sites at 96 and 146, forming one disulfide bond for structural stabilization. The tertiary structure of zokor Epo was based on a model of human Epo (1BUY) (Cheatham et al., 1998) from the Protein Data Bank. The motifs predicted in the Epo structure indicated that the protein kinase C phosphorylation sites (PKC) and the casein kinase II phosphorylation sites (CK2) are conserved among these lineages and an N-glycosylation site existed only in zokor Epo and *Spalax* Epo. There are nine convergent evolution sites existing between zokor and *Spalax*: 13V, 14T, 15F, 19S, 61F, 79T, 97F, 159F, 170I (Fig. 3).

On the basis of its primary amino acid sequence and disulfide bonds, Epo is predicted to have a four-antiparallel amphipathic α -helical bundle structure in common with other members of the cytokine family (Wen et al., 1994). We have shown that deletions in nonhelical regions at the N terminus, the C terminus, and in the loops between helices resulted in the formation of Epo proteins that are readily secreted from the cell and are biologically active. These regions (Fig. 4) can be ruled out as domains essential for function, such as the sites involved in the binding of Epo to its receptor.

4. Discussion

In the present study, we have compared the entire coding sequence of Epo from different lineages of representative mammals in order to identify the variation of functional sites and to understand the mechanism of functional evolution of zokor Epo. Although *M. baileyi* belongs to the genus *Myospalax* of the family Cricetidae and

Spalax allospecies to the family Spalacidae, the phylogenetic tree of Epo showed that these two subterranean species converged into one branch, indicating such convergent evolution was driven by similar hypoxic ecological niches.

To determine the nature of variation that occurred at these loci in zokor Epo, a set of evolutionary analysis was performed. A comparison between one-ratio and two-ratio branch-specific models revealed the ω ratio of the zokor lineage was significantly different from that of all other lineages. Among site-specific models, both the M8 (beta and ω) and M2a (positive selection, $K=3$) models demonstrated varied ω ratios among sites, yielding ω ratio of 1.000, and predicted some common sites under positive selection in zokor (116M and 144A) and *Spalax* (102R, 116M, 144A and 152P). The ancestral sequence reconstructed by the models of Goldman and Yang (1994) and of Yang et al. (1994) suggested the following amino acid changes along the zokor Epo: 116M and 144A. Obviously, sites under positive selection, inferred by Bayes prediction, were highly consistent with these from the reconstruction of the ancestral sequence. Therefore, our evolutionary analysis confirmed the previous hypothesis that adaptive evolution occurred in zokor Epo.

The establishment of new or modified function of a protein under specific stress is derived from the adaptive evolution in this protein. We speculate the possible effect of positive selection sites on the functional evolution of zokor Epo interpreted from the analysis of literature on the evolutionary, functional-structural, and biochemical information concerning the zokor protein. Previous investigations indicated that interhelical loops (AB: $\Delta 32-36$, $\Delta 53-57$; BC: $\Delta 78-82$; CD: $\Delta 111-119$) of mature Epo protein play an important role in biological activity (Boissel et al., 1993). According to the algorithm of Emini et al. (62), the residues 111–119 are predicted to be at the surface of the molecule. Since the $\Delta 111-119$ mutant is readily secreted and has full biological activity, it seems unlikely that the putative β -sheet segment in the CD loop is an important determinant of molecular stability. Our results showed the codes for important sites in the biological activity of EPO.

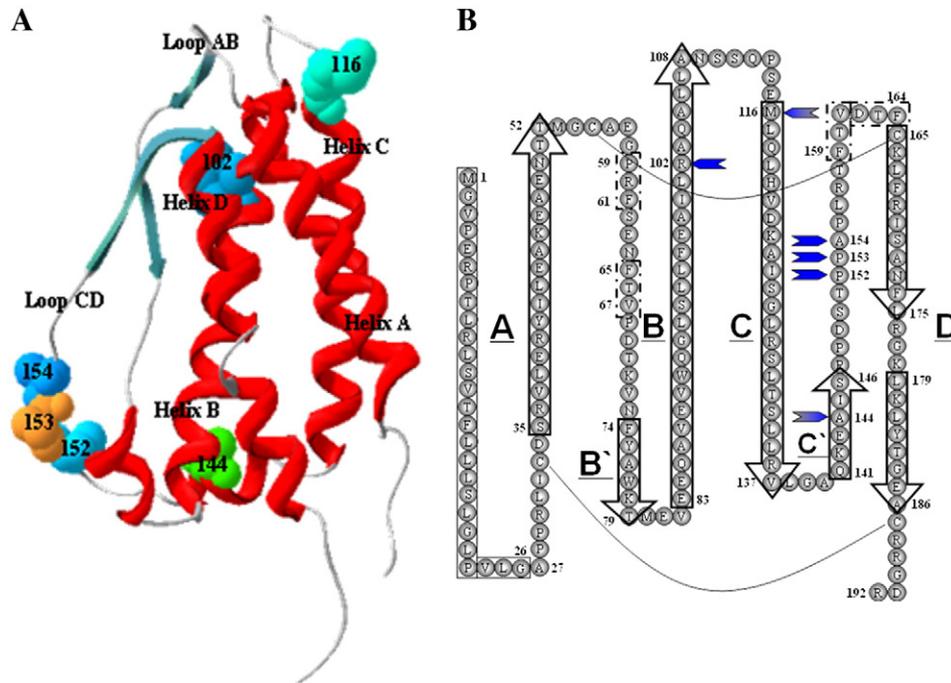


Fig. 4. Model of the three-dimensional structure of erythropoietin. A, ribbon diagram of the predicted Epo tertiary structure. An apparent signal peptide sequence of 26 amino acids with signal cleavage site between Gly-26 and Ala-27 is delineated by the solid rectangle. The four α -helices are labeled A–D (red); Loops between helices are named for the helices they interconnect. Two regions of extended structure which could form hydrogen bonds between Loop AB and Loop CD are also presented (cyan). Disulfide bonds bridge residues 29–139 and 7–161. This folding pattern is strongly suggested by the large size of the two interconnecting loops AB and CD. B, schematic representation of Epo's primary structure depicting predicted up-up-down-down orientation of the four antiparallel α -helices (boxes with arrowhead). This folding pattern is strongly suggested by the large size of the two interconnecting loops AB and CD. The limits of each helix were drawn accordingly to Fig. 1. A predicted short region of β -sheet is delineated by the dashed rectangle.

Previous studies demonstrated blunted erythropoietic response in human populations living at high-altitude, which is considered to be beneficial (Beall, 2007; Garruto et al., 2003; Wu et al., 2005). It is possible that Andean highlanders have not evolved a similar mechanism for attenuating the erythropoietic response to hypoxia because of their shorter history of residence at high altitude (Beall, 2007; Moore et al., 2001). In laboratory conditions, by adjusting the P_{O_2} , we found that the number of red blood cells in plateau zokors increased with hypoxia (unpublished data), which is contrary to the results of previous human study. However, whether such function is consistent in the natural population is not known. Further research is required to test whether the structural change in EPO of zokor results in increased number of red blood cells or increased efficiency in oxygen utilization in natural populations.

The EPO gene expressions are regulated by GATAs, which bind to the GATA site of the EPO gene promoters (Aird et al., 1994; Imagawa, et al., 1997). This type of regulation is not considered in present study, the expressions of EPO gene will be conducted in the further research.

In conclusion, this study determined the predicted amino acid sequence of EPO of the plateau zokor so that it can be compared to that of other rodents and vertebrates and the intrinsic properties of zokor EPO in adaptation to underground life at high altitude can be identified. This genetic diversity in zokor EPO may be one of the reasons of adaptation to the life underground in harsh plateau environment.

Acknowledgments

This work was supported by the National Science Foundation of China (no. 31140077), Key Innovation Research Programs of the Chinese Academy of Sciences (no. KSCX2-EW-N-05), grants from the "973" Project of China Ministry of Science and Technology (no. 2012CB722506) and Key Projects in the National Science & Technology Pillar Program (no. 2009BAI83B01). We would like to thank Dr. Ji Weihong of Massey University for her comments and careful editing of this manuscript.

References

- Aird, W.C., Parvin, J.D., Sharp, P.A., Rosenberg, R.D., 1994. The interaction of GATA binding proteins and basal transcription factors with GATA box containing core promoters. *J. Biol. Chem.* 269, 883–889.
- Baldi, P., Pollastri, G., 2003. The principled design of large-scale recursive neural network architectures DAG-RNNs and protein structure prediction problem. *J. Mach. Learn. Res.* 4, 575–602.
- Beall, C.M., 2007. Two routes to functional adaptation: Tibetan and Andean high altitude natives. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8655–8660.
- Boissel, J.P., Lee, W.R., Presnell, S.R., Cohen, F.E., Bunn, H.F., 1993. Erythropoietin structure-function relationships. Mutant proteins that test a model of tertiary structure. *J. Biol. Chem.* 268, 15983–15993.
- Cheatham, J.C., et al., 1998. NMR structure of human erythropoietin and a comparison with its receptor bound conformation. *Nat. Struct. Biol.* 5, 861–866. doi:10.1038/2302.
- Ebert, B.L., Bunn, H.F., 1999. Regulation of the erythropoietin gene. *Blood* 94, 1864–1877.
- Erslev, A.J., 1974. In vitro production of erythropoietin by kidneys perfused with a serum-free solution. *Blood* 44, 77–85.
- Felsenstein, J., 2006. PHYLIP: Phylogeny Inference Package, Version 3.66. University of Washington, Seattle.
- Garruto, R.M., Chin, C.-T., Weitz, C.A., Liu, J.-C., Liu, R.-L., 2003. Hematological differences during growth among Tibetans and Han Chinese born and raised at high altitude in Qinghai, China. *Am. J. Phys. Anthropol.* 122, 171–183.
- Goldberg, M.A., Dunning, S.P., Bunn, H.F., 1988. Regulation of the erythropoietin gene: evidence that the oxygen sensor is a heme protein. *Science* 242, 1412–1415.
- Goldman, N., Yang, Z.H., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11, 725–736.
- Guex, X., Peitsch, M.C., 1997. SWISS-MODEL and Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723.
- Imagawa, S., Yamamoto, M., Miura, Y., 1997. Negative regulation of the erythropoietin gene expression by the GATA transcription factors. *Blood* 89, 1430–1439.
- Jacobson, L.O., Goldwasser, E., Fried, W., Plzak, L., 1957. Role of the kidney in erythropoiesis. *Nature* 179, 633–634.
- Kurtz, A., Eckardt, K.E., Tannahill, L., Bauer, C., 1988. Regulation of erythropoietin production. *Contrib. Nephrol.* 66, 1–16.
- Moore, L.G., Young, D., McCullough, R.E., Droma, T., Zamudio, S., 2001. Tibetan protection from intrauterine growth restriction (IUGR) and reproductive loss at high altitude. *Am. J. Hum. Biol.* 13, 635–644.
- Nielsen, R., Yang, Z.H., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936.
- Norris, R.W., Zhou, K.Y., Zhou, C.Q., Yang, G., Kilpatrick, C.W., Honeycutt, R.L., 2004. The phylogenetic position of the zokors (Myospalacinae) and comments on the families of muroids (Rodentia). *Mol. Phylogenet. Evol.* 31, 972–978.
- Pollastri, G., Baldi, P., 2002. Prediction of contact maps by recurrent neural network architectures and hidden context propagation from all four cardinal corners. *Bioinformatics* 18, 62–70.
- Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Qi, X.Z., Wang, X.J., Zhu, S.H., Rao, X.F., Wei, L., Wei, D.B., 2008. Hypoxic adaptation of the hearts of plateau zokor (*Myospalax baileyi*) and plateau pika (*Ochotona curzoniae*). *Acta Physiol. Sin.* 60, 348–354 (in Chinese).
- Robinson, R.M., Huchon, D., 2000. RR tree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* 16, 296–297.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Schooley, J.C., Mahlmann, L.J., 1972. Evidence for the de novo synthesis of erythropoietin in hypoxic rats. *Blood* 40, 662–670.
- Shoemaker, C.B., Mitscock, L.D., 1986. Murine erythropoietin gene: cloning, expression and human gene homology. *Mol. Cell. Biol.* 6, 849–858.
- Swofford, D.L., 2000. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4.0. Sinauer Associates Inc., Sunderland, Massachusetts.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aides by quality analysis tools. *Nucleic Acids Res.* 24, 4876–4882.
- Wang, Z.W., Zeng, J.X., Han, Y.C., 1979. Studies on the metabolism rates of the mouse hare (*Ochotona curzoniae*) and the mole rat (*Myospalax fontanierii*). *Acta Zool. Sin.* 25, 75–85 (in Chinese).
- Wang, X.J., Wei, D.B., Wei, L., Zhang, J.M., Yu, H.Y., 2008. Physiological character of erythrocyte adapting to hypoxia in plateau zokor and plateau pika. *Sichuan J. Zool.* 27, 1100–1104 (in Chinese).
- Wei, D.B., Wei, L., Zhang, J.M., Yu, H.Y., 2006. Blood-gas properties of plateau zokor (*Myospalax baileyi*). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 14, 372–375.
- Wen, X., Zheng, S.M., Chen, Q., 1994. Studies on the production of poly-β-hydroxybutyric acid from beet molasses by *Alcaligenes eutrophus*. *Microbiol.* 21, 71–75.
- Wong, W.S., Yang, Z.H., Goldman, N., Nielsen, R., 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168, 1041–1051.
- Wu, T., et al., 2005. Hemoglobin levels in Qinghai–Tibet: different effects of gender for Tibetans vs. Han. *J. Appl. Physiol.* 98, 598–604.
- Yang, Z.H., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13, 555–556.
- Yang, Z.H., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19, 908–917.
- Yang, Z.H., Kumar, S., Nei, M., 1994. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141, 1641–1650.
- Yang, Z.H., Nielsen, R., Goldman, N., Pedersen, A.M., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.
- Yang, Z.H., Wong, W.S., Nielsen, R., 2005. Bayes empirical Bays inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118.
- Zanjani, E.D., Poster, J., Burlington, H., Mann, L.I., Wasserman, L.R., 1977. Liver as the primary site of erythropoietin formation in the fetus. *J. Lab. Clin. Med.* 89, 640–644.
- Zhang, Y.M., Liu, J.K., 2003. Effects of plateau zokors (*Myospalax fontanierii*) on plant community and soil in an alpine meadow. *J. Mammal.* 84, 644–651.
- Zhang, Y.M., Zhang, Z.B., Liu, J.K., 2003. Burrowing rodents as ecosystem engineers: The ecology and management of plateau zokors *Myospalax fontanierii* in alpine meadow ecosystems on the Tibetan Plateau. *Mammal. Rev.* 33, 284–294.
- Zhang, J., Nielsen, R., Yang, Z.H., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22, 2472–2479.