*Pakistan J. Zool., vol. 54(4), pp 1819-1827, 2022* **DOI**: https://dx.doi.org/10.17582/journal.pjz/20191113051142

# Novel SNP Markers for Non-invasive Samples in Tibetan Macaque (*Macaca thibetana*)

**Hua Liu, Wei Hou, Li Peng, Changjun Peng, Yansen Cai and Jing Li***

*Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, Sichuan, P.R. China*

## ABSTRACT

Owing to the characteristic of high-throughput, easy transferability between laboratories and low genotyping error, single nucleotide polymorphism (SNPs) are considered to be preferable molecular markers for genetic population analysis. SNP markers that are suitable for non-invasive samples are important to wild population investigations in endangered species. Based on the whole genome sequences of *Macaca thibetana*, we successfully developed 26 SNP loci that were sensitive to non-invasive samples, then based on which, genetic diversity and population structure of three populations of *M. thibetana* were estimated. Our results showed the novel SNP loci were polymorphic and bi-allelic. The observed and expected heterozygosity across 38 individuals varied from 0.184-0.605 and 0.405-0.506, respectively. The minor allele frequency were ranged from 0.277 to 0.487 with an average of 0.417 per locus. We detected the lowest genetic diversity in the HS population compared with the EM and FJ population. Population analysis based on Structure and neighbour-joining tree revealed a unique genetic cluster of HS population and a similar genetic background and close relatedness between the EM and FJ population. It is the first reported polymorphic SNP markers in *M. thibetana* that can be applied to non-invasive samples, and the SNPs based analysis on the population structure and genetic diversity has provided new insight into the relationships among wild populations of *M. thibetana*.

**Article Information**
**Received 13 November 2019**
**Revised 02 January 2020**
**Accepted 15 February 2020**
**Available online 24 August 2021 (early access)**
**Published 22 April 2022**

**Authors' Contribution**
**HL performed the laboratory work, analysed the data and wrote the manuscript. WH carried out the bioinformatics analyses. LP designed the primers. CP screened candidate loci. YC collected the samples and performed the experiments. JL supervised the laboratory work, conceived the study and participated in its design.**

**Key words**
***M. thibetana*, SNP markers, Non-invasive samples, Population structure, Genetic diversity**

## INTRODUCTION

The Tibetan macaque (*Macaca thibetana*), commonly considered to be endemic to China, has wild distributions in Sichuan, Yunnun, Guizhou, Hunan, Guangxi, Anhui, Zhejiang, Guangdong, Jiangxi, and Fujian provinces (Jiang *et al*., 1996; Zhong *et al*., 2012). This species is categorized as a near threatened species in the IUCN Red List and is listed in CITES in Appendix II (Wang, 1998). Based on external morphological and anatomical variation analysis, *M. thibetana* has been divided into four subspecies: *M. thibetana huangshanensis*, specifically distributed at Huangshan mountain in the east of China; *M. thibetana guizhouensis*, mainly distributed in the south and southwest of China; *M. thibetana thibetana*, mostly distributed in the southwest of China; and *M. thibetana pullus*, wildly distributed in the south and east of China (Jiang *et al*., 1996). However, continued anthropogenic interference or natural habitat isolation has resulted in a declination of the wild population size of *M. thibetana* (Li *et al*., 2014), which subsequently lead to reduction of genetic diversity and modification of population structure. To better understand the status of wild populations of *M thibetana* in China, it is necessary to develop molecular markers that were appropriate for non-invasive samples to estimate the genetic diversity and analyze population structure of *M. thibetana*.

Abundant molecular genetic markers can provide great potential for conservation research in wildlife biology (Taberlet *et al*., 1999; Bastos *et al*., 2010). Previously, mitochondrial DNA (mtDNA) has been used to investigate genetic differentiation among four subspecies (Liu *et al*., 2006; Li *et al*., 2008; Sun *et al*., 2010; Yao *et al*., 2013). Another study used microsatellite (SSR) markers to evaluate genetic diversity of the Jianyang captive population (Li *et al*., 2014), in addition, 26 polymorphic *Alu* mobile elements were identified in the Tibetan macaques and were applied to investigate genetic diversity between two Tibetan macaque populations (Guo *et al*., 2015). Compared with mtDNA, SSR, and other markers, single nucleotide polymorphism (SNPs), the third-generation markers, were considered as preferable tools to assay the genetic variations and to support science-based conservation designs (Varshney *et al*., 2007) because of their abundances in the genomes, high levels of stability

* Corresponding author: ljtjf@126.com
0030-9923/2022/0004-1819 $ 9.00/0

and polymorphism (O'Brien *et al*., 1999; Defaveri *et al*., 2013). However, so far, only 29 SNP markers have been reported in *M. thibetana* based on the transcriptome sequencing (Zhang *et al*., 2018). These SNP loci were identified from one individual and their polymorphism among populations were unsure. To date, most reported markers such as mtDNA, SSR, *Alu* and SNPs in *M thibetana* are based on blood or other tissue samples which have limited their applications for non-invasive samples from wild populations.

Many researches indicate that non-invasive samples such as shed skin, hair, saliva and feces, are necessary alternatives to tissue samples without handing, capturing or even hurting individuals, which held great promise for the wildlife genetic analysis (Taberlet *et al*., 1999). However, researchers also are aware of potential weakness of non-invasive samples such as bad DNA contamination and low DNA quantity (Waits and Paetkau, 2005; Ruell and Crooks, 2007; Norman and Spong, 2015). Null alleles, false alleles and allelic dropout thus may be likely to be common in PCR amplification using such templates and lead to error genotyping (Taberlet *et al*., 1999). Hence, it is imperative to develop SNP markers with high sensibility and accuracy for non-invasive samples so as to apply to wild populations of endangered species. Taking the advantages of whole genome resequencing data from multiple individuals of *M. thibetana,* this study aims to identify polymorphic SNP loci, from which genetic markers that are sensitive and responsible to non-invasive samples could be developed. Base on the fecal samples collected from wild populations of different geographic distribution, we also aim to investigate the genetic diversity and population structure of *M. thibetana.*

## MATERIALS AND METHODS

*Sample collections and DNA extraction*

In total, we collected 38 samples from wild *M. thibetana,* including two muscle samples, three blood samples, and 33 fecal samples. All the fecal samples were collected in 2018 from three wild populations at Emeishan-Shengtaihouqu in Sichuan Province (EM, n=11), Fanjingshan-huilongwan in Guizhou Province (FJ, n=11) as well as Huangshan-Yulinkeng in Anhui Province (HS, n=11). Sample collection sits are shown in Figure 1. These three populations are representatives of three subspecies of *M. thibetana*. Caution was taken to select fresh and individual-based feces according to its shape, color, size as well as moisture level of surface to avert sample repetition, and fecal samples split in different separate pellets to avoid and cross-individuals contamination (Yao *et al*., 2013). SSR markers together with sex markers were used for individual determination (data not shown). Blood samples provided by Administration Department of Emei Mountain Scenic Area, were taken from six injured individuals of *M. thibetana.* While two muscle samples were provided by Prof. Binhua Sun at Anhui University and from two dead individuals in 2017 and 2018, respectively. Genomic DNA was extracted using TIANamp Blood and Tissue Genomic DNA Kit or TIANamp Stool DNA Kit flowing the manufacturer protocol instructions. This study was approved by the Animal Care and Use Committee of the Institute of Laboratory Animal Science and Sichuan University Ethics Association.

*Polymorphic SNPs isolation and PCR amplification*

A dataset containing whole genome resequencing data of eight individuals of *M. thibetana*, provided by Sichuan Key Laboratory of Conservation Biology on Endangered Wildlife, was used for polymorphic SNP identification. *M. thibetana* genome was sequenced using the Illumina HiSeq X Ten platform generating ~444 million clean reads for the 8 individuals, with an average 31.09-fold genome coverage ranging from 28.41-fold to 39.09-fold (unpublished data). Due to the close relatedness between *M. thibetana* and *M. mulatta*, all the reads of eight *M. thibetana* were aligned to the *M. mulatta* reference genome rhcMac8 by bowtie software, the putative SNP loci were identified using GATK and bcftoos (De Summa *et al*., 2017; Danecek and Mccarthy, 2017). The polymorphic SNP loci were screened with the following strict criteria: (1) SNP loci were not located in repeat sequence; (2) SNP loci appeared as bi-allelic in the 8 individuals. (3) Each SNP locus showed highly polymorphism with minimum allele frequency (MAF) ranging from 0.50 to 0.75 among the 8 individuals. (4) The franking regions of SNPs (250bp both sides) were conservative for primer designs. Candidate SNPs were randomly selected for primer designs using Primer Premier 5. The primers were firstly tested by In-Silico PCR (http://genome.ucsc.edu) using rhcMac8 as reference and then validated by PCR amplification on a DNA panel from blood and muscle samples. Those primers generated one single band of expected size were further applied to PCR amplification for non-invasive samples.

The PCR amplifications were carried out in a total volume of 20μL reaction mixtures, including 40-400ng DNA, 2.5μL 10×Tag Buffer (1.5-2mM $Mg^{2+}$ plus), 40um dNTP Mix (10mM each), 80μm of each primer (10μM), 0.2U Tag DNA Polymerase, 15pmom bovine serum albumin (BSA) and dd$H_2O$. The PCR amplification procedures were conducted as follows: predenaturation at 95°C for 5 min, followed by 40 cycles of denaturation at 95°C for 30s, annealing at locus-specific temperature (54°C, 56°C, 58°Cand 60°C) for 30s and extension at 72°Cfor 40s, and final extension at 72°C for 10min. The PCR products were sequenced in both forward and reverse directions with the primers used for 2-5 times independent

PCR amplifcations on the ABI3730XL Genetic Analyser (Applied biosystems) (Sangong, Chengdu, China). Subsequently, the sequenced fragments were aligned using Vector NTI 10.3.0 (Invitrogen, Carlsbad, CA, USA), and putative SNPs were identified manually. The consensus genotypes per locus were determined when each loci were amplified in at least two independent PCR results.

*Data analysis*

*Genetic diversity:* To characterize the newly developed SNP loci, the observed (Ho) and expected (He) heterozygosities as well as minor allele frequency (MAF) were calculated using CERVUS 3.0 (Kalinowski, 2007). The PIC, the P-value for the Hardy–Weinberg equilibrium (HWE) together with inbreeding coefficient index (Fis) for each locus among loci, were estimated by GENEPOP v4.2 (François, 2008). To test its discrimination power of individual identification, probability of identity (PI), cumulative probability of identity (CDP) and the cumulative probability of identity among sibs (PIDsib) across all the samples were first calculated using GIMLET software (Valière, 2010). Further individual identification simulations were performed using CERVUS software across all the SNP and sex markers (SRY and Dead box1) (Affara *et al*., 1993; Villesen *et al*., 2006).

*Genetic structure*: To infer the population structure among EM, FJ, FJ geographical populations, an admixture-model-based clustering method for inferring population structure based on a set of allele frequency at each locus was employed using the STRUCTURE v 2.3.4 program (Pritchard *et al*., 2000). In this computational approach, it assumed that all the loci are at Hardy-Weinberg equilibrium and not significant in Linkage Disequilibrium within populations. Simulations were implemented using 10,000 Markov chain Monte Carlo (MCMC) iterations, 100,000 burn-in period and 10 independent runs for each value of K (where K may be unknown). The value of K was set from 1 to 6, the maximum number of putative subpopulations (FJ, HS, and EM) plus 3 (Evanno *et al*., 2005). Meanwhile, STRUCTURE HARVESTER (Earl and von Holdt, 2012) (http://taylor0.biology.ucla.edu/structureHarvester/#) enable us selecting the most posterior probabilities of the K values. Next, neighbour-joining tree (Saitou and Nei, 1987), a phylogenetic tree of individuals was constructed based on the Nei's standard genetic distance, (Nei, 1972) with 1000 bootstraps using POPULATION v1.2 (Takezaki *et al*., 2010) software and visualized by SPLITREE v4.1 (Huson, 1998). Owing to its highly resolved networks in the Neighbor-Net, it provides a detailed informative for genetic differentiations analysis (Bryant and Moulton, 2002).

## RESULTS

*Polymorphic SNPs isolation*

With the criteria mentioned above, a total of 169 putative SNP loci were identified from the genome resequencing data. We randomly selected 87 polymorphic SNP loci among the eight genomes to examine their potentiality as genetic markers. We designed primers for the 87 loci, 62 (72%) of them could be genotyped successfully in both blood and muscle samples. However, when they were applied to fecal samples, 36 of the 62 loci generated either no amplification result or multiple results. Whereas the remaining 26 loci (30%, Table I) showed stable and sensitive in amplification in fecal samples indicating the prospect as genetic markers for non-invasive samples.

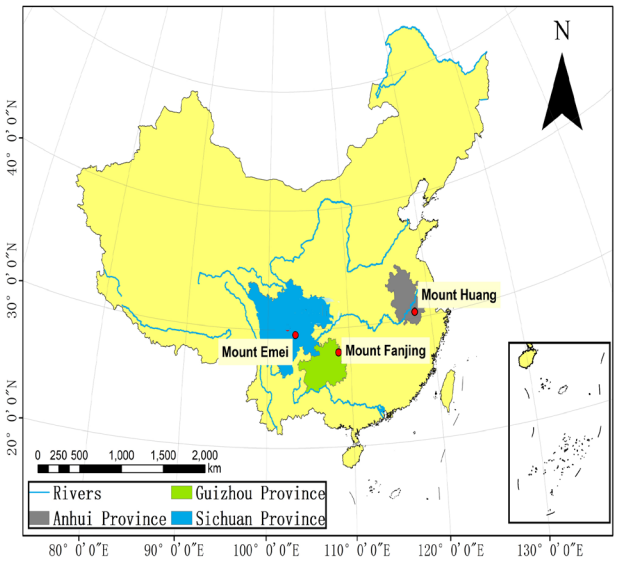


Fig. 1. The distribution of our sampling sites.

*Characterization of the novel SNP loci*

The 26 novel SNP markers were bi-alleles and highly polymorphic in *M. thibetana* individuals. Characteristics of the 26 loci in the non-invasive samples of *M. thibetana* were examined. The average minor allele frequency (MAF) were 0.417 per locus ranging from 0.2763 to 0.4868 (Table I). The observed heterozygosity (Ho) varied from 0.184 to 0.605 with an average of 0.401 per locus, and the expected heterozygosity (He) were from 0.405 to 0.506 with an average of 0.486 per locus. The Genepop analysis exhibited that the Fis across all the SNP loci were ranged from 0.1986 to 0.5915 and the HW test showed all loci were in line with Hardy-Weinberg equilibrium with P values varied from 0.1060 to 1.000 (P>0.01) after Bonferroni correction in the overall samples.

**Table I. Identification and characteristics of 26 novel SNP loci in 38 individuals of *M. thibetana*.**

| Locus name | Primer sequence (5'-3') | Tm (°C) | Fragment size (bp) | SNP | MAF | Ho | He | PIC | Fis | HWEP | Single-locus PI values | Multi-locus PI values |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mt10-3 | F:CCTACATCCCTTCTCCACTG R:CTCAACCCACATCTTCCTG | 58 | 400 | G/A | 0.487 | 0.395 | 0.506 | 0.375 | 0.223 | 0.2091 | 0.369 | $3.69\times10^{-1}$ |
| Mt15-4 | F:CCCCAGACCTAAGGGAAGT R:GGACACGGAAGGATAAATG | 56 | 280 | G/A | 0.487 | 0.605 | 0.506 | 0.375 | -0.199 | 0.3299 | 0.369 | $1.36\times10^{-1}$ |
| Mt1-1 | F:CGGGAACTCAGCCTAAGGGACC R:CACAAAACCTGGCAAGACGGGA | 56 | 296 | A/C | 0.474 | 0.421 | 0.505 | 0.374 | 0.169 | 0.3396 | 0.37 | $5.04\times10^{-2}$ |
| Mt15-1 | F:CTGATGCCAGTGGGAGTT R:CAGGGAGTGAATAAATAGAAGG | 56 | 325 | C/T | 0.474 | 0.368 | 0.505 | 0.374 | 0.274 | 0.1117 | 0.37 | $1.86\times10^{-2}$ |
| Mt15-2 | F:CAGAACCCGCTTCCATTT R:GACTGCTCCAACCCTCGT | 56 | 404 | G/A | 0.474 | 0.368 | 0.505 | 0.374 | 0.274 | 0.1164 | 0.37 | $6.89\times10^{-3}$ |
| Mt18-1 | F:GGTTTGTTCCCTCAGCAT R:GCAATCACCCGACTTTAT | 56 | 496 | C/T | 0.474 | 0.526 | 0.505 | 0.374 | -0.042 | 1.000 | 0.37 | $2.55\times10^{-3}$ |
| Mt10-1 | F: CACTCAAGGCAGTAATCAGC R:TGTGGTCCGACTTCATCC | 58 | 290 | G/A | 0.474 | 0.474 | 0.505 | 0.374 | 0.063 | 0.7525 | 0.37 | $9.41\times10^{-4}$ |
| Mt13-1 | F:TGTGGCAACATAGACCCG R:GCCCTGCTAAGAAGACAAAA | 56 | 433 | A/T | 0.474 | 0.263 | 0.505 | 0.374 | 0.483 | 0.0644 | 0.37 | $3.48\times10^{-4}$ |
| Mt20-2 | F:TTGGGTATCACAGTTAGGG R:CTCTGCGTAGCAACAAGC | 56 | 278 | G/A | 0.474 | 0.505 | 0.374 | 0.633 | 0.752 | 0.1286 | 0.37 | $1.29\times10^{-4}$ |
| Mt8-1 | F:GGGGTCCTGTGGTTCTGA R:AGCCGCAAGCGATGGTAG | 56 | 273 | C/T | 0.461 | 0.447 | 0.504 | 0.373 | 0.113 | 0.5271 | 0.37 | $4.76\times10^{-5}$ |
| Mt9-2 | F:AGAGTCAATACTGAACCCAACC R:AAGGCATTGCTCTGCTGATA | 60 | 470 | G/A | 0.461 | 0.395 | 0.504 | 0.373 | 0.218 | 0.199 | 0.37 | $1.76\times10^{-5}$ |
| Mt3-1 | F:GGAACAGCCGTGAAGTGA R:CCAGATAGACCAGCCAGAAT | 56 | 272 | T/C | 0.434 | 0.5.00 | 0.498 | 0.371 | -0.004 | 1.000 | 0.373 | $6.57\times10^{-6}$ |
| Mt10-2 | F:AAAGCACCAAGACCTCAA R:AATCACAAACTTATCCACGA | 56 | 250 | C/T | 0.421 | 0.421 | 0.494 | 0.369 | 0.149 | 0.5007 | 0.374 | $2.46\times10^{-6}$ |
| Mt16-2 | F:GACAGCATCATTGGGAGAC R:AAGCCAGTGAGCAGAGTG | 56 | 345 | G/A | 0.421 | 0.474 | 0.494 | 0.369 | 0.042 | 1.000 | 0.374 | $9.19\times10^{-7}$ |
| Mt4-1 | F:GCTGCTATCTCAGTCTCC R:ACATCCTGCTATGCTTTC | 56 | 248 | A/G | 0.408 | 0.289 | 0.489 | 0.366 | 0.412 | 0.0186 | 0.376 | $3.46\times10^{-7}$ |
| Mt7-1 | F:CCCCAGATTCCCTAACTT R:TTTGTGGTGACCTGCTACT | 56 | 262 | G/A | 0.408 | 0.237 | 0.489 | 0.366 | 0.520 | 0.019 | 0.376 | $1.30\times10^{-7}$ |
| Mt12-1 | F:TCTCCTACTCTTACCCTCAA R:CTGGGCTCCTTCTGTCTC | 56 | 264 | T/G | 0.395 | 0.263 | 0.484 | 0.364 | 0.460 | 0.0638 | 0.378 | $4.92\times10^{-8}$ |
| Mt3-2 | F:CAAATCAGTCAATCCCAGAA R:CCCACTGCCTTCGTTTAT | 56 | 346 | A/T | 0.382 | 0.395 | 0.478 | 0.361 | 0.177 | 0.3202 | 0.381 | $1.87\times10^{-8}$ |
| Mt3-3 | F:GCCAGGCACTTGAACACT R:GCCTTTGAAGAAGCACCA | 58 | 343 | C/T | 0.382 | 0.5 | 0.478 | 0.361 | -0.046 | 1.000 | 0.381 | $7.14\times10^{-9}$ |
| Mt2-1 | F:GTGTGAGGGGAGTGGTAT R:GAGGCTGTGGAAAGTGTA | 56 | 374 | G/A | 0.369 | 0.526 | 0.472 | 0.357 | -0.118 | 0.5038 | 0.384 | $2.75\times10^{-9}$ |
| Mt9-1 | F:CATTTCCGCACCTTGAGA R:CTGGAGTTGAACGCCCTA | 56 | 248 | A/G | 0.356 | 0.289 | 0.464 | 0.353 | 0.380 | 0.0334 | 0.388 | $1.07\times10^{-9}$ |
| Mt6-1 | F:GAACAGGGTCACAAGAGC R:AGTGGACTGGTGGCTAAA | 56 | 284 | G/A | 0.356 | 0.447 | 0.464 | 0.353 | 0.037 | 1.000 | 0.388 | $4.14\times10^{-10}$ |
| Mt15-3 | F:AGCCCAACACTACCCTCG R:CGTGCTGTCTACCCACTTCT | 56 | 349 | C/T | 0.356 | 0.553 | 0.464 | 0.353 | -0.194 | 0.2993 | 0.388 | $1.61\times10^{-10}$ |
| Mt11-1 | F:GGATCTCCAGCTCTGCGTTTG R:TGTGCCACCACCCACCTTC | 56 | 309 | G/A | 0.342 | 0.368 | 0.456 | 0.349 | 0.194 | 0.2852 | 0.392 | $6.30\times10^{-11}$ |
| Mt16-1 | F:GGAAAGTCAGGTGGAAACA R:CACAGGCTTACCAGAGGC | 56 | 359 | C/T | 0.329 | 0.184 | 0.447 | 0.344 | 0.592 | 0.0106 | 0.397 | $2.50\times10^{-11}$ |
| Mt7-2 | F:GGAGCCTTACACTTCACC R:TCGTATTCTCAGCAACCC | 56 | 298 | C/T | 0.277 | 0.237 | 0.405 | 0.32 | 0.420 | 0.0160 | 0.425 | $1.06\times10^{-11}$ |

The anneal temperature (*Tm*), minor allele frequency (*MAF*); observed (*H*o) and expected (*H*e) heterozygosity; polymorphic information content (*PIC*); inbreeding coefcient (*Fis*); and P-value for Hardy–Weinberg equilibrium tests (*HWEP*).

**Table II. The simulation of individual identification of 38 individuals based on different number of SNP loci.**

| Numbers of loci | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|
| PID(sib) | $4.41\times10^{-2}$ | $2.62\times10^{-2}$ | $1.56\times10^{-2}$ | $9.27\times10^{-3}$ | $5.52\times10^{-3}$ | $3.29\times10^{-3}$ | $1.97\times10^{-3}$ |
| PI | $2.55\times10^{-3}$ | $9.4\times10^{-4}$ | $3.48\times10^{-4}$ | $1.29\times10^{-4}$ | $4.76\times10^{-5}$ | $1.76\times10^{-5}$ | $6.57\times10^{-6}$ |
| Individual simulations | 34 | 36 | 37 | 38 | 38 | 38 | 38 |

In addition, the individual simulation results suggested that the cumulative identity probability (PI) of the 26 SNP loci was $1.06\times10^{-11}$, and the most informative 9 loci were enough for discriminating 38 individuals with PI value $1.29\times10^{-4}$ (Table II).

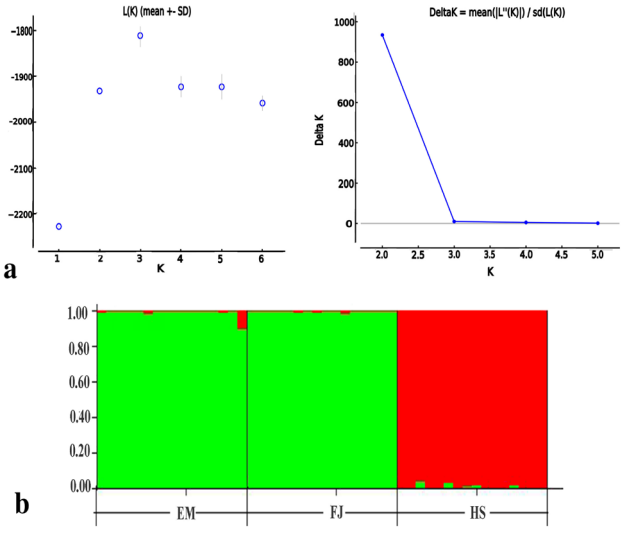


Fig. 2. Genetic structure analysis of *M. thibetana* based on 26 SNP loci. (a) Estimation number of genetic clusters (K) from 10 independent runs for K=1-6. The left panel shows the rate of change (ΔK). The right panel shows the mean of test Ln probability of data. K=2 shows the most posterior probabilities of the K values (b) The colors indicated the likely proportion of inferred clusters that each individual may assigned to. The EM and FJ individuals were assigned the similar genetic clusters, while HS individuals assigned the other distinct cluster. The sampling areas of the individuals are shown at the bottom. HS, Husangshan-Yulinkeng; EM, Emei mountain-Shengtai houqu; FJ, Fanjingshan-huilongwan.

*Genetic diversity analysis*

We assessed genetic diversity level among the three wild populations of *M. thibetana* based on newly identified polymorphic SNP loci. Statistical tests showed the mean Ho of EM, FJ and HS populations were 0.360, 0.531 and 0.315, respectively; The mean He of EM, FJ and HS populations were 0.473, 0.453 and 0.327, respectively. Moreover, the average polymorphic information content (PIC) of three populations were 0.326, 0.333 and 0.250 (Table III), indicating a moderate high level of genetic diversity in the wild populations (0.25<PIC<0.5).

*Population structure inferred from SNP data*

SNP based population structure analysis showed consistent results over ten repeated runs for each tested K values (1-6) (data not shown). K=2, where ΔK reached the maximum value, indicated the likely presence of two genetic clusters (Fig. 2). The EM and FJ individuals shared a same cluster, while the HS individuals were assigned to another unique genetic cluster. To exclude interference resulting from samples of other groups, we also ran separate structure analysis on the each subset of item pool (EM, FJ or HS population), which generated similar results as shown in Figure 2 (data not shown) suggesting a close relatedness between the EM and FJ population. Furthermore, neighbour-joining tree based on Nei's standard genetic distance provided a visualization of genetic differentiation among populations (Fig. 3). The partitioning of HS individuals into a unique cluster was consistent with the results of the structure analysis. However, the 11 individuals from the EM population were separately clustered with individuals from the FJ population forming a cluster distinct from the HS population.

**Table III. The genetic diversity among three populations of *M. thibetana*.**

| Sampling sites | Ho | He | PIC |
|---|---|---|---|
| EM | 0.360 | 0.473 | 0.326 |
| FJ | 0.531 | 0.453 | 0.333 |
| HS | 0.315 | 0.327 | 0.250 |
| Overall | 0.401 | 0.486 | 0.364 |

## DISCUSSION

*Development of polymorphic SNP loci for non-invasive samples*

Given the development of next-generation sequencing technology, the availability of whole genome sequences of the *M. thibetana* enables us to screen polymorphic SNP markers more rapidly and more efficiently. Previously, our lab reported the first genome of *M. thibetana* from one individual and identified about 11.9 million single nucleotide variants compared with the genome of *M. mulatta* (Fan *et al*., 2014). In this study, genome sequences

of different individuals have provided a large database of SNPs, by which the minimum allele frequency of each SNP locus could be estimated to identify SNP loci with high polymorphism. Using strict screening criteria, we successfully identified 169 SNP loci that were bi-allelic and highly polymorphic in the eight individuals, which will greatly benefit development of novel polymorphic SNP markers in *M. thibetana*.

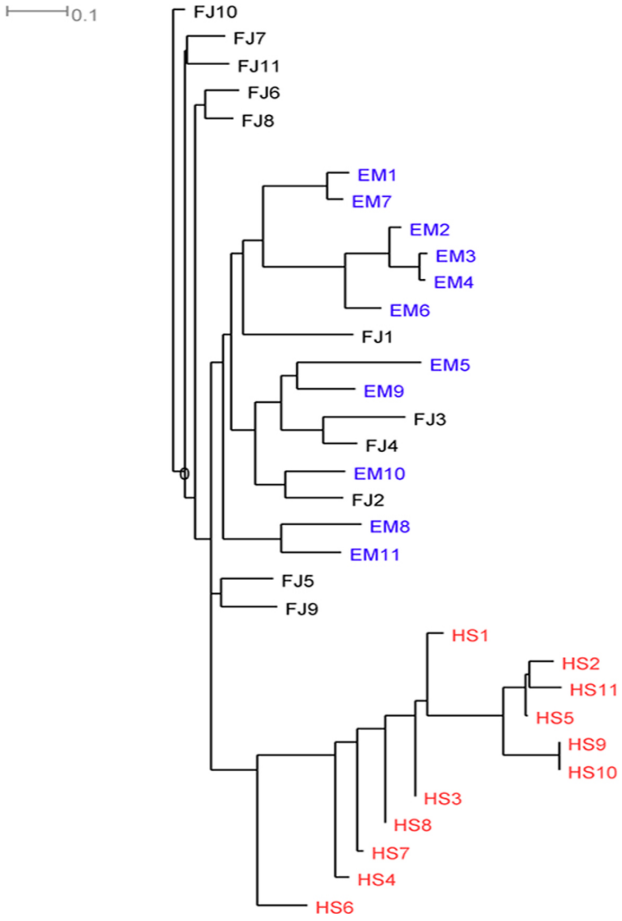


Fig. 3. Neighbour-joining tree of 33 *M. thibetana* individuals constructed based on 26 SNP loci. The color of samples were associated with the geographical origins. HS, Husangshan-Yulinkeng; EM, Emei mountain-Shengtai houqu; FJ, Fanjingshan-huilongwan.

In the genetic analysis for the wild populations, non-invasive sampling provided an alternative approach for research and management conservation in wildlife biology (Taberlet *et al*., 1999). Researchers can extracted DNA from a variety of discarded sources including shed feather, hair and feces. Nevertheless, DNA extracted from non-invasive samples have serious limitations for PCR amplification resulting from low DNA quality, or poor DNA quantity with plant-containing DNA and microbe-containing DNA (Mondol *et al*., 2009). Considering the practicability of non-invasive samples, a feasibility analysis of the newly identified SNP loci for the fecal samples was conducted. Although 72% of examined SNP loci could be successfully amplified in all tissue samples, most of them were excluded since their amplification success rate were less than 95% in the non-invasive samples or they generated inconsistent genotypes in three independent PCR results. Zhang *et al*. (2018) reported 29 polymorphic SNP markers in *M thibetana* based on blood, their applicability to non-invasive samples are unsure. Finally, we identified 26 SNP loci that showed a moderate polymorphism (0.25<PIC<0.05) in 38 *M thibetana* individuals, moreover, they had strong stability and sensitivity for all non-invasive samples. In addition, the newly developed SNP loci showed a good ability for individual discrimination. According to Hara *et al*., (2010), the subset of SNP markers with PI values ranging from $3\times10^{-6}$ to $1\times10^{-7}$ was enough for individual identification. The 26 SNP loci had a cumulative PI of $1.06\times10^{-11}$ and only 9 loci were enough for discriminating all the individuals. Thus, the newly identified SNP loci were effective markers that could be applied to further genetic study on the wild population of *M. thibetana*.

*Genetic diversity and population structure inferred from SNP data*

Evaluation of genetic diversity was significant and important for the species conservation (Avise and Hamrick, 1996). In the present study, we used the observed Heterozygosity and expected Heterozygosity to evaluate genetic diversity levels. The higher the heterozygosity is, the richer genetic variation and the stronger ability of the species adapt to the environment. Our results shows that the Ho and the He of the wild *M. thibetana* are 0.401 and 486, respectively (Table III), which are higher than those of a captive population non-human primate breeding centre of the Institute of Laboratory Animal Science, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital (Chengdu, China) based on 29 SNP markers with Ho and He of 0.331 and 0.332, respectively (Zhang *et al*., 2018). It may suggest a higher genetic diversity level in the wild *M. thibetana* than the captive population, although comparison of genetic diversity parameters based on different loci could produce biased assessment if the higher polymorphic loci were included for analysis.

Similar to other primates, *M. thibetana* has now faced a severe threat of survive resulting from human exploitation and natural habitat fragmentation (Roos *et al*., 2014; Jiang *et al*., 1996). It is wildly recognized that continued population isolation caused by anthropogenic or natural habitat fragmentation can lead to a decline of the total effective population size, reduction of genetic diversity

and modification of genetic structure among populations (Gaines *et al*., 1997; Lemer and Planes, 2014). However, besides several reports about genetic differentiation among *M. thibetana* populations based on the mtDNA sequences (Sun *et al*., 2010; Li *et al*., 2008), very little information has been known about genetic diversity and population structure among different geographic populations. In this study, we collected feces from three geographical populations of *M. thibetana* in China, and they are representatives of three different subspecies according to the external morphological characteristics or mitochondrial markers (Jiang *et al*., 1996; Yao *et al*., 2013). Overall, we detected a moderate level of genetic diversity in the three populations, however HS population had the lowest level in all *Ho, He* and PIC indicating the lowest genetic diversity of HS population compared with the EM and FJ population. On the other hand, genetic diversity levels in the EM and the FJ population are similar. Our results about genetic diversity of *M. thibetana* based on SNP markers are congruent with that based on mtDNA variations (Sun *et al*., 2010).

Furthermore, two distinct genetic clusters have been detected among the three populations including one from EM+FJ population and one from HS population, respectively. Moreover, in the NJ tree, individuals from the EM population were separately clustered with those from FJ population forming a cluster distinct from the HS population. To our surprises, our results indicate that there is no significant genetic differentiation between the EM and the FJ population, which is inconsistent with previous study based on the mtDNA variations (Sun *et al*., 2010). Their results suggested that the EM Tibetan macaques had distinct genetic background from other geographical populations including Guizhou macaques. Several reasons may contribute to the differences on the population structure of *M. thibetana.* From one hand, both the EM and FJ population are located in the southwest of China compared with eastern population of HS, which may result in close genetic relatedness between the two western populations. As suggested in Yao *et al*. (2013) study, eastern populations of *M. thibetana* had distinct genetic background from the western populations. On the other hand, although EM Tibetan macaques and Guizhou Tibetan macaques belong to different subspecies, there is possible to exist subpopulations within each of the subspecies that have different genetic background. At present, the fragmentation of habitat in these Tibetan macaque populations makes it impossible to gene flow among them, however, genetic communication might have occurred between some subpopulations of the subspecies in the evolutionary history. In this case, more populations from the two subspecies should be investigated in future to confirm the hypothesis. At last, differences between genetic markers of mtDNA variations and SNP polymorphism, exhibiting different genetic information and different mutation rates, also may lead to incongruent relationship in population genetic inference (Weiran *et al*., 2015; Wang *et al*., 2017).

## CONCLUSIONS

In conclusion, we have developed 26 highly polymorphic SNP markers based on the genome resequencing data of *M. thibetana.* These markers are stable and sensitive to non-invasive samples, which has provided powerful tools for population genetic analysis on the wild populations of *M. thibetana*. Based on the polymorphism of the SNPs among populations of *M. thibetana,* we found a low level of genetic diversity in the HS population. Furthermore, no population differentiation has been detected between the EM population and the FJ population although they are categoried as different subspecies of *M. thibetana*. The results are interesting and worthy of further investigations on the more wild populations of *M. thibetana*.

## ACKNOWLEDGEMENTS

This work was supported by the State Key Program of National Natural Science Foundation of China (Grants No. 31530068), and partly by the National Natural Science Foundation of China (Grants No. 31770415).

*Statement of conflict of interest*

The authors have declared no conflict of interest.

## REFERENCES

Affara, N.A., Chalmers, I.J. and Fergusonsmith, M.A., 1993. Analysis of the SRY gene in 22 sex-reversed XY females identifies four new point mutations in the conserved DNA binding domain. *Hum. mol. Genet*., **2**: 785-789. https://doi.org/10.1093/hmg/2.6.785

Avise, J.C. and Hamrick, J.L., 1996. Conservation genetics: Case histories from nature. *J. appl. Ecol*., **78**: 471−501. https://doi.org/10.1007/978-1-4757-2504-9

Bastos, H.B., Gonçalves, E.C., Ferrari, S.F., Silva, A. and Schneider, M.P.C., 2010. Genetic structure of red-handed howler monkey populations in the fragmented landscape of Eastern Brazilian Amazonia. *Genet. mol. Biol*., **33**: 774-780. https://doi.org/10.1590/S1415-47572010000400027

Blair, M.E., Eleanor, J.S. and Martha, H., 2011. Taxonomy and conservation of vietnam\"s primates: a review. *Am. J. Primatol*., **73**: 1093-

1106. https://doi.org/10.1002/ajp.20986

Bryant, D. and Moulton, V., 2002. *Neighbor Net: An agglomerative method for the construction of planar phylogenetic networks*. International Workshop on Algorithms in Bioinformatics. https://doi.org/10.1007/3-540-45784-4_28

Defaveri, J., Viitaniemi, H., Leder, E. and Merilä, J., 2013. Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. *Mol. Ecol. Resour*., **13**: 377–392. https://doi.org/10.1111/1755-0998.12071

Danecek, P. and Mccarthy, S.A., 2017. Bcftools/csq: Haplotype-aware variant consequences. *Bioinformatics*, **33**: 2037. https://doi.org/10.1093/bioinformatics/btx100

De Summa, S., Malerba, G., Pinto, R., Mori, A., Mijatovic, V. and Tommasi, S., 2017. GATK hard filtering: tunable parameters to improve variant calling for next generation sequencing targeted gene panel data. *BMC Bioinfo.*, **18**: 119. https://doi.org/10.1186/s12859-017-1537-8

Earl, D.A. and Vonholdt, B.M., 2012. Structure harvester: A website and program for visualizing structure output and implementing the Evanno method. *Conserv. Genet. Resour*., **4**: 359-361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G.S., Regnaut, S.J. and Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol*., **14**: 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fan, Z.X., Zhao, G., Li, P., Osada, N., Xing, J., Yi, Y., Du, Du., Silva, P., Wang, H., Ryuichi Sakate, Zhang, X.Y., Xu, H. and Li, J., 2014. Whole-genome sequencing of tibetan macaque (*Macaca thibetana*) provides new insight into the macaque evolutionary history. *Mol. Biol. Evol*., **31**: 1475-1489. https://doi.org/10.1093/molbev/msu104

François, R., 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour*., **8**: 103-106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Gaines, M.S., Diffendorfer, J.E., Tamarin, R.H. and Whittam, T.S., 1997. The effects of habitat fragmentation on the genetic structure of small mammal populations. *J. Hered*., **88**: 294-304. https://doi.org/10.1093/oxfordjournals.jhered.a023107

Guo, H., Jiang, J., Cui, Y., Yi, Y., Jia, X., Wang, H., Yue, B.S. and Li, J., 2015. Identification and characterization of polymorphicaluinsertions in the tibetan macaque (*Macaca thibetana*). *Eur. J. Wildl. Res*., **61**: 143-149. https://doi.org/10.1007/s10344-014-0887-z

Hara, K., Watabe, H., Sasazaki, S., Mukai, F. and Mannen, H., 2010. Development of SNP markers for individual identification and parentage test in a Japanese black cattle population. *Anim. Sci. J*., **81**: 152-157. https://doi.org/10.1111/j.1740-0929.2009.00720.x

Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, **14**: 68-73. https://doi.org/10.1093/bioinformatics/14.1.68

Jia, X.D., Yang, B.D., Yue, B.S., Yin, H.L., Wang, H.X. and Zhang, X.Y., 2011. Isolation and characterization of twenty-one polymorphic microsatellite loci in the thibetan macaque (*Macaca thibetana*). *Russ. J. Genet*., **47**: 884-887. https://doi.org/10.1134/S1022795411070088

Jiang, X.L., Wang, Y.X. and Wang, Q.S., 1996. Taxonomy and distribution of Tibetan macaque. *Zool. Res*., **17**: 361-369.

Kalinowski, S.T., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol*., **16**: 1099-1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Lalitha, S., 2000. Primer premier 5. *Biotech. Softw. Inter. Rep*., **1**: 270-272. https://doi.org/10.1089/152791600459894

Lemer, S. and Planes, S., 2014. Effects of habitat fragmentation on the genetic structure and connectivity of the black-lipped pearl oysterpinctada margaritiferapopulations in french polynesia. *Mar. Biol*., **161**: 2035-2049. https://doi.org/10.1007/s00227-014-2484-9

Li, D., Fan, L., Li, D., Ran, J.H., Yin, H.L., Wang, H.X., Wu, S.B. and Yue, BS., 2008. Genetic diversity analysis of *Macaca thibetana* based on mitochondrial DNA control region sequences. *DNA Sequence*, **19**: 446-452. https://doi.org/10.1080/19401730802449196

Li, P., Yang, C., Zhang, X., Li, J., Yi, Y. and Yue, B.S., 2014. Development of eighteen tetranucleotide microsatellite markers in Tibetan macaque (*Macaca thibetana*) and genetic diversity analysis of captive population. *Biochem. Syst. Ecol*., **57**: 293-296. https://doi.org/10.1016/j.bse.2014.08.015

Liu, Y., Li, J. and Zhao, J., 2006. Sequence variation of mitochondrial DNA control region and population genetic diversity of Tibetan macaques (*Macaca thibetana*) in the Huangshan Mountain. *Acta Zool. Sin*., **52**: 724-730.

Mondol, S., Karanth, K.U., Kumar, N.S., Gopalaswamy, A.M., Andheria, A. and Ramakrishnan, U., 2009. Evaluation of non-invasive genetic sampling methods for estimating tiger population size. *Biol. Conserv*., **142**: 2350-2360. https://doi.org/10.1016/j.

biocon.2009.05.014
Nei, M., 1972. Genetic distance between populations. *Am. Nat.,* **106**: 283–292. https://doi.org/10.1086/282771
Nei, M., 1975. Evolutionary genetics. (book reviews: molecular population genetics and evolution). *Science*, **190**: 372-373.
Norman, A.J. and Spong, G., 2015. Single nucleotide polymorphism-based dispersal estimates using noninvasive sampling. *Ecol. Evol*., **5**: 3056-3065. https://doi.org/10.1002/ece3.1588
O'Brien, S.J., Menotti, R.M., Murphy, W.J., William, G.N., Johannes, W., Roscoe, S., Neal, G.C., Nancy, A.J., James, E.W. and Jennifer, A.M.G., 1999. The promise of comparative genomics in mammals. *Science*, **286**: 479-481. https://doi.org/10.1126/science.286.5439.458
Pritchard, J.K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, **155**: 945–959.
Roos, C., Boonratana, R., Supriatna, J. and Fellowes, J.R. 2014. An updated taxonomy and conservation status review of Asian primates. *Asian Primate. J.,* **4**: 2-38.
Ruell, E.W. and Crooks, K.R., 2007. Evaluation of Noninvasive Genetic Sampling Methods for Felid and Canid Populations. *J. Wildl. Manage*., **71**: 1690-1694. https://doi.org/10.2193/2006-061
Saitou, N. and Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*., **4**: 406–425.
Sun, B., Li, J., Zhu, Y. and Xia, D., 2010. Mitochondrial DNA variation in Tibetan macaque (*Macaca thibetana*). *Folia Zool*., **59**: 1861. https://doi.org/10.25225/fozo.v59.i4.a5.2010
Taberlet, P., 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic. Acids. Res.,* **24**: 3189-3194. https://doi.org/ 10.1093/nar/24.16.3189
Taberlet, P., Waits, L.P. and Luikart, G., 1999. Noninvasive genetic sampling: look before you leap. *Trends Ecol. Evol*., **14**: 323-327. https://doi.org/10.1016/S0169-5347(99)01637-7
Takezaki, N., Nei, M. and Tamura, K., 2010. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with windows interface. *Mol. Biol. Evol*., **27**: 747-752. https://doi.org/10.1093/molbev/msp312
Valière, N., 2010. GIMLET: A computer program for analysing genetic individual identification data. *Mol. Ecol. Resour*., **2**: 377-379. https://doi.org/10.1046/j.1471-8286.2002.00228.x
Varshney, R.K., Chabane, K., Hendre, P.S., Aggarwal, R.K. and Graner, A., 2007. Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. *Pl. Sci*., **173**: 638-649. https://doi.org/10.1016/j.plantsci.2007.08.010
Vignal, A., Milan, D., Sancristobal, M. and André, E., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol*., **34**: 275-280. https://doi.org/10.1186/1297-9686-34-3-275
Villesen, P. and Fredsted, T. 2006. Fast and non-invasive pcr sexing of primates: Apes, old world monkeys, new world monkeys and strepsirrhines. *BMC Ecol.*, **6**: 8. https://doi.org/10.1186/1472-6785-6-8
Waits, L.P. and Paetkau, D., 2005. Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *J. Wildl. Manage*., **69**: 1419-1433. https://doi.org/10.2193/0022-541X(2005)69[1419:NGSTFW]2.0.CO;2
Wang, S., 1998. *China red data book of endangered animal (Mammalia).* Science Press, Beijing.
Wang, W., Qiao, Y., Li, S., Pan, W. and Yao, M., 2017. Low genetic diversity and strong population structure shaped by anthropogenic habitat fragmentation in a critically endangered primate, *Trachypithecus leucocephalus*. *Heredity*, **118**: 542–553. https://doi.org/10.1038/hdy.2017.2
Weiran, W., Yu, Q., Wenshi, P., Meng, Y. and Russello, M.A., 2015. Low genetic diversity and strong geographical structure of the critically endangered white-headed langur (*Trachypithecus leucocephalus*) inferred from mitochondrial DNA control region sequences. *PLoS One*, **10**: e0129782. https://doi.org/10.1371/journal.pone.0129782
Yao, Y., Zhong, L., Liu, B., Li, J., Ni, Q. and Xu, H., 2013. Genetic variation between two tibetan macaque (*Macaca thibetana*) populations in the eastern china based on mitochondrial DNA control region sequences. *DNA Sequence*, **24**: 267-275. https://doi.org/10.3109/19401736.2012.748040
Zhang, X., Niu, L., Zhou, L., Yan, Z., Yang, F., Wu, A., Min, S.K., Li, X., Tao, K.S., Dou, K.F. and Chen, H., 2018. Development and characterization of 29 SNP markers for the Tibetan macaque (*Macaca thibetana*). *Conserv. Genet. Resour*., **9**: 325-328.
Zhong, L.J., Zhang, M.W., Yao, Y.F., Ni, Q.Y., Jun, M., Li, C.Q. and Xu, H.L., 2012. Genetic diversity of two Tibetan macaque (*Macaca thibetana*) populations from Guizhou and Yunnan in China based on mitochondrial DNA D-loop sequences. *Gene Genom*., **35**: 205-214. https://doi.org/10.1007/s13258-012-0048-2