Gene 501 (2012) 206–212

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

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, Hennan, 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.

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 Mitsock, 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 PO_2 (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 O_2 and CO_2 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., 20e06). 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@nwipb.ac.cn (Y. Zhang).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2012.02.045

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	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-μl volume containing 5× First-Strand Buffer, Oligo(dT)18 (20 μg/ml), 10 Mm dNTPs, 0.1 M DTT, RNaseOUT (40 units/μl) and Reverse Transcriptase (40 units/μl). 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 μM dNTP, 0.3 μM primers (Beijing Genomics Institute, Beijing, China) and 2× GC Buffer I (Mg^{2+} plus) with 5 μg 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-lyonl.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 $2\Delta\ell$ 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