Genetic analyses on the differentiation of the Himalayan marmot, Marmota himalayana, in the Qinghai-Tibet Plateau

Jingyan Yan

Abstract

After total genomic DNA was extracted, complete cytochrome *b* (cyt *b*) gene and 11 microsatellite loci were examined for all sampled individuals. Based on the cyt *b* sequences, phylogenetic analyses with maximum likelihood and Bayesian inference methods were conducted and a median-joining haplotype network was calculated.The divergence time between major lineages was estimated using BEAST. Genetic structure based on the microsatellite dataset were also analyzed to detected admixture between major lineages in the mixed region.

Citation: Jingyan Yan Genetic analyses on the differentiation of the Himalayan marmot, Marmota himalayana, in the Qinghai-Tibet Plateau. **protocols.io**
dx.doi.org/10.17504/protocols.io.iyxcfxn
Published: 17 Jul 2017

Protocol

Step 1.

DNA extraction

Tissues were stored in 95% ethanol immediately after collection. Total genomic DNA was extracted using the commercial kit (Ezup Column Animal Genomic DNA Purification Kit, Sangon Biotech).

Step 2.

Mitochondrial DNA amplification and sequencing

The complete sequence (1140 bp) of the cyt b gene and partial flanking segments (partial ND6 and tRNA-Glu for 5’-end, partial tRNA-Pro and tRNA-Thr for 3’-end) were amplified and sequenced using a pair of self-designed primers: CY-F (5’-ATCCTAAGCCTCCGTAAATAGGA-3’) and CY-R (5’-CAGGGAATAGTT TAAGTAAGAAATGTCA-3’) (referring to the M. himalayana mitochondrial genome under accession No. JX069958).

PCRs were carried out in a 40 μL reaction volume containing 1 μL genomic DNA (approximately 60 ng), 5.0 μL of 10×PCR buffer (with Mg2+), 2.5 μL of dNTPs (2.5 mM), 1 μL of each primer (10 μM), 2.5 units Taq DNA polymerase (5 U/μL) and 29 μL sterilized double-distilled water.

The PCR conditions used were as follows: an initial denaturation at 95 oC for 5 minutes followed by 30 cycles of denaturation at 95 oC for 45 seconds, annealing at 57 oC for 45 seconds, and extension at 72 oC for 1 minute 45 seconds, and a final extension at 72 oC for 10 minutes. PCR products were then sequenced in both directions by the Shanghai Sangon Biological Engineering Technology & Services Co., Ltd (Shanghai, China).

Step 3.

Microsatellite amplification and genotyping

All specimens were genotyped at 11 microsatellite loci, B, D, E, G, H, I, J, M, R, T, and W. (Ma H. The study of *Marmota himalayana* migrant among different families within small scale: Chinese Center for Disease Control and Prevention; 2011.) Detailed information was listed in the following table.

PCRs were conducted using fluorescently-labelled primers in a total volume of 20 μL that contained 0.5 μL genomic DNA (approximately 30 ng), 2 μL of 10×PCR buffer (with Mg2+), 1 μL of dNTPs (2.5 mM), 0.4 μL of each primer (10 μM), 1 unit Taq DNA polymerase (5 U/μL) and 15.5 μL sterilized double-distilled water.

The amplification profile was: an initial denaturation at 95 oC for 4 min followed by 40 cycles of denaturation at 95 oC for 45 sec, annealing at Tm (annealing temperature, as listed in the table) for 25 sec and extension at 72 oC for 20 sec, and a final extension at 72 oC for 2 min. Fragments were then analyzed by the Shanghai Sangon Biological Engineering Technology & Services Co., Ltd (Shanghai, China).

Locus	Accession number	Repeat motif	Primer sequence (5'→3')		AT (°C)	AN
			Forward	Reverse		
B	JQ317689	$(GT)_{16}$	TTTTTGGCTAACATAGTGGT	AGTGAAGGCTAAAGCAGAGT	53	12
D	JQ317691	$(GT)_{19}$	ATGGGGACAAACATGGGACT	CGGTTGCTATGGAGACTGGA	56	8
E	JQ317692	$(TC)_{24}$	CTTGTTCAGGATTTGGCTAT	AATGTCTTGAAAATGGTGTT	52	14
G	JQ317694	$(AG)_{20}$	ATGGCAGAGAATATAAAATGG	CTGGTGGAACTTGTTAGGAG	53	14
H	JQ317695	$(GT)_{14}$	GGAAGACCACAGAGGAACAG	CCTTGAAGAGCAAGAGCATA	54	8
I	JQ317696	$(TG)_{12}$	TAATATCCCCCAAAGAAGTA	TAGACCTTGCTGTGAAAAAT	48	13
J	JQ317697	$(TG)_{12}$	ATGGGACAGAACTCTTGATT	CCTTATAGTTTTACCTCCTCC	56	13
M	JQ317700	$(AC)_{22}$	CATTGGAAGACAGAAAATACA	CAGTCCTTTGAAACTTGAGTA	48	14
R	JQ317704	$(AC)_{11}$	ACAAAACTTCTTCGTCTC	GTCTTCCACTACTCCTCT	50	7
T	JQ317706	$(TG)_{11}$	AATAGCCAGTTCAACCTC	ATGCTAACTTCAGCAACA	53	11
W	JQ317709	$(CA)_{14}$	TTTCCACAGCAGCACTCT	GGTTCCTTACCCAGACCA	55	10

Step 4.

Phylogenetic analysis

Phylogenetic relationship of the detected haplotypes was reconstructed by maximum likelihood (ML) and Bayesian inference (BI) methods. Jmodeltest v2.1.6 and the Akaike Information Criterion (AIC) were used to select a best-fit DNA substitution model. Maximum-likelihood analyses were conducted in PhyML v3.0 from the command line (phyml -i myfile.phy -d nt -n 1 -b 1000 --run_id TIM2+G -m 010232 -f m -v e -c 4 --no_memory_check -o tlr -s BEST). The branch support values of the ML trees were estimated using non-parametric bootstrap (BSP) with 1000 replicates. The Bayesian analyses were performed using MrBayes v3.2.3 . Two simultaneous Markov Chain Monte Carlo (MCMC) analyses were run for 10 million generations, and trees were sampled every 100 generations. The first 25% of trees were discarded as ‘burn-in’, and the remaining samples were used to generate the consensus tree and the Bayesian posterior probability (BPP).Phylogenetic trees were visualized with Figtree v1.4.3.

To indicate the phylogenetic features more clearly, another two cyt *b* sequences with confirmed taxonomic identities from Genbank, one for *M. h. himalayana* (from the genome under Accession No. JX069958) and the other for *M. h. robusta* (Accession No. AF143928), were added to the final data set, as well as a homogenous sequence for *M. sibirica* (Accession No. AF143937) as an outgroup to root the phylogenetic trees.Besides, a median-joining haplotype network was calculated based on the maximum parsimony criterion with NETWORK v5.0 since this approach was suggested to be more efficient than classical phylogenetic methods for representing intraspecific evolution . Ambiguous connections were removed to show the most definite connections clearly under the criteria described in Posada and Crandall (2001) .

Step 5.

Divergence time estimation

Divergence times of the well-supported basal nodes were then estimated using a strict molecular clock method in BEAST v1.8.3 , with a constant-size coalescent tree prior and HKY + I substitution model. Three calibrations were used. The 1st was a fossil calibration regarding *Marmota minor* as the oldest marmot fossil at 10.3 mya (J. Alroy, Macquarie University, Sydney, pers. comm.; Paleobiology Database, http://www.paleodb.org, accessed 1 January 2010). Following the protocol of Steppan et al. (2011) , a prior normal distribution with a mean of 10.3 mya and a standard deviation (SD) of 0.9 mya was assigned to the most recent common ancestor (MRCA) of marmots and their sister group, the clade of *Callospermophilus lateralis*, *C. saturatus*, and *Otospermophilus beecheyi*. Another two calibrations were based on estimates of Steppan et al. (2011) . One was a normal distribution with a mean of 6.0 mya and an SD of 0.8 mya for the MRCA of subgenera Petromarmota and Marmota, and another was a normal distribution with a mean of 4.9 mya and an SD of 0.6 mya for the MRCA of subgenus Marmota. For the calibrations, cyt *b* sequences of another 16 relative species, *M. camtschatica* (AF100715), *M. baibacina* (AF143915), *M. kastschenkoi* (AF143914), *M. bobak* (AF143917), *M. menzbieri* (AF143931), *M. caudata* (AF143925), *M. marmota* (AF143929), *M. broweri* (AF143919), *M. monax* (AF143934), *M. olympus* (JF313271), *M. vancouverensis* (AF143939), *M. caligata* (AF143920), *M. flaviventris* (AF143927), *C. lateralis* (AF157887), *C. saturatus* (AF157917) and *O. Beecheyi* (AF157919), were added to the data set.The MCMC was run for 10 million generations with parameters sampled every 1000 generations. With the first 10% samples discarded as burn-in, the convergence of the stationary distribution and stability of estimated parameters were examined using TRACER v1.6 by inspection of plotted posterior estimates and the effective sample sizes of all parameters. The maximum clade credibility tree was summarized in TreeAnnotator v1.8.3 . A tree with ages for main nodes and their 95% highest posterior density intervals was displayed using FigTree v1.4.3 .

Step 6.

Genetic structure analyses using microsatellite data

A Bayesian model-based clustering method implemented in structure v2.3 was employed to detect the genetic structure based on the microsatellite dataset. To determine the optimal number of genetic clusters (K), five independent runs were performed for each K-value ranging from 1 to 20, based on the admixture model with correlated allele frequencies. The length of the MCMCs was 500 000 steps after a burn-in period of 50 000 steps. The most likely value of K was estimated with Structure Harvester using the statistic ΔK (http://taylor0.biology.ucla.edu/struct_harvest/). Membership coefficients were permuted using CLUMPP v1.1.2 . Plots of individual assignment

probabilities at the optimal value of K and several other K-values were generated with DISTRUCT v1.1
.