J Exp Biol Advance Online Articles. First posted online on 19 January 2016 as doi:10.1242/jeb.133272
Access the most recent version at http://jeb.biologists.org/lookup/doi/10.1242/jeb.133272

Gene expression and adaptive evolution of *ZBED1* in the hibernating great horseshoe bat (*Rhinolophus ferrumequinum*)

Yanhong Xiao[1], Yonghua Wu[1,2]*, Keping Sun[1], Hui Wang[1], Tinglei Jiang[1], Aiqing Lin[1], Xiaobin Huang[1], Xinke Yue[1], Limin Shi[3], Jiang Feng[1]*

[1]Jilin Provincial Key Laboratory of Animal Resource Conservation and Utilization, Northeast Normal University, 2555 Jingyue Street, Changchun 130117, China

[2]School of Life Sciences, Northeast Normal University, 5268 Renmin Street, Changchun 130024, China

[3]School of Life Science, Yunnan Normal University, Chenggong District, Kunming 650500, China

*Corresponding authors: Jiang Feng; Yonghua Wu

Phone: +86-0431-85099866

Fax: +86-0431-89165677

E-mail: fengj@nenu.edu.cn and wuyh442@nenu.edu.cn

Journal of Experimental Biology • Advance article

Summary statement

We for the first time investigated the expression and adaptive evolution of *ZBED1* gene, an important cell proliferation-related transcription factor, and detected the potential role of *ZBED1* gene in hibernation.

Abstract

Mammalian hibernators experience physiological extremes, e.g. ischemia, muscle disuse, hypothermia, which are lethal to non-hibernators, implying the existence of underlying mechanisms that allow hibernators to withstand these physiological extremes. Increased cell proliferation is suggested to be such a strategy, while its molecular basis remains unknown. In this study, we characterized the expression pattern of *ZBED1* (zinc finger, BED-type containing 1), a transcription factor that plays a crucial role in regulating cell proliferation, in five tissues of the great horseshoe bat (*Rhinolophus ferrumequinum*) during pre-hibernation, deep hibernation and post-hibernation. Moreover, we investigated the *ZBED1* genetic divergence from individuals with variable hibernation phenotypes that cover all three known mtDNA lineages of the species. Expression analyses showed that *ZBED1* is only over-expressed in brain and skeletal muscle, but not in the other three tissues, suggesting an increased cell proliferation in these two tissues during deep hibernation. Evolutionary analyses showed that *ZBED1* sequences were clustered into two well-supported clades with each one dominated by hibernating and non-hibernating individuals, respectively. Positive selection analyses further showed some positively selected sites and a divergent selection pressure among hibernating and non-hibernating groups of *R. ferrumequinum*. Our results suggest that *ZBED1* may be a potential candidate gene that regulates cell proliferation for hibernators to face physiological extremes during hibernation.

Keywords: Zinc finger, BED-type containing 1; cell proliferation; gene expression; adaptive evolution; protective mechanism

Introduction

Hibernation is a strategy taken by hibernators to survive the pressures of a harsh environment such as low ambient temperature and/or scarce food availability (Wang et al., 1996). As the cost of survival, hibernators must suffer potential adverse effects induced by long periods of low body temperature, repeated ischemia and perfusion, some of which are highly stressful and even lethal for non-hibernators (Carey et al., 2003). However, hibernators obviously have abilities to protect themselves from these potential harms, and several potential underlying mechanisms are proposed, such as induced expression of stress (heat shock) proteins to combat cellular stress (Carey et al., 1999; Carey et al., 2000; Lei et al., 2014), and increased protein-ubiquitin conjugate concentration to help degrade damaged proteins during hibernation (Carey et al., 2000).

Among protective mechanisms that underlie the tolerance of physiological extremes, an increased cell proliferation is one suggested mechanism (Cerri et al., 2009). It is reported that both the number of apoptotic cells and the number of cells that undergo proliferation in the main brain areas of hibernating frogs increase compared with active frogs, suggesting a compensatory effect of increased cell proliferation for increased apoptotic cells during hibernation (Cerri et al., 2009). During hibernation, it is stressful to cells to be in a long period of low temperature due to its effects on such processes as protein stability, membrane function, ATP synthesis, activity of key regulatory enzymes, and cytoskeletal integrity (Kandror and Goldberg, 1997; Sonna et al., 2002). Additionally, ischemia-induced oxidative stress during hibernation is considered to play a crucial role in causing cell death (Dave et al., 2012). Regarding the high risk of cell death, an increased cell proliferation might be an important protective mechanism for hibernating mammals to cope with potential cell damage during hibernation.

ZBED1 (zinc finger, BED-type containing 1), also called DREF (DNA replication-related element-binding factor), is a transcription regulatory factor playing a key role in the control of cell proliferation (Matsukage et al., 2008). ZBED1 regulates cell proliferation by controling many cell proliferation-related genes, such as encoding PCNA, Orc 2, DNA polymerase α 180 kDa and 73 kDa subunits, RFC140, raf, E2F, cyclin A, Skp A, etc (Hirose et al., 1991; Hirose et al., 1993; Ohno et al., 1996; Takahashi et al., 1996; Kandror and Goldberg, 1997; Ryu et al., 1997; Sawado et al., 1998;

Phuong Thao et al., 2006; Tsuchiya et al., 2007; Matsukage et al., 2008). To date, no study about the role of *ZBED1* in hibernation has been reported. Given the important role of *ZBED1* in cell proliferation, it can be hypothesized to be a potential factor for helping physiological adaption in hibernators. If it is the case, an increased expression of *ZBED1* in tissues vulnerable to stresses induced by physiological extremes during hibernation would be expected and adaptive evolution of the *ZBED1* gene in hibernating taxa may occur compared with non-hibernating taxa.

The greater horseshoe bat, *Rhinolophus ferrumequinum*, has a wide distribution in China and is found from southern China to northern China, and three divergent mtDNA evolutionary groups (Northeast, Central-East, Southwest) are identified (Sun et al., 2013). The Northeast group, which is distributed in the northeast of China, strictly hibernates from December to April during winter, while the Southwest group, which is distributed in the southwest of China, where the weather is warm and food is available even during winter, is active year-round (Aulagnier et al., 2008). Many individuals in the Central-East group (e.g., Beijing and Jinan), also hibernate during winter. The different hibernation phenotypes therefore provide us an ideal model system to study the potential role of *ZBED1* in hibernation.

In this study, we characterized relative mRNA expression levels of the *ZBED1* gene in five tissues including heart, kidney, brain, liver and skeletal muscle, from Northeast individuals of *R. ferrumequinum* during pre-hibernation, deep hibernation and post-hibernation, respectively. We also analyzed adaptive evolution of *ZBED1* based on individuals from all three mtDNA groups. We aimed to (i) determine the expression pattern of the *ZBED1* gene in different tissues across the hibernation cycle and (ii) determine the potential genetic divergence and possible adaptive evolution of the *ZBED1* gene among hibernating and non-hibernating groups of *R. ferrumequinum*.

Materials and methods

Gene expression analyses of *ZBED1*

Samples collection

A total of 16 female individuals, belonging to Northeast group, were used in this study and these individuals were collected at different time points in Jilin Province of northeast China (Table 1). Of which, three individuals were sampled in September in the Ground Cave (125°50'25" E, 41°4'8" N), referred to as pre-hibernation individuals, three and five individuals were sampled in December and April from Ground Cave and Pigeon Cave (125°56'19" E, 41°28'29" N), respectively, referred to as deep hibernation individuals, and five individuals were sampled in May from Pigeon Cave, referred to as post- hibernation individuals. The pre-hibernation individuals and post-hibernation individuals were active and had obvious responses to stimuli (sound, touch and light) and high body temperatures, while the deep hibernation individuals were in a torpid state, and had no response to stimuli with body temperatures close to ambient temperature. All bats were euthanized by decapitation as soon as they were caught in the caves to ensure RNA profiles most representative of their active or hibernation state as well as to minimize pain and suffering. Surgical procedures were promptly performed to protect RNA from degradation. Five tissue types (heart, kidney, brain, liver and skeletal muscle) of each animal were rapidly excised, flash frozen in liquid nitrogen, and then stored at −80°C until processed for RNA isolation. All experimental procedures carried out in this study were approved by the National Animal Research Authority of Northeast Normal University, China (approval number: NENU-20080416) and the Forestry Bureau of Jilin Province of China (approval number: [2006]178).

Total RNA isolation

Total RNA was isolated from bat tissues using TRIzol Reagent (Life Technologies Inc., Carlsbad, CA). Approximatly 50 mg of tissue was ground in liquid nitrogen using a pestle and mortar, and then dissolved in 1 ml of ice-cold TRIzol Reagent and incubated for 5 minutes. The cell debris was pelleted by centrifuging the homogenates at 12000 × g, 4°C for 10 minutes. 0.2 ml 100% chloroform was added

to the supernatant, and then tubes were shaken vigorously by hand for 15 seconds and incubated at room temperature for 5 minutes. Then the aqueous phase was transferred to a fresh tube after centrifuging at 12000 × g, 4°C for 15 minutes and an equal volume of 100% chloroform was then added. Shake tubes vigorously by hand for 15 seconds and incubate them on ice for 2–3 minutes and then centrifuge at 12000 × g, 4°C for 10 minutes to extract again. Next, 0.5 ml of isopropyl alcohol was added to the aqueous phase transferred into a fresh tube and the RNA was precipitated at −80°C overnight. The next day, the supernatant was removed from the tube after centrifuged at 12000 × g, 4°C for 10 minutes. Then the RNA pellet was washed with 1 ml of 75% ethanol through mixing the sample by vortex and centrifuging at 7500× g, 4°C for 5 minutes. Finally, the RNA pellet was air-dried for 5-10 minutes, and then dissolved in RNase-free water. Non-denaturing agrose gel electrophoresis and a NanoDrop spectrophotometer (Thermo Fischer Scientific Inc., Waltham, MA) were used to assess the quality and quantity, respectively, of the isolated RNA. A260/280 values were all above 2.0 and electrophoresis of the RNA samples demonstrated that 28S and 18S rRNA were intact.

cDNA synthesis

A total of 80 RNA samples were converted to cDNA by using *TransScript* One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech Co., Ltd, Beijing, CN) following the manufacturer's instruction. The reaction mixture contained approximatly 1μg RNA template, 0.5 μg Anchored Oligo $(dT)_{18}$ primer, and 6 μl RNase-free water and was incubated at 65 °C for 5 min, then on ice for 2 minutes. In the next step, a mixture containing 2×TS Reaction Mix, *TransScript* RT/RI Enzyme Mix, and gDNA Remover was added into the mixture from the first step. The final 20 μl reaction mixture was incubated at 42 °C for 30 minutes, and finally 85 °C for 5 minutes.

Real-time quantitative PCR

Quantitative RT-PCR was performed using StepOne Real-Time PCR System (Applied Biosystem) and an automatic threshold calculated by the StepOne software v2.1 was used. The primer pairs of *ZBED1* gene and reference gene *β-actin* specified for *R. ferrumequinum* are listed in Table S1, including their sequences, annealing temperatures and product lengths. The efficiencies of these two genes are all approximately 100%, which can ensure the ΔΔCt equation to be used correctly. The total 10 μl PCR

mixture contained 5 μl THUNDERBIRD SYBR qPCR Mix (TOYOBO), 0.2 μl 50×ROX reference dye, 1 μl cDNA template and 0.25 μM of each primer. The PCR was performed under the following conditions: pre-denaturation at 95 °C for 1minutes, then 40 cycles: 95 °C, 15s; 60 °C for 1minutes, data collection after each cycle, then melting curve.

The relative quantity was calculated by using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). The fold changes of *ZBED1* gene in five tissue types between September (pre-hibernation) and December (deep hibernation) were calculated relative to September; between April (deep hibernation) and May (post-hibernation) were calculated relative to April. The relative expression folds were expressed as mean ± SE and statistical significance between active and torpid samples was analyzed using Independent-Samples T Test. In the further analysis, the fold changes of *ZBED1* in kidney, brain, liver and skeletal muscle were calculated relative to heart. The statistical significance among the five tissues was analyzed by one-way ANOVA followed by post hoc Tukey's test (SPSS software 19.0). Differences between means were statistically significant when $P < 0.05$.

Adaptive evolution analyses of *ZBED1*

Samples collection, cloning and sequencing

Forty-six individuals of *R. ferrumequinum* were sampled from 13 different localities covering all three known mtDNA groups (Fig. 1, Table S2). One individual of *Rhinolophus macrotis* was also sampled to be used as an outgroup of phylogenetic analyses. Biopsy punches (3 mm diameter) were used to acquire bat wing membrane (plagiopatagium) samples as outlined by Worthington Wilmer and Barratt (Worthington Wilmer and Barratt, 1996). The 3 mm holes in each wing heal in approximately 4 weeks (Worthington Wilmer and Barratt, 1996). Genomic DNA was extracted from the wing biopsies using the Animal Genomic DNA Isolation Kit (Sangon, China). Primer pairs for amplification of *ZBED1* coding domain sequence (CDS) were designed according to cDNA sequence (KU312321) of *ZBED1* and whole genome shotgun sequence of *R. ferrumequinum* (Accession: AWHA01106959.1 and AWHA01106960.1) by using Primer Premier 5.0. Sense primer sequence was 5'-GCTGTTTGTCTCACTGCTT-3' and anti-sense primer sequence was 5'-ACCCCATCTGTCTGTCACT-3'. To reduce mutations introduced by amplification, Q5™ super

fidelity 2× Master Mix (Biolabs) was used according to the manual instruction. Purification of PCR product was completed by using TaKaRa MiniBEST Agarose Gel DNA Extraction Kit. Cloning of *ZBED1* CDS was performed using pEASY^{TM} Blunt Zero Cloning Kit (TransGen Biotech Co., Ltd, Beijing, CN). Five positive clones of each individual were randomly selected to be sequenced by Shanghai Sangon Biotechnology Co., Ltd. All CDS sequences of *ZBED1* were deposited in GenBank with accession numbers KR868578-KR868649.

Data analyses

All forty-six individuals sampled were sequenced successfully. Some individuals were determined to be heterozygotes and both allele sequences from the heterozygote individuals were used for subsequent analyses. Sequences were aligned using ClustalW implemented in MEGA 6 (Tamura et al., 2013). Polymorphic site detection and haplotype analysis were implemented in Dnasp 5.0 (Librado and Rozas, 2009). The sequence alignment was analyzed with RDP 3 (Martin et al., 2010) implementing several algorithms to detect recombinant sequences.

The phylogenetic relationship of the *ZBED1* gene was constructed using MrBayes 3.2.0 (Ronquist and Huelsenbeck, 2003). The DNA substitution model that best fit the data was estimated using the Akaike information criterion using Modeltest 3.7 (Posada and Crandall, 2005). Bayesian Inference was run with four simultaneous chains, each of 1×10^6 generations, sampled every 100 generations and the first 25% of trees were discarded as 'burn-in'. *ZBED1* CDS of *R. macrotis* was also cloned and sequenced, which was used as the outgroup.

The adaptive evolution of the *ZBED1* gene was determined by positive selection analyses. For the positive selection analyses, the CODEML program in PAML package was used to derive maximum-likelihood estimates of the rate of nonsynonymous substitutions (d_N) and synonymous substitutions (d_S) (Yang, 2007). A d_N/d_S ratio (ω) of >1 signifies positive selection, a ratio of ~ 1 signifies neutrality and a ratio of <1 signifies purifying selection. Branch models, Branch-site models, site-specific models (M1a, M2a, M3, M7, M8), and clade model (Model C) were implemented in the CODEML program. Each model was fit to the data multiple times with different initial omega values to help ensure local optima were avoided. Likelihood ratio tests (LRTs) were used to compare the model fit of complex models against simpler, nested models. LRTs were carried out by comparing twice the difference in *ln*

likelihood scores of nested models against a χ^2 distribution with the degrees of freedom equal to the number of extra parameters estimated by the more complex model. The species tree was constructed based on the published mtDNA phylogenetic tree of *R.ferrumequinum* (Sun et al., 2013), and the inner relationships within each of the three mtDNA groups were respectively constructed by MrBayes 3.2.0 using nucleotide sequences of *ZBED1* CDS. The best-fit models selected for 1st, 2nd and 3rd codon sites within the Northeast group were TIM, HKY and HKY, respectively. Similarly, the best-fit models selected for 1st, 2nd and 3rd codon sites within the Central-East and Southwest group were K81uf, TIM ,HKY+I+G , TIM, F81 and HKY+G, respectively. Based on the species tree, branch models were used to identify branches under positive selection. For this, a one-rate mode and a free rate model were initially employed, and then two-rate modes that assumed the positive selection of the ancestral branches of the Northeast group and the Central-East group of *R. ferrumequinum* were used, respectively. In addition, we also applied two-rate models to the branches with $\omega>1$ identified under the free-ratios model, which assumes an independent ω ratio for each branch, to identify potential positively selected branches in the phylogeny of *R. ferrumequinum*.

Branch-site models were applied to the ancestral branches of the Northeast and the Central-East groups to identify the certain sites under positive selection. Estimates of site-wise ω values in these foreground branches were then compared with estimates across the remaining background branches in the phylogeny. Under Model A, the ω values were assigned to four predefined site classes. Site class 0 was estimated from data but constrained ($0<\omega_0<1$), site class 1 $\omega_1=1$, site class 2a ω_{2a} could exceed 1 on the foreground but is constrained to be under purifying selection on the background, and site class 2b ω_{2b} could exceed 1 on the foreground but not on the background. The null model fixes $\omega_2=1$. The comparison between these two models is so-called test 2 with one degree of freedom (d.f.). In addition, test 1 was also used for the analyses, which compares Model A with M1a (Nearly Neutral), with d.f. =2.

Clade model was applied to test potential divergent selection between hibernating and non-hibernating groups of *R. ferrumequinum*. We compared ω ratios averaged across branches within focal clades (foreground) with ω ratios estimated for the rest of the tree (background) (Bielawski and Yang, 2004). In clade models, we applied a 'multi-clade' approach, which allows for more than two tree partitions to examine complex patterns of divergence in selection across the phylogeny by comparing

such models against simpler, nested models with fewer tree partitions (Weadick and Chang, 2012). Models M1a and M2a_rel were used as null models for Model C in LRTs. M1a is the standard null model for Model C. However, Model C versus M1a LRT was reported prone to false positive test results when faced with moderate among-site variation in selective constraint (Weadick and Chang, 2012). M2a_rel null model proposed by Weadick et al. (2012a) was therefore also used in this study.

Positively selected sites were analyzed using both PAML and HyPhy. With PAML, we compared site models M2a (selection) versus M1a (neutral), M7 (beta) versus M8 (beta & ω). We also compared models M3 (discrete) versus M0 (one ratio) to test change of ω ratio among sites. The Bayes empirical Bayes (BEB) was used to identify sites under positive selection. In HyPhy we used open-source software packages available under the datamonkey web-server (http://www.datamonkey.org/) (Pond and Frost, 2005a, b; Pond and Muse, 2005; Murrell et al., 2012), to implement five different methods including single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), random effects method (REL), mixed effects model of evolution (MEME) and fast unbiased Bayesian approximation (FUBAR).

Results

Differential expression of *ZBED1* gene

The Northeast group of *R. ferrumequinum* is known to have a strict hibernation behavior during winter, and here we analyzed its *ZBED1* expression level in five different tissues across pre-hibernation, deep hibernation and post-hibernation. Three tissues (heart, kidney and liver) consistently presented a decreased expression during deep hibernation (December and April) compared with both pre-hibernation (September) and post-hibernation (May). In contrast, the two other types of tissue, brain and skeletal muscle, showed an increased expression during deep hibernation and decreased expression in both pre-hibernation and post-hibernation (Fig. 2A,B). However, only the expression differences of *ZBED1* in brain and skeletal muscle between April and May were statistically significant with *p*-values of 0.011 and 0.017, respectively. The smaller sample size in September and December than in April and May may limit the statistical difference of *ZBED1* expression in brain and skeletal muscle between September and December.

We further analyzed the relative expression levels of the *ZBED1* gene for each of the five tissue types and for the four time points (September, December, April, and May) (Fig. 3). The results showed that *ZBED1* was expressed strongest in skeletal muscle compared with the other four tissues (Fig. 3). Liver showed the least expression compared with the other four tissues (Fig. 3). Expression of *ZBED1* in kidney, brain and liver showed no significant difference among each other at the three time points: May, December, and April, with *p*-values >0.05 (Fig. 3B,C,D), except for the expression at the active state in September, which showed significant differences among the three tissues (Fig. 3A). The expression differences between heart and kidney were not significant at the four time points (Fig. 3). Hence, significant differences were found for the expression of *ZBED1* in the different types of tissue.

Genetic analysis

Among 46 individuals analyzed, 26 individuals were determined to be heterozygous, and a total of 39 haplotypes were detected (Table S3). Among the 39 haplotypes, 64 polymorphic sites were found, of which 7 sites (1092, 1206, 1281, 1335, 1344, 1555 and 1626) were homozygous and identical between the Northeast group and the Central-East group but different from the Southwest group (Table 2). No evidence of recombinant sequence was found in our dataset. To determine the genetic divergence of the *ZBED1* gene within the species, the Bayes tree based on the complete CDS of the *ZBED1* gene was constructed (Fig. 4A). The best-fit models for 1st, 2nd and 3rd codon sites were TIM, TIM and K81uf+I+G respectively. Accordingly, 39 haplotypes of the *ZBED1* gene were clustered into two well-supported clades, with one clade containing only individuals from the Southwest group, and the other one containing individuals from both the Northeast group and the Central-East group. Unlike the mtDNA result, the *ZBED1* gene of individuals from the Northeast group and the Central-East group did not differentiate but mixed with each other.

Positive selection analysis

To determine the potential adaptive evolution of the *ZBED1* gene, different models implemented in PAML were employed. In branch models, one-ratio model (M0) showed a single ω ratio (ω_0=0.01926) for the entire tree (Table 3). The free-ratio model showed high variable ω ratios among branches, of which seven branches (a-g in Fig. 4B) were considerably higher than one (Table 3),

though the LRT statistic of the free-ratio model vs. one rate was not significant with $2\Delta\ell = 60.90$ ($p = 0.924$ and df = 78). The seven branches were further designated as the foreground branches and a two-ratio branch model was then analyzed (Table 3). Among seven branches, only five branches had a significantly better fit than the one-ratio model (Table 3). Interestingly, all these five branches belonged to the Northeast group and the Central-East group. To test whether the ancestral branches of the Northeast group (branch NE) and the Central-East group (branch CE) underwent positive selection, branches NE and CE were then respectively defined as foreground branches with others being background branches, and ω ratios of these two branches were calculated to be 2.01765 and 1.48366, respectively. LRT statistic showed that the two-ratio models were not significantly better than the one-ratio model. We further used branch-site Model A to identify possible positively selected sites of NE and CE branches, while the LRT statistics of both test 1 and test 2 were not significant, the ω_{2a} and ω_{2b} ratios for the ancestral NE and CE branches were higher than one (Table 4).

Subsequently, different site-specific models (M1a, M2a, M3, M7 and M8) were further analyzed (Table 5). M2a (positive selection) identified one amino acid (324A) under positive selection at the 72.4% probability, while the LRT statistic of the M1a-M2a comparison was not significant ($p = 0.664$). M8 showed that five sites were under positive selection with ω = 1.00952, and one (324A) of the five sites showed a high probability of 93.2%. The difference between M7 and M8 was statistically significant ($2\Delta\ell = 13.70$, df = 2, $p = 0.001$). The analyses based on datamonkey web-server also supported that the amino acid 324A was under positive selection, as at least two methods, REL and FUBAR, significantly identified 324A to be under positive selection with ω >1 (Table S4). The discrete model (M3) with K=3 site classes suggested that 2.067% of sites were under positive selection with ω_2= 1.00949 and identified ten amino acid under positive selection at 100% probability. The LRT statistic of M3(K=3)-M0 comparison was significant ($2\Delta\ell = 45.62$, df=4, p< 0.001).

We used clade models (branch-site Model C) to test for potential divergent selection pressure among our three focal groups (Table 6). We first applied a multi-clade Model C assuming three partitions: Northeast group, Central-East group and Southwest group. Under this model, which we call 'Model $C_{CE\&NE}$' model, a small proportion of the data set (only 1.819%) evolved under divergent selective constraint across the three partitions, with a ω ratio less than one along the SW branches ($\omega_{2(SW)}$=0.00000), above one along the Northeast group ($\omega_{2(NE)}$=2.07667), and slightly higher than one

along the Central-East group ($\omega_{2(CE)}$= 1.07229). Then we employed Model C with either the Northeast group ('Model C_{NE}') or the Central-East group ('Model C_{CE}') as the foreground partition and all others comprised the background partition. Using the 'Model C_{NE}' model, the ω ratio of the Northeast group was estimated to be 2.83291, while under the 'Model C_{CE}' model, the ω ratio of the Central-East group was less than one ($\omega_{2(CE)}$= 0.59606). LRT statistics suggested 'Model $C_{CE\&NE}$' fits the data better compared with 'Model C_{NE}' and 'Model C_{CE}'. The comparison between 'Model $C_{CE\&NE}$' and M1a was not statistically significant, but the comparison between 'Model $C_{CE\&NE}$' and M2a_rel was statistically significant. We also compared 'Model C_{CE}' and 'Model C_{NE}' versus M1a and M2a_rel, respectively, neither of these comparisons were statistically significant, suggesting neither 'Model C_{CE}' nor 'Model C_{NE}' fit the data better than the null models. These results suggested that 'Model $C_{CE\&NE}$' was more suitable for the data, and the Northeast and Central-East groups were under positive selection.

Discussion

Mammalian hibernators experience physiological extremes, e.g. ischemia, muscle disuse, and hypothermia, that are lethal to non-hibernators but have no deleterious effects on hibernators. This may suggest the existence of underlying mechanisms that allow hibernators to withstand these physiological extremes. To date, several possible protective mechanisms have been proposed (Carey et al., 2000; Carey et al., 2003; Cerri et al., 2009; Lei et al., 2014) and among them, an increased cell proliferation appears to represent a major protective mechanism to compensate for increased apoptotic cells during hibernation (Cerri et al., 2009).

In this study, using *R. ferrumequinum* as a model system, we evaluated the expression and adaptive evolution of an important transcription factor (*ZBED1*) playing a key role in the control of cell proliferation (Matsukage et al., 2008). Gene expression analyses showed that out of the five tissues analyzed, the *ZBED1* gene only over-expressed in the brain and skeletal muscle tissue during deep hibernation when compared to pre-hibernation and post-hibernation. The tissue-specific over-expression in the brain and skeletal muscle may suggest their high level of cell proliferation. The increased expression of *ZBED1* in the brain observed in this study is consistent with the study of Cerri et al. (2009), in which an increased cell proliferation in the brain of a species of hibernating frog (*Rana esculenta L.*) is demonstrated. In addition, an increased cellular proliferation in tests in some

hibernators, e.g., ground squirrels and woodchucks, is suggested to accounting for their full sexual activity during hibernation (Wimsatt, 1969). The increased cell proliferation in the certain tissues found in these studies and our own study might suggest tissue heterogeneity, since many other specific tissues (e.g., lymphocytes, liver and hippocampus) are suggested to undergo a reduced cellular proliferation during hibernation (Shivatcheva and Hadjioloff, 1987; Kruman, 1992; Kiss et al., 2011; Popov et al., 2011). Different tissues may suffer different pressures during hibernation, and hence might need variable protective mechanisms. Given the stress that the brain suffers during hibernation, such as ischemia-induced oxidative stress and thermal stress (Carey et al., 2003; Dave et al., 2012), the increased *ZBED1* expression in the brain during deep hibernation observed in this study may be associated with cell proliferation and hence may be helpful to compensate for cell death and the onset of neurological damages. We also found an obvious over-expression of the *ZBED1* gene in skeletal muscle tissue of *R. ferrumequinum* during deep hibernation**.** Further analyses showed that the expression of the gene in the skeletal muscle is consistently higher than that of the other four tissues used in this study. The over-expression of the *ZBED1* gene in skeletal muscle tissue may imply a high level of cell proliferation. The possible cell proliferation may partly account for the lack of muscle disuse atrophy in hibernators after extended periods of disuse during hibernation. As is known, inactivity in humans, such as confined bed rest, weightlessness or limb immobilization, can lead to atrophy of skeletal muscle, loss of muscle tone and impaired strength (Harlow et al., 2001). However, in hibernating animals, such as ground squirrels and bats, though their skeletal muscle mass is reduced by 14-65% depending on muscle type and species, neither atrophy nor dysfunction is found in the skeletal muscle of these hibernators (Steffen et al., 1991; Wickler et al., 1991). The possible increase of cell proliferation in skeletal muscle could therefore help to maintain the function of skeletal muscle during hibernation.

Phylogenetic analyses of the *ZBED1* gene showed a strong genetic divergence between the Northeast group and the Southwest group of *R. ferrumequinum*, which represent hibernating and non-hibernating groups, respectively. This would imply a functional divergence between them. Positive selection analyses further showed some evidence of the adaptive evolution of the *ZBED1* gene. The branch model detected five positive selected branches, which were found to be restricted to the Northeast and the Central-East groups that are dominated by hibernating individuals. Under the site

model, several positively selected sites were detected, of which, one site, 324A, was found to be with a high probability under positive selection. The amino acid 324A is putatively located in the Ribonuclease H-like catalytic domain in the protein structure of ZBED1 (Mitchell et al., 2015). Our clade model supported the existence of divergent selection pressures of the *ZBED1* gene among the three mtDNA groups. Estimates of ω were >1 for ~2% of sites when the Northeast and the Central-East group (ω_{NE}=2.07667, ω_{CE}=1.07229) were defined as the foreground clades. This may indicate a strong positive selective pressure on the Northeast group and a slight positive selective pressure on the Central-East group. Nevertheless, in the Southwest group, we detected no evidence for positive selection. These findings remained unchanged when the *ZBED1* phylogenetic tree (Fig. 4A) was used for our positive selection analyses (data not shown). Identification of the positively selected branches and positive selective pressure on the Northeast and the Central-East group support our prediction that adaptive evolution of *ZBED1* may occur in the hibernating groups of *R. ferrumequinum*.

In this study, using *R. ferrumequinum* as a model system, we investigated the expression and adaptive evolution of the *ZBED1* gene, an important cell proliferation-related transcription factor, and found a potential role of the *ZBED1* gene in hibernation. We found an over-expression of the *ZBED1* gene only in brain and skeletal muscle, and detected some evidence of adaptive evolution of this gene in hibernating groups of *R. ferrumequinum*. Our results suggest that *ZBED1* may be a potential candidate gene that regulates cell proliferation for hibernators to face the physiological extremes of hibernation.

List of symbols and abbreviations

mtDNA = mitochondrial DNA

°C = degree Celsius

mg = milligram

ml = milliliter

µg = microgramme

µl = microlitre

µM = micromole

mm = millimeter

PCR = Polymerase Chain Reaction

ANOVA = analysis of variance

Acknowledgements

We are thankful to Hongjun Lin, Bo Luo for their great contributions to samples collection and Yuyang Hao for her help in the experiments. We are grateful to Katie Andrea Solari for her help in the language revision.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JF, YW and KS conceived and designed the experiments. YX, YW, HW, TJ, AL, XH, XY, and LS collected the samples and realized the experiments. JF contributed reagents, materials and analysis tools. YX wrote the paper. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China [grant numbers 91131003, 31270414, 31370399].

Reference

Aulagnier, S., Hutson, A. M., Spitzenberger, F., Juste, J., Karataş, A., Palmeirim, J. and Paunovic, M (2008). *Rhinolophus ferrumequinum*. The IUCN Red List of Threatened Species 2008: e.T19517A8947355.

Bielawski, J. P. and Yang, Z. (2004). A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* **59**, 121-132.

Carey, H. V., Rhoads, C. and Aw, T. (2003). Hibernation induces glutathione redox imbalance in ground squirrel intestine. *J. Comp. Physiol. B* **173**, 269-276.

Carey, H. V., Sills, N. S. and Gorham, D. A. (1999). Stress proteins in mammalian hibernation. *Am. Zool.* **39**, 825-835.

Carey, H. V., Frank, C. L. and Seifert, J. P. (2000). Hibernation induces oxidative stress and activation of NK-kappaB in ground squirrel intestine. *J. Comp. Physiol. B* **170**, 551-559.

Carey, H. V., Anderws, M. T. and Martin, S. L. (2003). Mammalian hibernation: cellular and molecular responses to depressed metabolism and low temperature. *Physiol. Rev.* **83**, 1153-1181.

Cerri, S., Bottiroli, G., Bottone, M. G., Barni, S. and Bernocchi, G. (2009). Cell proliferation and death in the brain of active and hibernating frogs. *J. Anat.* **215**, 124-131.

Chen, J., Yuan, L., Sun, M., Zhang, L. and Zhang, S. (2008). Screening of hibernation-related genes in the brain of *Rhinolophus ferrumequinum* during hibernation. *Comp. Biochem. Phys. B* **149**, 388-393.

Dave, K. R., Christian, S. L., Perez-Pinzon, M. A. and Drew, K. L. (2012). Neuroprotection: Lessons from hibernators. *Comp. Biochem. Phys. B* **162**, 1-9.

Harlow, H. J., Lohuis, T., Beck, T. D. I. and Iaizzo, P. A. (2001). Muscle strength in overwintering bears. *Nature* **409**, 997.

Hirose, F., Yamaguchi, M., Handa, H., Inomata, Y. and Matsukage, A. (1993). Novel 8-base pair sequence (Drosophila DNA replication-related element) and specific binding factor involved in the expression of Drosophila genes for DNA polymerase alpha and proliferating cell nuclear antigen. *J. Biol. Chem.* **268**, 2092-2099.

Hirose, F., Yamaguchi, M., Nishida, Y., Masutani, M., Miyazawa, H., Hanaoka, F. and Matsukage, A. (1991). Structure and expression during development of Drosophila melanogaster gene for DNA polymerase alpha. *Nucleic. Acids. Res.* **19**, 4991-4998.

Kandror, O. and Goldberg, A. L. (1997). Trigger factor is induced upon cold shock and enhances viability of Escherichia coli at low temperatures. *P. Natl. Acad. Sci. USA.* **94**, 4978-4981.

Kiss, A. J., Muir, T. J., Lee, R. E. and Costanzo, J.P. (2011). Seasonal Variation in the Hepatoproteome of the Dehydration- and Freeze-Tolerant Wood Frog, *Rana sylvatica*. *Int. J. Mol. Sci.* **12**, 8406-8414.

Kruman, I. I. (1992). Comparative analysis of cell replacement in hibernators. *Comp. Biochem. Physiol.* **101**, 11-18.

Lei, M., Dong, D., Mu, S., Pan, Y.-H. and Zhang, S. (2014). Comparison of Brain Transcriptome of the Greater Horseshoe Bats (*Rhinolophus ferrumequinum*) in Active and Torpid Episodes. *PLOS ONE* **9**, e107746.

Librado, P. and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. and Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462-2463.

Matsukage, A., Hirose, F., Yoo, M. and Yamaguchi, M. (2008). The DRE/DREF transcriptional regulatory system: a master key for cell proliferation. *BBA-Gene Regul. Mech.* **1779**, 81-89.

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S. et al. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic. Acids. Res.* **43**, D213-D221.

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K. and Pond, S. L. K. (2012). Detecting individual sites subject to episodic diversifying selection. *PLOS Genet.* **8**, e1002764.

Ohno, K., Hirose, F., Sakaguchi, K., Nishida, Y. and Matsukage, A. (1996). Transcriptional regulation of the Drosophila CycA gene by the DNA replication-related element (DRE) and DRE binding factor (DREF). *Nucleic. Acids. Res.* **24**, 3942-3946.

Phuong Thao, D. T., Ida, H., Yoshida, H. and Yamaguchi, M. (2006). Identification of the Drosophila skpA gene as a novel target of the transcription factor DREF. *Exp. Cell Res.* **312**, 3641-3650.

Pond, S. L. K. and Frost, S. D. (2005a). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**, 1208-1222.

Pond, S. L. K. and Frost, S. D. (2005b). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531-2533.

Pond, S. L. K. and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. In *Statistical methods in molecular evolution*, pp. 125-181: Springer.

Popov, V. I., Kraev, I. V., Ignat'ev, D. A. and Stewart, M. G. (2011). Suspension of mitotic activity in dentate gyrus of the hibernating ground squirrel. *Neural Plast.* **2011**, Article ID 867525, 7 pages.

Posada, D. and Crandall, K. (2005). Modeltest 3.7. *Program distributed by the author. Universidad de Vigo, Spain*.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

Ryu, J. R., Choi, T. Y., Kwon, E. J., Lee, W. H., Nishida, Y., Hayashi, Y., Matsukage, A., Yamaguchi, M. and Yoo, M. A. (1997). Transcriptional regulation of the Drosophila-raf proto-

oncogene by the DNA replication-related element (DRE)/DRE-binding factor (DREF) system. *Nucleic. Acids. Res.* **25**, 794-799.

Sawado, T., Hirose, F., Takahashi, Y., Sasaki, T., Shinomiya, T., Sakaguchi, K., Matsukage, A. and Yamaguchi, M. (1998). The DNA replication-related element (DRE)/DRE-binding factor system is a transcriptional regulator of the Drosophila *E_2F* gene. *J. Biol. Chem.* **273**, 26042-26051.

Shivatcheva, T. M. and Hadjioloff, A. I. (1987). Seasonal involution of gut-associated lymphoid tissue of the european ground squirrel. *Dev. Comp. Immunol.* **11**, 791-799.

Sonna, L. A., Fujita, J., Gaffin, S. L. and Lilly, C. M. (2002). Invited review: Effects of heat and cold stress on mammalian gene expression. *J. Appl. Physiol.* **92**, 1725-1742.

Steffen, J. M., Koebel, D. A., Musacchia, X. J. and Milsom, W. K. (1991). Morphometric and metabolic indices of disuse in muscles of hibernating ground squirrels. *Comp. Biochem. Phys. B* **99**, 815-819.

Sun, K., Luo, L., Kimball, R. T., Wei, X., Jin, L., Jiang, T., Li, G. and Feng, J. (2013). Geographic variation in the acoustic traits of greater horseshoe bats: testing the importance of drift and ecological selection in evolutionary processes. *PLOS ONE* **8**, e70368.

Takahashi, Y., Yamaguchi, M., Hirose, F., Cotterill, S., Kobayashi, J., Miyajima, S. and Matsukage, A. (1996). DNA replication-related elements cooperate to enhance promoter activity of the drosophila DNA polymerase alpha 73-kDa subunit gene. *J. Biol. Chem.* **271**, 14541-14547.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725-2729.

Tsuchiya, A., Inoue, Y. H., Ida, H., Kawase, Y., Okudaira, K., Ohno, K., Yoshida, H. and Yamaguchi, M. (2007). Transcriptional regulation of the Drosophila *rfc1* gene by the DRE-DREF pathway. *Febs J.* **274**, 1818-1832.

Wang, L. C., Lee, T., Fregley, M. and Blatteis, C. (1996). Torpor and hibernation in mammals: metabolic, physiological, and biochemical adaptations. In *Handbook of physiology: environmental physiology*, (eds. J. R. Pappenheimer, M. J. Fregly and C. M. Blatties). New York: Oxford University Press.

Weadick, C. J. and Chang, B. S. W. (2012). Complex patterns of divergence among green-sensitive (RH2a) African cichlid opsins revealed by Clade model analyses. *Bmc Evol. Biol.* **12**, 206.

Weadick, C. J. and Chang, B. S. W. (2012). An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Mol. Biol. Evol.* **29**, 1297-1300.

Wickler, S. J., Hoyt, D. F. and Van Breukelen, F. (1991). Disuse atrophy in the hibernating golden-mantled ground squirrel, *Spermophilus lateralis*. *Am. J. Physiol.-Reg. I.* **261**, R1214-R1217.

Wimsatt, W. A. (1969). Some interrelations of reproduction and hibernation in mammals. In *Dormancy and Survival* (ed A. H. W. Woolhouse). pp. 511-549. London: Cambridge University Press.

Worthington Wilmer, J. and Barratt, E. (1996). A non-lethal method of tissue sampling for genetic studies of chiropterans. *Bat Research News* **37**, 1-3.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586-1591.

Figures

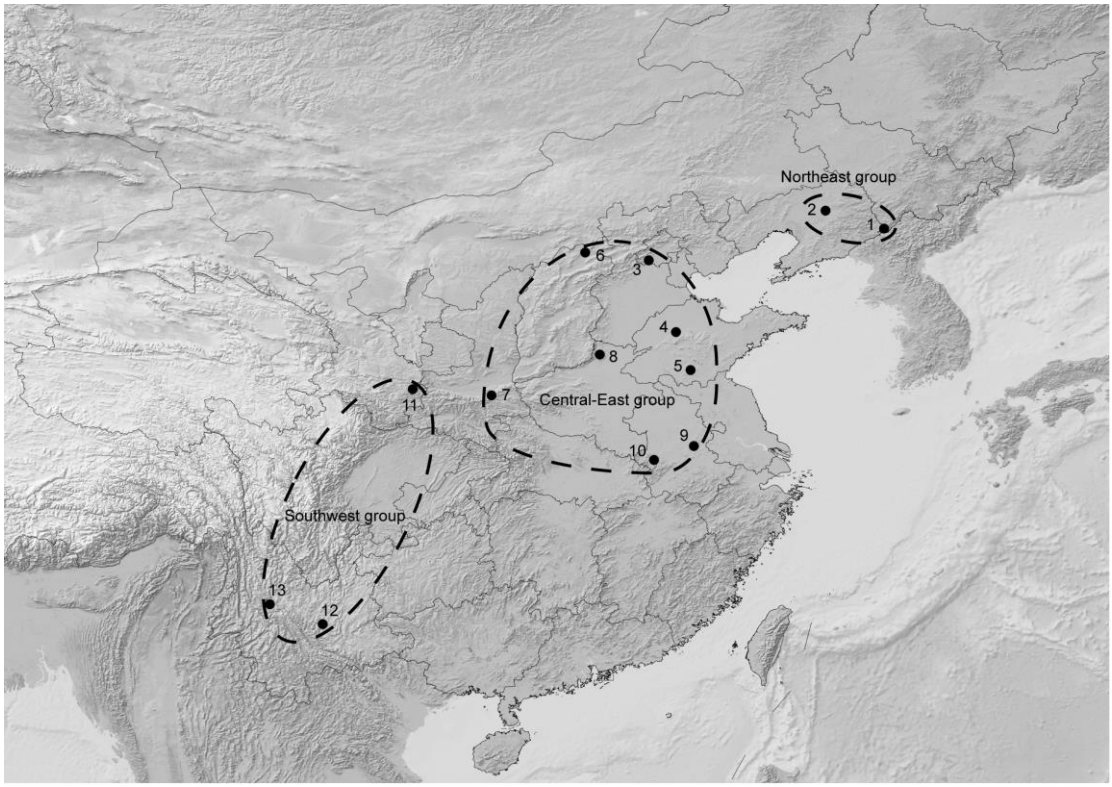

Fig. 1. **Distribution of sampling sites in China.** The corresponding three mtDNA groups are outlined by hatched lines.

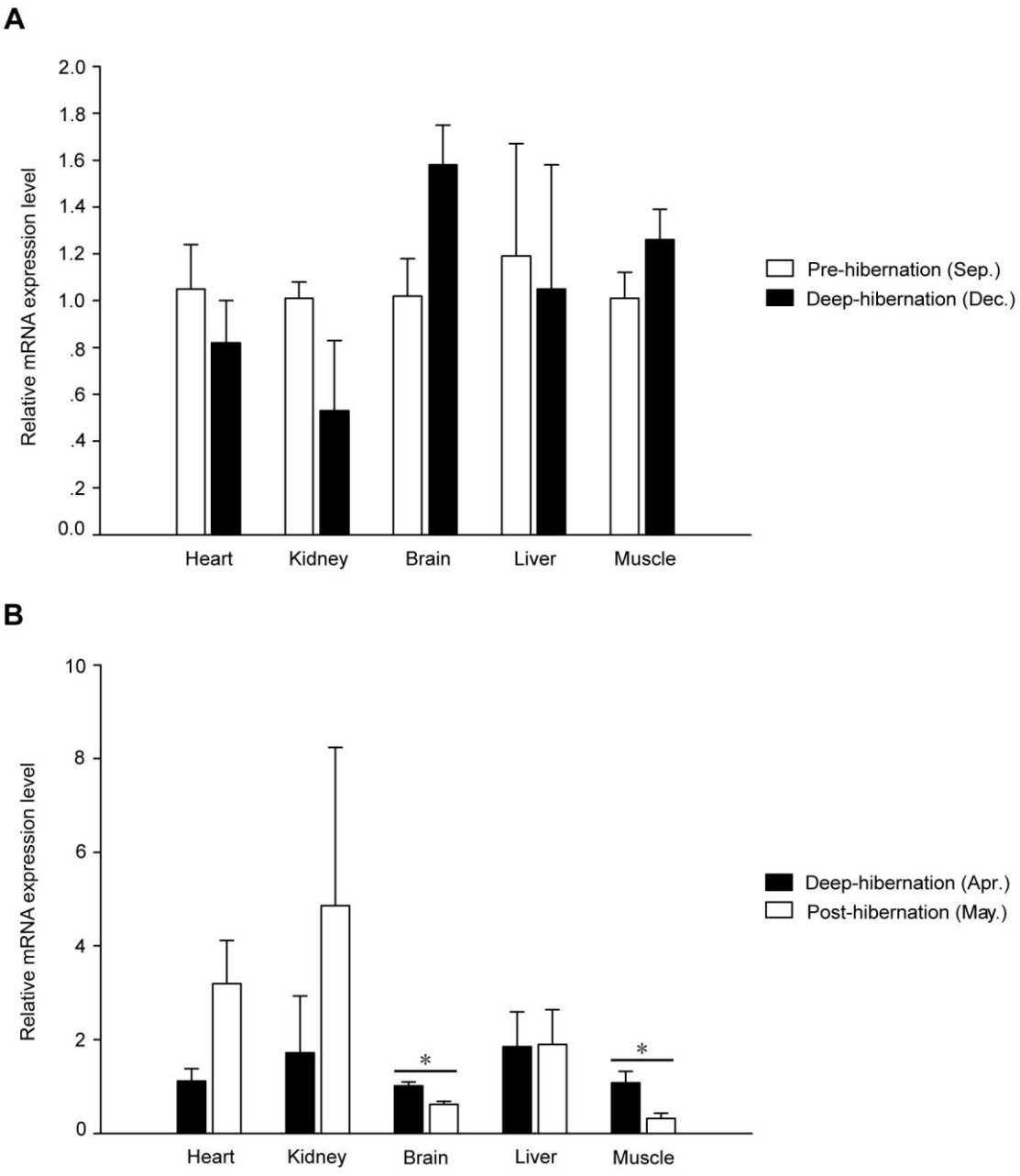

Fig. 2. **Relative quantification of the *ZBED1* gene expression in five tissues (heart, kidney, brain, liver and skeletal muscle) from pre-hibernation (September) to deep hibernation (December) (A) and from deep hibernation (April) to post-hibernation (May) (B).** The statistical significance was analyzed using Independent-Samples T Test. Asterisks indicate significant differences between the two states. Error bars in the figure are SEs.

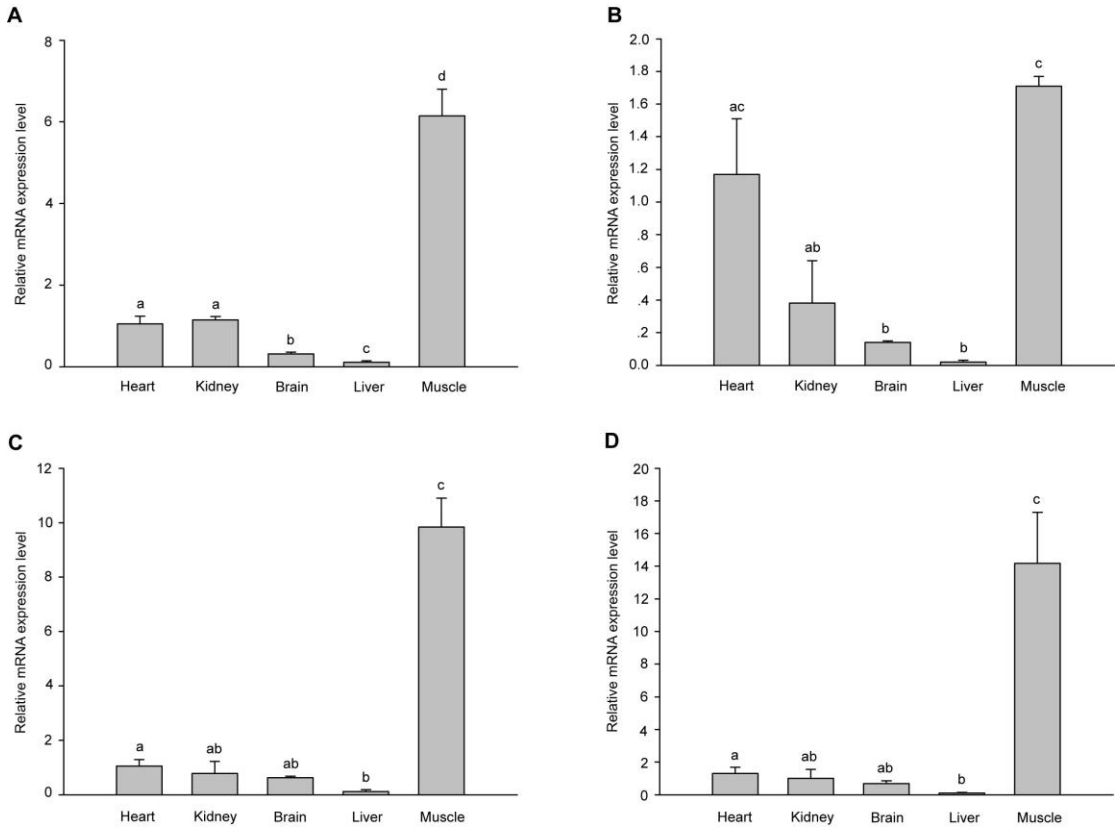

Fig. 3. **Differential expression of the *ZBED1* gene in different tissues respectively collected in September (A), May (B), December (C) and April (D).** Individual mRNA expression values are normalized to the mean value in heart, and calculated as $2^{-\Delta\Delta Ct}$. Normalized mRNA expression values for multiple individuals were averaged by tissue. Different letters indicate significant differences among tissues (Tukey's test, $p<0.05$). Error bars in the figure are SEs.

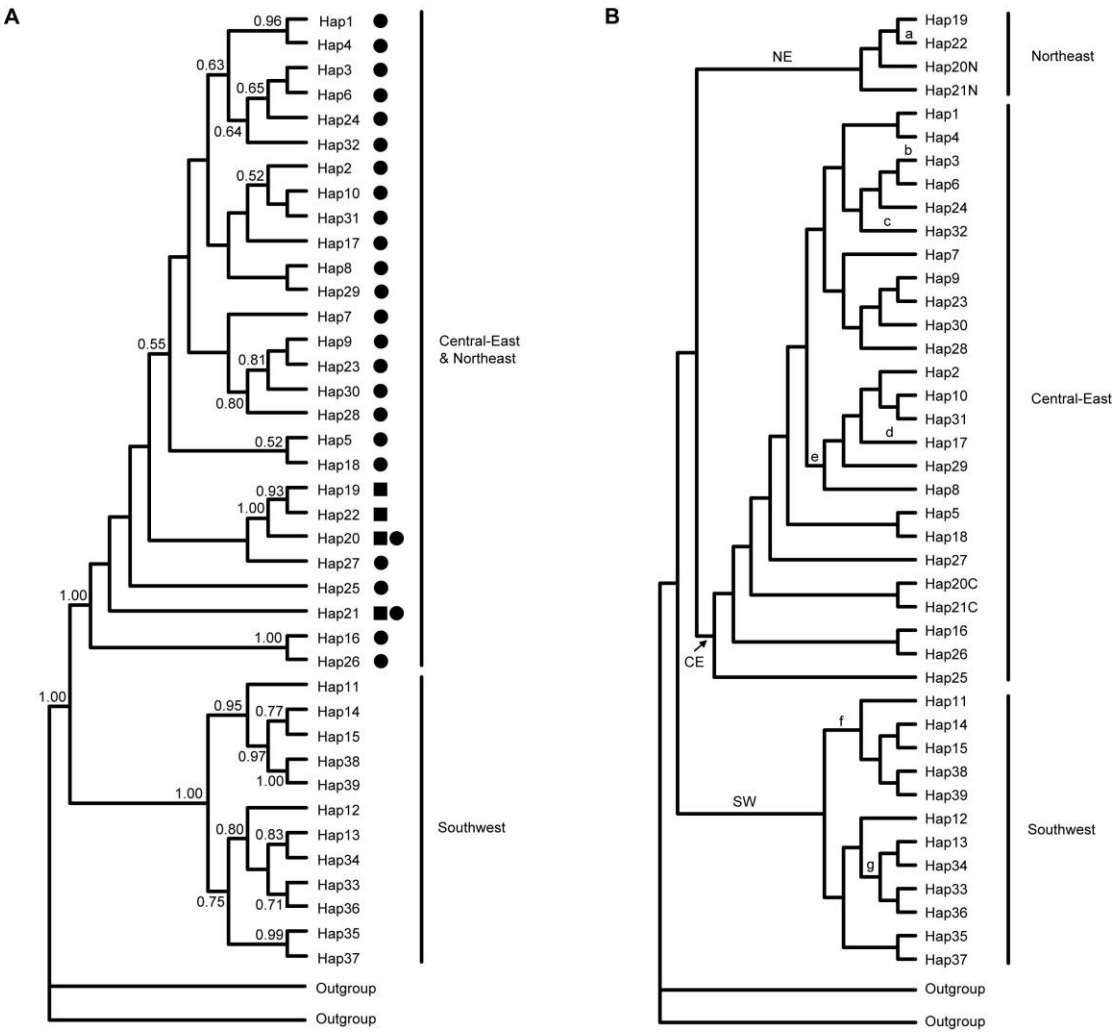

Fig. 4. **The phylogenetic tree. (A) The Bayes phylogenetic tree of the *ZBED1* gene of *R. ferrumequinum*.** The gene tree was based on complete *ZBED1* CDS. Posterior probabilities over 0.50 are shown on tree branches. Squares and circles indicate haplotypes that belong to the Northeast group and the Central-East group, respectively. The haplotypes shared between the Northeast and the Central-East groups are indicated by both squares and circles. **(B) The phylogenetic tree of *R. ferrumequinum*.** The topologic relationship among the Northeast group, the Central-East group and the Southwest group is based on previous mtDNA phylogenetic results (Sun et al. 2013). The inner relationship within each group was constructed by MrBayes 3.2.0 using *ZBED1* CDS in this study. Hap 20 and Hap 21, marked with 'N' and 'C', are haplotypes shared between the Northeast and the Central-

East groups. Different from the haplotypes from the Northeast and the Central-East groups forming one clade labeled with Central-East & Northeast in the gene tree (A), the haplotypes from the Northeast and the Central-East groups formed two clades labeled with Northeast and Central-East in the species tree (B), respectively.

Tables

Table 1. Sample information used for gene expression analyses. Sampling sites, sampling time (month) and corresponding state of bats, mean body temperature (T_b) and number of individuals are shown.

Sampling Site	**Sampling Time**	**State**	**Mean T_b ± SD**	**Number of Samples**
Ground Cave	September	Pre-hibernation (active)	34.6 ± 0.6 °C	3
Ground Cave	December	Deep hibernation (torpor)	6.9 ± 0.5 °C	3
Pigeon Cave	April	Deep hibernation (torpor)	8.4 ± 0.3 °C	5
Pigeon Cave	May	Post-hibernation (active)	32.2 ± 2.6 °C	5

Table 2. Single nucleotide polymorphism of *ZBED1* CDS among three groups, NE, CE and SW. Variations within group are listed as degenerate bases. Y: C/T, R: A/G, S: C/G, M: A/C, K: G/T. Variations that are identical between NE and CE groups but different from SW group are marked with gray colour.

Group	Sites of nucleotides															
	192	243	261	291	318	348	358	375	417	426	441	447	450	462	510	590
NE	C	G	C	C	G	G	G	Y	R	C	C	C	Y	C	Y	Y
CE	C	G	C	S	G	R	R	Y	R	Y	S	Y	Y	Y	Y	Y
SW	Y	R	Y	S	R	R	G	Y	G	C	C	C	Y	C	C	C
	609	633	654	675	683	684	693	699	714	728	735	750	759	768	774	810
NE	C	Y	G	C	C	C	T	C	C	Y	G	C	C	T	G	C
CE	C	Y	R	C	Y	C	T	Y	C	Y	G	C	S	Y	R	C
SW	Y	C	G	S	C	Y	Y	C	M	C	K	Y	C	C	G	Y
	837	855	894	934	954	964	971	979	1017	1066	1092	1161	1206	1281	1308	1335
NE	S	C	C	G	A	C	C	G	C	G	A	C	A	C	A	T
CE	R	Y	C	R	R	C	S	R	C	G	A	C	A	C	A	T
SW	G	C	Y	G	R	Y	S	G	Y	R	C	Y	G	G	R	C
	1344	1536	1548	1555	1563	1587	1590	1607	1626	1638	1671	1713	1834	1881	2085	2103
NE	C	A	R	C	G	G	C	C	A	G	C	C	C	R	C	K
CE	C	R	R	C	G	G	C	Y	A	G	Y	C	C	R	M	T
SW	G	R	G	T	R	R	S	C	G	R	C	Y	Y	R	C	T

Table 3. The positive selection analyses of the *ZBED1* gene based on branch models. The branches that were defined as the foreground branches in the two-ratio model are shown in Fig 4B.

Branch model	**Estimates of parameters**	ℓ	**Likelihood ratio test (LRT)**		
			$2\Delta\ell$	**df**	***P*-value**
M0: one-ratio	ω= 0.01926	-3430.70	-	-	-
Free-ratio	-	-3400.25	-	-	-
Branch model (Two-ratio)					
Branch a	ω_0= 0.01813, ω_1=999.00000	-3428.73	-	-	-
Branch b	ω_0= 0.01812, ω_1=999.00000	-3428.74	-	-	-
Branch c	ω_0= 0.01812, ω_1=999.00000	-3428.77	-	-	-
Branch d	ω_0= 0.01587, ω_1=999.00000	-3425.16	-	-	-
Branch e	ω_0= 0.01812, ω_1=999.00000	-3428.76	-	-	-
Branch f	ω_0= 0.01814, ω_1=999.00000	-3429.01	-	-	-
Branch g	ω_0= 0.01821, ω_1=999.00000	-3429.18	-	-	-
Branch NE	ω_0= 0.01885, ω_1=2.01765	-3431.55	-	-	-
Branch CE	ω_0= 0.01928, ω_1=1.48366	-3431.10	-	-	-
Branch model (Two-ratio, ω_1=1)					
Branch a	ω_0= 0.01811, ω_1=1.00000	-3428.84	-	-	-
Branch b	ω_0= 0.01812, ω_1=1.00000	-3428.85	-	-	-

Branch c	ω_0= 0.01812, ω_1=1.00000	-3428.87	-	-	-
Branch d	ω_0= 0.01585, ω_1=1.00000	-3425.07	-	-	-
Branch e	ω_0= 0.01812, ω_1=1.00000	-3428.87	-	-	-
Branch f	ω_0= 0.01815, ω_1=1.00000	-3429.08	-	-	-
Branch g	ω_0= 0.01821, ω_1=1.00000	-3429.26	-	-	-
Branch NE	ω_0= 0.01885, ω_1=1.00000	-3431.55	-	-	-
Branch CE	ω_0= 0.01926, ω_1=1.00000	-3430.70	-	-	-
LRT of variable ω values among branches					
Free-ratio vs. M0			60.90	78	0.924
LRT of ω at specific lineages (. two ratio vsone ratio)					
Branch a			3.94	1	0.047
Branch b			3.92	1	0.048
Branch c			3.86	1	0.049
Branch d			11.08	1	0.001
Branch e			3.88	1	0.049
Branch f			3.38	1	0.066
Branch g			3.04	1	0.081
Branch NE			1.70	1	0.192
Branch CE			0.80	1	0.371
LRT of ω at specific lineages (two ratio vs. two ratio,ω_1=1)					

Branch a	0.22	1	0.639
Branch b	0.22	1	0.639
Branch c	0.20	1	0.655
Branch d	0.18	1	0.671
Branch e	0.22	1	0.639
Branch f	0.14	1	0.708
Branch g	0.16	1	0.689
Branch NE	0.00	1	1.000
Branch CE	0.80	1	0.371

Table 4. The positive selection analyses of the *ZBED1* gene based on branch-site models. The branches that were defined as the foreground branches are shown in Fig 4B.

Branch-site model	**Estimates of parameters**	ℓ	**Test**	**Likelihood ratio test (LRT)**		
				$2\triangle\ell$	**df**	***P*-value**
M1a: Nearly Neutral	p_0=0.97925, p_1=0.02075	-3407.48	-	-	-	-
	ω_0=0.00000, ω_1=1.00000					
Model A (alternative hypothesis)	p_0=0.97924, p_1=0.02075, p_{2a}=0.00000, p_{2b}=0.00000	-3407.89	Test 1	0.82	2	>0.05
for Branch NE	Background:					
	ω_0=0.00000, ω_1=1.00000, ω_{2a}=0.00000, ω_{2b}=0.00000					
	Foreground:					
	ω_0=0.00000, ω_1=1.00000, ω_{2a}=1.00000, ω_{2b}=1.00000					
Model A (null hypothesis)	p_0=0.97925, p_1=0.02075, p_{2a}=0.00000, p_{2b}=0.00000	-3407.89	Test 2	0.00	1	>0.05
for Branch NE	Background:					
	ω_0=0.00000, ω_1=1.00000, ω_{2a}=0.00000, ω_{2b}=0.00000					
	Foreground:					
	ω_0=0.00000, ω_1=1.00000, ω_{2a}=1.00000, ω_{2b}=1.00000					
Model A (alternative hypothesis)	p_0=0.91841, p_1=0.01947, p_{2a}=0.06083, p_{2b}=0.00129	-3407.89	Test 1	0.82	2	>0.05
for Branch CE	Background:					
	ω_0=0.00000, ω_1=1.00000, ω_{2a}=0.00000, ω_{2b}=1.00000					
	Foreground:					

	ω_0=0.00000, ω_1=1.00000, ω_{2a}=3.16380, ω_{2b}=3.16380						
Model A (null hypothesis) for Branch CE	p_0=0.91672, p_1=0.01943, p_{2a}=0.06252, p_{2b}=0.00133 Background: ω_0=0.00000, ω_1=1.00000, ω_{2a}=0.00000, ω_{2b}=1.00000 Foreground: ω_0=0.00000, ω_1=1.00000, ω_{2a}=1.00000, ω_{2b}=1.00000	-3407.89	Test 2	0.00	1	>0.05

Table 5. The positive selection analyses of the *ZBED1* gene based on site models.

Model	**Estimates of parameters**	ℓ	**Likelihood ratio test (LRT)**			**Positively selected sites**
			$2\Delta\ell$	**df**	***P*-value**	
M0: one-ratio	ω=0.01926	-3430.70	-	-	-	None
M1a: Nearly Neutral (K=2)	p_0=0.97925 (p_1=0.02075) ω_0=0.00000, ω_1=1.00000	-3407.48	-	-	-	Not allowed
M2a: Positive selection (K=3)	p_0=0.97930, p_1=0.01783 (p_2=0.00287) ω_0=0.00000, ω_1=1.00000, ω_2=1.04714	-3407.89	-	-	-	324A (0.724)
M2a_rel	p_0=0.75622, p_1=0.02075 (p_2=0.22303) ω_0=0.00000, ω_1=1.00000, ω_2=0.00000	-3407.48	-	-	-	None
M3: discrete (K=3)	p_0=0.15772, p_1=0.82161 (p_2=0.02067) ω_0=0.00000, ω_1=0.00000, ω_2=1.00949	-3407.89	-	-	-	120A (1.000), 197A (1.000), 228A (1.000), 243T (1.000), 312G (1.000), 324A (1.000), 327A (1.000), 356G (1.000), 536A (1.000), 701F (1.000)
M7: beta	p=0.00945, q=0.22253	-3414.74	-	-	-	None
M8: beta&ω>1	p_0=0.97934, p=0.00500, q=2.75586, (p_1=0.02066), ω=1.00952	-3407.89	-	-	-	197A (0.535), 243T (0.528), 324A (0.932), 356G (0.525), 536A (0.534)

LRT of variable ω values among sites

M0 vs. M3(K=3)	-	45.62	4	<0.001	-
M1a vs. M2a	-	0.82	2	0.664	-
M7 vs. M8	-	13.70	2	0.001	-

Note: The numbers in the bracket following amino acid sites are posterior probability of positively selected sites. The reference sequence is Hap39.

Table 6. The positive selection analyses of the *ZBED1* gene based on clade models. The three clades (Northeast, Central-East, and Southwest) are shown in Fig 4B.

Clade model	**Estimates of parameters**	ℓ	**Likelihood ratio test (LRT)**		
			$2\Delta\ell$	**df**	***P*-value**
Model $C_{CE\&NE}$	p_0=0.97517, p_1=0.00664 (p_2=0.01819)	-3403.76	-	-	-
	ω_0=0.00000, ω_1=1.00000, $\omega_{2(SW)}$=0.00000, $\omega_{2(NE)}$=2.07667, $\omega_{2(CE)}$= 1.07229				
Model C_{CE}	p_0=0.97003, p_1=0.01458 (p_2=0.01540)	-3406.50	-	-	-
	ω_0=0.00000, ω_1=1.00000, $\omega_{2(SW-NE)}$=0.00000, $\omega_{2(CE)}$= 0.59606				
Model C_{NE}	p_0=0.97881, p_1=0.01060 (p_2=0.01059)	-3407.00	-	-	-
	ω_0=0.00000, ω_1=1.00000, $\omega_{2(SW-CE)}$=0.72245, $\omega_{2(NE)}$= 2.83291				
LRT of variable ω values among sites and branches					
Model $C_{CE\&NE}$ vs. M1a		-	7.44	4	0.114
Model $C_{CE\&NE}$ vs. M2a_rel		-	7.44	2	0.024
Model $C_{CE\&NE}$ vs. Model C_{CE}		-	5.48	1	0.019
Model $C_{CE\&NE}$ vs. Model C_{NE}		-	6.48	1	0.011
Model C_{CE} vs. M1a		-	1.96	3	0.581
Model C_{CE} vs. M2a_rel		-	1.96	1	0.162
Model C_{NE} vs. M1a		-	0.96	3	0.811
Model C_{NE} vs. M2a_rel		-	0.96	1	0.327

Supplementary information

Table S1. Primers of *ZBED1* and *β-actin* used for real time PCR.

Gene	Primer	Base sequence (5'- 3')	Tm (°C)	Products (bp)
ZBED1	Sense	TGACGTAGGTCCTGTTCTGGTT	60.0	238
	Anti-sense	CTCATCTGCGATGGGCTGT		
β-actin *	Sense	GACCTCTATGCCAACACAG	60.0	190
	Anti-sense	CATCTGCTGGAAGGTGGAC		

*Primer pair of *β-actin* was obtained in Chen et al. 2008.

Table S2. Sample information used for phylogenetic and positive selection analyses. The sampling sites and their corresponding mtDNA clades are shown. NE, Northeast group; CE, Central-East group; SW, Southwest group.

Sites No.	Sites	Evolutionary Group	Location		Samples No.
			Latitude	Longtitude	
1	Ji'an, Jilin	NE	41°03′	125°50′	JL07, JL09, JL10, JL11, JL12, JL13
2	Shenyang, Liaoning	NE	41°48′	123°23′	LN1307, LN1308, LN1309, LN1310, LN1344
3	Fangshan, Beijing	CE	39°43′	115°59′	BJ1311, BJ1312, BJ1315, BJ1316, BJ1317
4	Jinan, Shandong	CE	36°43′	117°07′	SD19, SD20
5	Feixian, Shandong	CE	35°08′	117°44′	SD1185, SD1188, SD1189
6	Datong, Shanxi	CE	40°03′	113°19′	SX0827, SX0830, SX0831
7	Lantian, Shaanxi	CE	34°04′	109°24′	SXi1216, SXi1217, SXi1218, SXi1220, SXi1221
8	Anyang, Henan	CE	35°47′	113°56′	HeN1201, HeN1202, HeN1204, HeN1208, HeN1211
9	Chaohu, Anhui	CE	31°57′	117°52′	AH0801
10	Lu'an, Anhui	CE	31°22′	116°12′	AH1002, AH1004
11	Tianshui, Gansu	SW	34°20′	106°06′	GS1005, GS1008, GS1009, GS1011
12	Kunming, Yunnan	SW	24°30′	102°20′	YN21, YN07205
13	Dali, Yunnan	SW	25°20′	100°07′	YN22, YN23, YN24

Table S3. 39 *ZBED1* haplotypes and their corresponding individuals (alleles). Different alleles from each individual are distinguished by adding 'A' or 'B' tags to their individual identifiers.

Haplotype	**Allele**	**Haplotype**	**Allele**	**Haplotype**	**Allele**
Hap 1	AH0801A, SX0827A	Hap 14	GS1011A	Hap 27	SD1189B
Hap 2	AH0801B, AH1002B, BJ1317A, HeN1208A, SD1185A, SD1188A, Sxi1218	Hap 15	GS1011B	Hap 28	SX0830A
Hap 3	AH1002A	Hap 16	HeN1202A	Hap 29	SX0831A
Hap 4	AH1004A	Hap 17	HeN1204A	Hap 30	SX0831B
Hap 5	AH1004B	Hap 18	HeN1204B	Hap 31	Sxi1216
Hap 6	BJ1311A, BJ1312A, BJ1315B, HeN1201B, Sxi1220B	Hap 19	JL07, JL09, JL11B, JL12, JL13B, LN1307, LN1308, LN1309, LN1310,LN1344	Hap 32	Sxi1220A
Hap 7	BJ1311B, BJ1317B, HeN1201A, HeN1208B, HeN1211	Hap 20	JL10, SD1189A	Hap 33	YN21A
Hap 8	BJ1312B	Hap 21	JL11A, SD19, Sxi1217	Hap 34	YN21B
Hap 9	BJ1315A, HeN1202B, Sxi1221	Hap 22	JL13A	Hap 35	YN22A
Hap 10	BJ1316	Hap 23	SD20A	Hap 36	YN22B
Hap 11	GS1005A, GS1009	Hap 24	SD20B	Hap 37	YN23, YN24
Hap 12	GS1005B	Hap 25	SD1185B	Hap 38	YN07205A
Hap 13	GS1008	Hap 26	SD1188B, SX0827B, SX0830B	Hap 39	YN07205B

Table S4. The positive selection analyses of the *ZBED1* gene based on SLAC, FEL, REL, MEME and FUBAR methods.

Codon	SLAC		FEL		REL		MEME		FUBAR	
	dN-dS	p-value	dN-dS	p-value	dN-dS	Bayes factor	ω^+	p-value	dN-dS	Post. Pr.
120	10.974	0.667	33.867	0.491	-	-	1.759	0.487	-	-
197	10.974	0.667	34.305	0.483	-	-	1.331	0.485	-	-
228	-	-	-	-	-	-	102.699	0.437	-	-
243	10.941	0.669	30.806	0.511	-	-	1.402	0.507	-	-
312	10.962	0.668	31.473	0.505	-	-	1.710	0.490	-	-
324	54.870	0.132	166.558	0.125	5.583	4.726E+06*	7.018	0.218	3.233	0.971*
327	10.974	0.667	34.068	0.292	-	-	1.825	0.280	-	-
356	10.717	0.683	29.944	0.531	-	-	1.623	0.500	-	-
536	10.974	0.667	34.292	0.483	-	-	1.794	0.484	-	-
701	10.228	0.715	37.965	0.276	-	-	1.545	0.320	-	-

*Statistically significant value.