Mitochondrion xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Mitochondrion

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

Mitogenomic analyses propose positive selection in mitochondrial genes for high-altitude adaptation in galliform birds

Taicheng Zhou [a,c], Xuejuan Shen [b], David M. Irwin [c,d], Yongyi Shen [b,c,*], Yaping Zhang [a,c,**]

[a] Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming 650091, China
[b] Joint Influenza Research Centre (SUMC/HKU), Shantou University Medical College, Shantou 515041, China
[c] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[d] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5S 1A8, Canada

ARTICLE INFO

Article history:
Received 25 February 2014
received in revised form 28 July 2014
accepted 31 July 2014
Available online xxxx

Keywords:
Mitochondrial genes
High-altitude adaptation
Positive selection
Galliform birds

ABSTRACT

Galliform birds inhabit very diverse habitats, including plateaus that are above 3000 m in altitude. At high altitude, lower temperature and hypoxia are two important factors influencing survival. Mitochondria, as the ultimate oxygen transductor, play an important role in aerobic respiration through oxidative phosphorylation (OXPHOS). We analyzed the mitochondrial genomes of six high-altitude phasianidae birds and sixteen low-altitude relatives in an attempt to determine the role of mitochondrial genes in high-altitude adaptation. We reconstructed the phylogenetic relationships of these phasianidae birds and relatives and found at least four lineages that independently occupied this high-altitude habitat. Selective analyses revealed significant evidence for positive selection in the genes *ND2*, *ND4*, and *ATP6* in three of the high-altitude lineages. This result strongly suggests that adaptive evolution of mitochondrial genes played a critical role during the independent acclimatization to high altitude by galliform birds.

1. Introduction

Galliform birds are a complex group, composed of about 290 species, which are distributed to essentially every part of the world, except the innermost deserts and polar regions. Most species inhabit areas of low altitude, although some can survive at very high altitudes, such as *Crossoptilon harmani* and *Tetraogallus tibetanus*. Cold climate, high levels of UV exposure, and hypoxia are the most important ecological factors that restrict the viability of animals that live on high plateaus. Animals that are native to the plateau, who have survived thousands of years in the highlands, must have evolved adaptive mechanisms during their evolutionary history to address these harsh environmental stresses (Scott et al., 2011a, 2011b; Xu et al., 2005).

Hypoxia is an unavoidable environmental stress for high-altitude birds (Altshuler and Dudley, 2006; Scott, 2011). The ambient partial pressure for oxygen decreases with increasing altitude, where, for example at 4000 m elevation the PO_2 of inspired air is approximately 60% of that of sea level (Beall, 2007). Although the atmosphere is oxygen-thin, high-altitude birds must sustain a high rate of metabolism to support the high energy costs of flight (Scott et al., 2009). Since there is little oxygen in the inspired air, only a low level can be carried by the bloodstream for delivery to cells and mitochondria for the production of energy (Beall, 2006; Viña, 2002). To adapt to hypoxia, every step in aerobic respiration in high-altitude birds must have experienced natural selection to increase efficiency (Scott and Milsom, 2006, 2007). The process of oxygen transport has become an area of focus in hypoxia research, especially the role of hemoglobin (Perutz, 1983; Storz and Moriyama, 2008; Storz et al., 2007, 2009; Weber, 2007). Mitochondria, the ultimate oxygen transductor, play an important role in aerobic respiration through oxidative phosphorylation (OXPHOS) (Chandel and Schumacker, 2000; Xu et al., 2007). For adaptation to high-altitude hypoxic conditions, in addition to enhancing oxygen uptake and transport, it is important to improve the efficiency of oxygen usage. Thus, the mitochondrial genome, which encodes 13 essential OXPHOS system proteins (7 subunits of the NADH dehydrogenase complex, the cytochrome b subunit of the cytochrome bc_1 complex, 3 subunits of the cytochrome c oxidase, and 2 subunits of ATP synthase), must have experienced natural selection during adaptation to hypoxic conditions at high-altitude.

In this research, we address the possible role of mtDNA in the adaption to hypoxia by galliform birds that live in a high-altitude environment. We sequenced the mitochondrial genome of two high-altitude species (*Perdix hodgsoniae* and *T. tibetanus*), and combined this new sequence data with four previously published genome sequence of galliform birds that live at high altitude (*Crossoptilon crossoptilon*, *Tetraophasis szechenyii*, *Tetraophasis obscurus*, and *Lophophorus lhuysii*),

* Correspondence to: Y. Shen, Joint Influenza Research Centre (SUMC/HKU), Shantou University Medical College, Shantou 515041, China.
** Correspondence to: Y. Zhang, Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming 650091, China.
E-mail addresses: shenyy@mail.kiz.ac.cn (Y. Shen), zhangyp@mail.kiz.ac.cn (Y. Zhang).

http://dx.doi.org/10.1016/j.mito.2014.07.012
1567-7249/

and their closest relatives that live at low altitude. This dataset was then analyzed the mtDNA genomes to examine their role in adaptation to high-altitude.

2. Material and methods

2.1. DNA samples and sequence determination

Mitochondrial genomes from two species of birds that live at high altitudes were sequenced, the genomes of *P. hodgsoniae*, which was collected from Zhongdian district (4000 m) in Yunnan Province, China, and *T. tibetanus*, collected from Lhasa district (3700 m) in Tibet.

Total genomic DNA was extracted using standard 3-step phenol/chloroform extraction methods (Sambrook et al., 1989). Segmental PCR amplification primer sets were as described previously (Nishibori et al., 2001), with additional species-specific primers being designed (Additional file 1). PCR amplifications were conducted in a 50 μl volume containing 5 μl of 10 × reaction buffer, 0.2 mM dNTPs, 0.2 μM each primer, 1.5 U Taq DNA polymerase (TaKaRa Biosystems), and approximately 50 ng template. PCR amplifications were carried out using the following parameters: 95 °C 4 min, 20 cycles of denaturation at 94 °C for 1 min, annealing at 60–50 °C (1 min; 0.5 °C/cycle), extension at 72 °C for 1 min, and finally 15 cycles of 94 °C 1 min, 50 °C 1 min, 72 °C 1 min. PCR products were cleaned using Watson RCR Purification Kits (Watson BioTechnologies, Shanghai). PCR products were sequenced at least twice in both directions on an ABI 3730 Sequencer (Applied Biosystems, Foster, CA, USA) using the ABI PRISM BigDye Terminator v3.0 sequencing kit. DNA sequences were edited using DNAstar Seqman software (DNASTAR Inc., Madison, WI, USA). Newly determined genomes were deposited into GenBank (GenBank accession numbers: KF027439, KF027440).

2.2. Sequence analysis

In addition to the mitochondrial genomes that we sequenced in this study, we collected 20 available mt genomes from GenBank. The taxonomic and geographical information for these species as well as their mitochondrial genomes GenBank accession numbers are provided in Additional file 2.

The concatenated nucleotide sequences of the 13 protein-coding genes, 22 tRNAs, 2 rRNAs and control region from the 22 galliform birds were aligned with Muscle 3.8.31 (Edgar, 2004) and refined by eye. Functional amino acid motifs were predicted using the MotifScan program in the PROSITE database of protein families and domains (http://www.expasy.org/prosite). Transmembrane and surface regions were verified using HMMTOP (v.2.0) (http://www.enzim.hu/hmmtop/). Secondary sequence structures were predicted using the consensus methods of Sspro, Sspro8, ACCpro and CONpro on the SCRATCH server (http://www.igb.uci.edu/tools/scratch/). Amino acid substitutions were tested for their possible effects on protein function with the program SIFT (Sorting Intolerant from Tolerant) (http://sift-dna.org). The cutoff value in the SIFT program is a tolerance index of ≥ 0.05.

2.3. Phylogenetic reconstruction

In contrast to the individual partitions (13 protein-coding genes, 12S rRNA, 16S rRNA, or CR), the combined sequence dataset (protein-coding, RNA and CR) showed greater power for phylogenetic inference (Shen et al., 2010), thus we pooled all of the mt genes (protein-coding, RNA and CR) to form a single dataset to establish a genomic-level phylogeny (Yu et al., 2007). Combined data (protein-coding, RNA and CR) were initially aligned using Muscle 3.8.31 (Edgar, 2004) with default parameters, with the resulting alignment subsequently adjusted manually.

Since compositional bias in the sequences among species can interfere with establishing a tree topology (Foster and Hickey, 1999; Mooers and Holmes, 2000), prior to the phylogenetic reconstruction we performed tests on the stationarity of the base composition with TREEPUZZLE 5.2 (Schmidt et al., 2002). Each gene and gene set was tested separately.

Phylogenetic trees were reconstructed from the concatenated nucleotide alignments (protein-coding, RNA and control region). We used the Anseriformes species (*Aythya americana* and *Anas platyrhynchos*) as outgroups in all of the phylogenetic analyses. ModelTest 3.7 (Posada and Buckley, 2004) was used to identify the preferred model of sequence evolution for the ML analysis, under the Akaike Information Criterion (Cowing et al., 2002). The GTR + I + G model was selected for the concatenated nucleotide sequence alignment. Since ML heuristic searches in PAUP* are very slow, we used two additional fast ML-based inference packages using 100 replicates each: RAxML (Stamatakis, 2006) and PHYML (Guindon and Gascuel, 2003). As the topologies generated by each method were identical, and differed only slightly in bootstrap values, we present just the trees with bootstrap values from PHYML.

MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) was used for Bayesian inference. Bayesian posterior probabilities (BPP) used models estimated with ModelTest 3.7 under AIC. The analysis was conducted twice with four Markov chains each. Runs were conducted for $5 * 10^6$ generations and were sampled every 100 generations. When the log-likelihood scores were found to stabilize, a consensus tree was calculated after omitting the first 25% of the trees as burn-in.

2.4. Analysis of selective pressures

The CODEML program from PAML4 (Yang, 2007) was used to analyze the selective pressures. Since the CODEML likelihood analysis is sensitive to the topology of tree, we used a phylogenetic tree that was modified according to previous studies (Wang et al., 2013; Zhao et al., 2012) as the guide tree for this analysis. Selective pressure was assessed under several models: (1) one-ratio model, which assumes identical ω values for all branches, where ω is the ratio of nonsynonymous to synonymous substitution rates; (2) free-ratio model, which assumes independent ω values for each branch, and provides a rough measure of the selective pressure on each branch; (3) two-ratio model, which two ω values are used, one for the tested branch and a second value that is equal for all other branches; (4) branch-site model, which was used to determine whether these genes have undergone positive selection on a foreground branch. Bayes Empirical Bayes (BEB) analysis was used to calculate the Bayesian posterior probability of the positively selected sites; and (5) three-ratio model, where three ω values are used, one ω value for all the deep branches, one ω value for the external branches of the low-altitude birds and one ω value for the external branches of the high-altitude birds. Then, we also tested for selection using two models on the Selection Server: MEC and M8 (http://selecton.bioinfo.tau.ac.il). Lastly, two models from the HYPHY program were performed to assess the selective pressure at synonymous sites (site-to-site synonymous rate variation) (Dimitrieva and Anisimova, 2014; Kosakovsky Pond et al., 2005; Pond and Muse, 2005): (1) MG94 × REV Nonsynonymous GDD 3 model (null), which assumes no variation in the synonymous rate (SRV) but allows three rate categories for nonsynonymous sites; (2) MG94 × REV Dual GDD 3 × 3 model (SRV-aware), where rate variation at both (dual) synonymous and nonsynonymous sites, with three rate categories for α_s and three for β_s. Values for α_s and β_s values are sampled from two given rate distributions.

3. Results

3.1. Characteristics of Galliform mitochondrial genomes

The general characteristics of the mitochondrial genomes are summarized in Table 1. The lengths of the complete mitochondrial genomes from the 22 bird species ranges from 16,586 bp to 16,836 bp, with the

Please cite this article as: Zhou, T., et al., Mitogenomic analyses propose positive selection in mitochondrial genes for high-altitude adaptation in galliform birds, Mitochondrion (2014), http://dx.doi.org/10.1016/j.mito.2014.07.012