

Ecological Genetics of Chinese Rhesus Macaque in Response to Mountain Building: All Things Are Not Equal

Shan-Jin Wu^{1,2,3,9}, Jing Luo^{1,9}, Qing-Qing Li⁴, Yan-Qin Wang¹, Robert W. Murphy^{3,5}, Christopher Blair⁶, Shi-Fang Wu³, Bi-Song Yue^{2*}, Ya-Ping Zhang^{3*}

1 Laboratory for Conservation and Utilization of Bio-resources, School of Life Science, Yunnan University, Kunming, China, **2** Sichuan Key Laboratory of Conservation Biology on Endangered Wildlife, College of Life Sciences, Sichuan University, Chengdu, Sichuan, China, **3** State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming, China, **4** School of Life Sciences, Yunnan Normal University, Kunming, China, **5** Centre for Biodiversity and Conservation Biology, Department of Natural History, Royal Ontario Museum, Toronto, Ontario, Canada, **6** Department of Biology, Duke University, Durham, North Carolina, United States of America

Abstract

Background: Pliocene uplifting of the Qinghai-Tibetan Plateau (QTP) and Quaternary glaciation may have impacted the Asian biota more than any other events. Little is documented with respect to how the geological and climatological events influenced speciation as well as spatial and genetic structuring, especially in vertebrate endotherms. *Macaca mulatta* is the most widely distributed non-human primate. It may be the most suitable model to test hypotheses regarding the genetic consequences of orogenesis on an endotherm.

Methodology and Principal Findings: Using a large dataset of maternally inherited mitochondrial DNA gene sequences and nuclear microsatellite DNA data, we discovered two maternal super-haplogroups exist, one in western China and the other in eastern China. *M. mulatta* formed around 2.31 Ma (1.51–3.15, 95%), and divergence of the two major matrilineages was estimated at 1.15 Ma (0.78–1.55, 95%). The western super-haplogroup exhibits significant geographic structure. In contrast, the eastern super-haplogroup has far greater haplotypic variability with little structure based on analyses of six variable microsatellite loci using STRUCTURE and GENELAND. Analysis using MIGRATE detected greater gene flow from WEST to EAST than vice versa. We did not detect signals of bottlenecks in most populations.

Conclusions: Analyses of the nuclear and mitochondrial datasets obtained large differences in genetic patterns for *M. mulatta*. The difference likely reflects inheritance mechanisms of the maternally inherited mtDNA genome versus nuclear biparentally inherited STRs and male-mediated gene flow. Dramatic environmental changes may be responsible for shaping the matrilineal history of macaques. The timing of events, the formation of *M. mulatta*, and the divergence of the super-haplogroups, corresponds to both the uplifting of the QTP and Quaternary climatic oscillations. Orogenesis likely drove divergence of western populations in China, and Pleistocene glaciations are likely responsible for genetic structuring in the eastern super-haplogroup via geographic isolation and secondary contact.

Citation: Wu S-J, Luo J, Li Q-Q, Wang Y-Q, Murphy RW, et al. (2013) Ecological Genetics of Chinese Rhesus Macaque in Response to Mountain Building: All Things Are Not Equal. PLoS ONE 8(2): e55315. doi:10.1371/journal.pone.0055315

Editor: Bengt Hansson, Lund University, Sweden

Received: August 6, 2012; **Accepted:** December 21, 2012; **Published:** February 6, 2013

Copyright: © 2013 Wu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the National Basic Research Program of China (973 Program, 2007CB411600 and 2007CB109106), National Natural Science Foundation of China, and Bureau of Science and Technology of Yunnan Province, and by a Visiting Professorship for Senior International Scientists from the Chinese Academy of Sciences and a Natural Sciences and Engineering Research Council Discovery Grant (A3148) to R. W. Murphy. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: bsyue@yahoo.com (BSY); zhangyp@mail.kiz.ac.cn (YPZ)

9 These authors contributed equally to this work.

Introduction

The Pliocene uplifting of the Qinghai-Tibetan Plateau (QTP) and Quaternary glaciation may be the primary drivers of genetic divergence and patterning of the Asian biota. With a mean elevation of about 4000 m above sea level and an area of approximately 2.5 million km², the QTP is the largest high-elevation plateau on Earth. Three phases—A, B, and C, and occurring about 3.6 million years ago (Ma), 2.6 Ma, and 1.7 Ma, respectively—characterize the last orogenic push of the QTP [1]. This orogenesis is associated with the formation of Asian monsoons, the beginning of Chinese loess (aeolian sedimentary deposits), and the appearance of the Yellow

(Huanghe) River [1]. Drastic environmental changes in Southeast Asia and especially in China are a response to these events. Dramatic climatic oscillations of the Quaternary Period profoundly affected the current distribution of most living organisms and their genetic structuring [2,3]. At times of Pleistocene glacial maxima, ice sheets did not cover most of East Asia [3,4], yet glaciation and its associated climatic oscillations profoundly affected some biota [5,6].

Geological and climatological events influenced speciation and spatial and genetic structuring, especially among ectotherms. However, much remains to be documented with respect to endothermic vertebrates. *Macaca mulatta*, the rhesus macaque, may be the most suitable model to investigate the relationship between

the non-human primates and the genetic consequences of orogenesis. Its geographic range extends from Afghanistan in the west to the coast of the East China Sea in the east. The species occurs southwards from the Himalayas and central China to central India and central Indochina. This range, which exceeds that of all other primates except for humans [7,8], is sufficiently extensive and old that major genetic structuring might have evolved [9].

Genus *Macaca* appears to have originated in North Africa approximately 5.5 Ma [10]. It seems to have dispersed to the Sawalik Hills of northern India, perhaps by a southern route, before 3 Ma and occupied Indonesia by 2 Ma [11]. Fossil evidence documents the presence of macaques in southeastern [12] and north-central [13] China before or in the early Pleistocene. Using the maternally, paternally, and biparentally inherited gene sequences, the evolutionary relationships of the *M. fascicularis* complex are well documented [14,15,16,17,18]. *Macaca fascicularis* forms the sister group of the other three species of *Macaca*, *M. cyclops*, *M. mulatta*, and *M. fuscata* [19]. The *fascicularis* species group, which also contains rhesus and cynomolgus macaques, shared a common ancestor about 3.2 Ma [20,21].

Diversity in the mitochondrial genome of rhesus macaques is reported from restriction fragment analyses [15,20], DNA sequencing of mitochondrial genes [9,22,23,24,25,26,27,28,29], and microsatellites (short tandem repeats; STRs) [30,31,32]. Some studies using single nucleotide polymorphisms (SNPs) also exist [33,34,35,36]. These studies provide evidence of regional variation within *M. mulatta*. For example, a large degree of genetic divergence occurs between Indian and Chinese populations. Smith and McDonough [9] characterize two major mtDNA haplogroups (CHIE and CHIW), which occur in eastern and western China, respectively. Satkoski et al. [28,36] add a third haplogroup—ChiS—from southwestern China and Viet Nam.

Captive populations form the basis of most genetic assessments. Relatively few wild samples are assessed. Often these samples are missing specific locality data other than their provinces of origin [9,29]. Satkoski et al. [28,36] provided evidence that breeding stock was exchanged among the breeding centers in China, especially those in eastern China where their genetic density was lower. Thus, it is particularly important to use wild samples in the genealogical study of Chinese rhesus macaques. Further, male rhesus macaques are highly mobile and they often leave one troop to join others. In contrast, females rarely do so [37]. For *Macaca mulatta*, the maternal genetic structure, especially in the form of maternal genealogy, is likely to better reflect the history of troop dispersal and colonization than that of paternal history.

We assess the extent of genetic diversity and the evolutionary history of the most widely distributed nonhuman primate, the rhesus macaque, by using a large dataset of maternally inherited mitochondrial DNA gene sequences and nuclear DNA microsatellite data. We test a suite of hypotheses related to the geographic structure of genetic diversity in wild Chinese rhesus macaques based on mtDNA and microsatellite markers. Further, we test hypotheses on the relative roles played by dramatic environmental changes, in particular orogenesis of the Qinghai-Tibetan Plateau (QTP), and Quaternary climatic oscillations in shaping genetic structure.

Materials and Methods

Ethics Statement

The Anhui experimental rhesus monkey Center, Chengdu pingan animal Breeding and Research Center and Primate experimental animal center of KIZ have provide the provided the blood samples. Blood samples were salvaged from routine

health checks for newly captured wild macaques at these breeding centers. We have obtained permission from the centers to analyze the blood samples for this study. These captures are usually as the founder populations of the breeding populations. And These captures of rhesus monkey were authorized by the specific government agency, for example, the Sichuan Department of Forestry, Yunnan Department of Forestry, or An Department of Forestry which is in charge of a certain sampling region. We collected the fecal samples from the Henan Taihang Macaque Nature Reserve, and the Zhangjiajie National Forest Park. We have obtained permission from the above administrative departments to observe the animal behavior of rhesus monkey and collect the stool samples. We also confirm that the behavioral observations did not impact the animals in any way.

Sampling

We collected 337 wild samples from most of the extant groups of *M. mulatta* across its entire range in China (Figure 1). Samples used for genetic analysis included feces (n = 78) and blood (n = 259). Blood samples were salvaged from routine health checks for newly captured wild macaques at breeding centers. Fecal samples, collected during direct behavioral observations, were stored in 95% ethanol. In order to avoid resampling of fecal material from the same individuals, we sampled each group once and on a specific day only. Each dropping was identified by size, shape, and color, and only one sample located <1.5 m apart was collected [38]. D-loop sequences from 168 *M. mulatta* and 56 haplotypes from Sichuan used by Li et al. [29] were downloaded from GenBank. The downloaded data included 76, 21, 13, 109, and 5 sequences originating in China (Guangxi, 20; Sichuan, 56), Nepal, Myanmar, India, and Viet Nam, respectively. The detailed sample data were listed in Table S1.

Laboratory Procedures

Blood samples were digested using proteinase K. This was followed by a standard 3-step phenol/chloroform extraction procedure. Fecal samples extracted using the 2CTAB/PCI protocol as modified by Vallet et al. [39]. DNA was then electrophoresed through an agarose gel to assess its quality and quantity. High-quality samples were diluted to ~20 ng/μL in TE buffer (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA).

To reconstruct the matrilineal genealogy, three fragments from the mitochondrial genome were selected for sequencing. A partial fragment of the D-loop included 510 base pairs (bp) in length. Another two mitochondrial fragments included part of the cytochrome b (*Cytb*) gene (882 bp) and part of rRNA 16S (861 bp). Partial fragments of the hypervariable D-loop were initially sequenced for all 337 specimens. Subsequently, 16S and *Cytb* fragments were sequenced for a subset of samples (66 specimens) that represented the major haplogroups identified from the genealogical reconstruction of the D-loop data (Table S1). The data from the 16S and *Cytb* fragments were used to estimate divergences deeper than those revealed by the D-loop data. PCR cycling conditions for D-loop consisted of an initial denaturation of 5 min at 94°C, followed by 35 cycles (94°C denaturation for 1 min, 50°C annealing for 1 min, and for 1 min extension at 72°C). Final extension at 72°C was conducted for 10 min. We used the same procedure for *Cytb* and 16S but with annealing at 55°C and 59°C, respectively. Primers sequences were summarized in Table 1.

The amplified DNA fragments were purified via spin columns and sequenced using an ABI PRISM 3730 DNA analyzer following the manufacturer's protocols. Sequences were determined in both directions for each individual. To preclude the inclusion of NUMTs, all fragments were submitted for BLAST

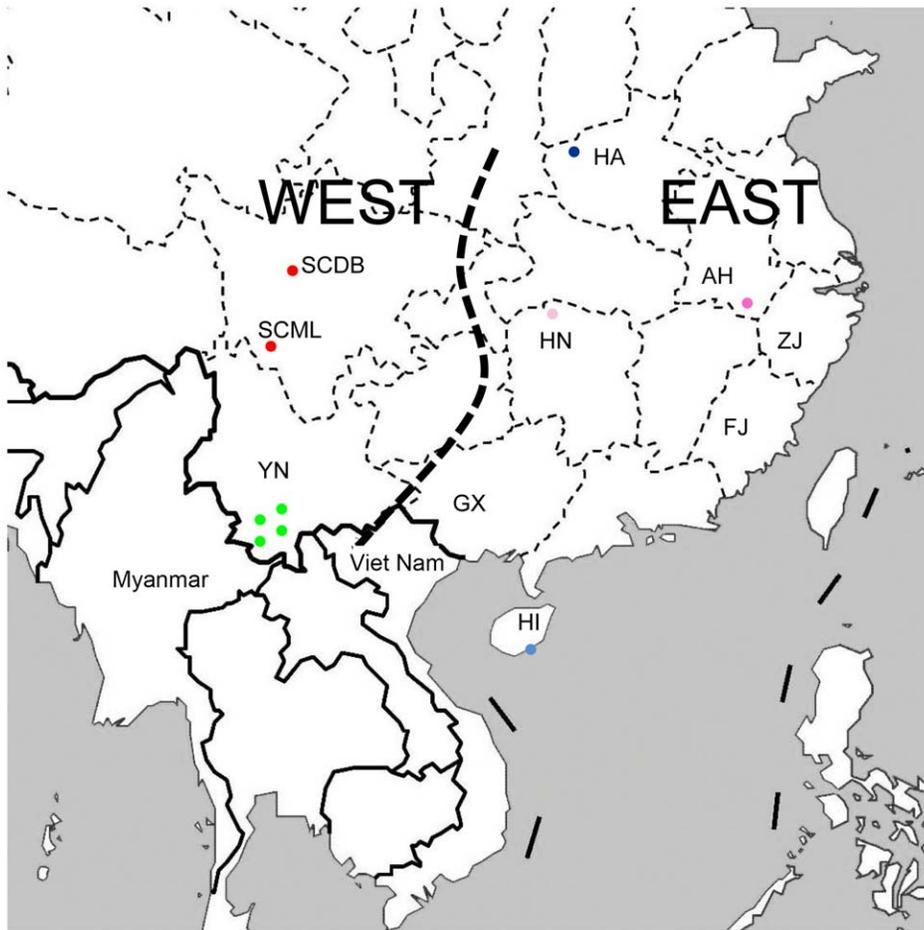


Figure 1. Map of sampling locations for the rhesus macaque, *Macaca mulatta*. Dots indicate specific locality data. SC = Sichuan (SCML = Muli; SCDB = Danba), YN = Yunnan (YNJG = Jinggu; YNSM = Simao (=Pu'er); YNZY = Zhenyuang; YNMJ = Mojiang), HI = Hainan, HA = Henan, HN = Hunan, GX = Guizhou, GX = Guangxi, FJ = Fujian, and ZJ = Zhejiang.
doi:10.1371/journal.pone.0055315.g001

searching [40] in GenBank to ensure that the required sequences had been amplified. Further, protein-coding nucleotide sequences

were translated to amino acids using MEGA 4.1 [41] to check for premature stop codons.

Table 1. Primers used in PCR and sequencing for the rhesus macaque, *Macaca mulatta*.

Locus	Primer name	Primer sequence	Size(bp)	Cited source
D-loop	1	5'-CCG CCC ACT CAG CCA ATT CCT GTT CT-3'	835	[9]
	2	5'-CCC GTG ATC CAT CGA GAT GTC TT-3'		[9]
	3	5'-TCC TAG GGC AAT CAG AAA GAA AG-3'	510	[90]
	4	5'-CCT GAA GTA GGA ACC AGA TG-3'		[90]
	5	5'-ATT CGT GCA TTA CTG CTA G-3'		this study
Cytb	6	5'-CAA CTA TAA AAG CAC CAA TGA C-5'	882	this study
	7	5'-GTT CGC TTC CAA TTC AGG T-3'		this study
	8	5'-CCT ACA CGA AAC AGG ATC AA-3'		this study
	9	5'-TGT AAG GTG AAG AAT CGT GTG-3'		this study
16S	10	5'-GTC CAA GGA GGA TTT AGC-3'	861	this study
	11	5'-TGT TAG GTA TCG GTG AGG-3'		this study
	12	5'-CGA AAC CAG ACG AGC TAC C-3'		this study
	13	5'-GGT AGC TCG TCT GGT TTC G-3'		this study

doi:10.1371/journal.pone.0055315.t001

We chose six previously reported tetranucleotide microsatellite loci (Table 2) to assess biparental gene flow. PCR procedures involved an initial denaturation at 94°C for 4 min followed by 30 cycles of 30 sec at 94°C, 40 sec at 52.1–60.0°C for annealing, 30 sec at 72°C, and a final extension at 72°C for 10 min. PCR products were diluted and mixed with LIZ 500 size standards. Microsatellites were obtained using an ABI PRISM 3730 DNA analyzer (ABI, USA). Allele scoring was performed using GENE-MAPPER [42].

MtDNA Data Analyses

DNA sequence alignment. Alignments were first conducted using CLUSTAL W [43] integrated into MEGA v.4.1 [41] while employing default parameters. When required, the initial alignments were subsequently adjusted by eye using MEGA v.4.1. Protein-coding nucleotide sequences were translated to amino acids using MEGA v.4.1 to confirm alignments. Gaps were considered to be a nucleotide variant (fifth state). Identical haplotypes were collapsed using DNASP, v.5.0 [44].

Matrilineal relationships. For phylogenetic analyses, several species of macaques were selected as outgroup taxa. Two mitochondrial datasets were analyzed separately. First, the D-loop fragments alone, including 214 unique haplotypes plus the four outgroup taxa from *M. fascicularis*, *M. thibetana*, *M. sylvanus*, and *M. arctoides*, were subjected to phylogenetic analyses. The second dataset included all three mtDNA fragments for 67 haplotypes plus 12 outgroup taxa including *M. fascicularis*, *M. thibetana*, *M. sylvanus*, *M. arctoides*, *M. assamensis*, *M. maura*, *M. nigra*, *M. ochreata*, *M. pagensis*, *M. radiata*, and *M. silenus*.

A heuristic maximum parsimony approach was conducted for both datasets. All characters were unordered and equally weighted. The search was conducted using PAUP* v.4.0b10 [45] with 100 random stepwise additions and TBR branch swapping. Nonparametric bootstrap proportions (BSP) based on 1000 pseudoreplicates were used to infer nodal support.

For the first dataset, associations among the 214 D-loop haplotypes were also visualized by a full median-joining network [46,47] with MP post-processing [48] as implemented in NETWORK v.4.5 (www.fluxus-engineering.com/sharenet.htm). For

the second dataset, phylogenetic relationships were also inferred using Bayesian inference as implemented in MRBAYES v.3.1.2 [49]. The best-fitting nucleotide substitution model for each gene (16S: GTR+I+G, *Cytb*: GTR+G, D-loop: HKY+I+G) was selected based on the Akaike Information Criterion as implemented in MODELTEST v.3.7 [50] using default parameters. Four Monte Carlo Markov chains (MCMC) were used and the dataset was run for 10 million generations and trees were sampled every 200 generations. The last 25% of the sampled trees were used to estimate the consensus tree and corresponding Bayesian posterior probabilities.

Isolation-by-distance and historical population analyses. Using the first dataset, we tested the null hypothesis of a correlation between genetic and geographical distances, i.e. isolation-by-distance (IBD), for the matrilineal using both all ten sampling localities and the six western sites that had specific locality data (Figure 1). Approximate geographic coordinates were determined using Google Earth. Distances between localities were estimated using the spherical distance between two points. The pairwise genetic differentiation values were assumed to measure the extent of DNA divergence between populations, and the significance was tested using 1,000 permutations with ARLEQUIN v.3.11 [51]. Correlation of geographic and genetic distances was determined using Mantel's permutation test with 10,000 permutations executed by IBDWS v.3.15 [52]. The strength of this relationship was determined by regressing all pairwise genetic differentiation values, F_{ST} , against the corresponding \log_{10} transformed geographical distance.

Historical population dynamics of the major haplogroups assigned in the median-joining network were examined with mismatch distribution analysis [53] as implemented in ARLEQUIN v.3.11. This analysis compared the observed frequencies of pairwise differences of haplotypes with those expected under a single sudden expansion model [53]. An expected distribution under a model of sudden demographic expansion was generated with a total of 1000 permutations. Demographic stability has produced multimodal distributions, and unimodal patterns have occurred during sudden population expansion [54]. The raggedness index was expected to have a higher value in relatively stable

Table 2. Primer sequences used in the microsatellite study.

STR marker	chromosome	Primer sequence	Reference	Size range (bp)	N_A	Annealing temperature
D7S1826	3	5'-[FAM]CATCCATCTATCTCTGTAATCTCTC-3'	[30]	93+115–155	12	58.4°C
		5'-TATTTAACACACCTGTCTCAATCC-3'				
D8S1106	8	5'-[FAM]TTGTTTACCCTGCATCACT-3'	[30]	131–191	10	58.0°C
		5'-TTCTCAGAATTGCTCATAGTGC-3'				
D1S548	1	5'-[FAM]GAACTCATTGGCAAAGGAA-3'	[91]	187–199	4	56.8°C
		5'-GCCTCTTGTGTCAGTGATT-3'				
D5S1457	6	5'-[FAM]TAGGTTCTGGGCATGTCTGT-3'	[91]	103–135	9	58.4°C
		5'-TGCTTGGCACACTTCAGG-3'				
D11S2002	14	5'-[FAM]CATGCCCTTCTTTTCATAG-3'	[30]	235–275	11	58.4°C
		5'-AATGAGGTCTTACTTTGTTGCC-3'				
D15S644	7	5'-[FAM]CCTTCATTGGCAGACTCACT-3'	[30]	174–262	23	58.4°C
		5'-GCAGACCAAGATGATAACG-3'				

The fluorescent dyes, original references, and annealing temperatures used for rhesus macaque-derived markers, with the number of alleles (N_A) and allele size range found in the 265 samples screened in this study are listed.^{a,b}

^aThe sequence of primers used to amplify each locus can be obtained from Research Genetics Inc.

^bThe chromosome number of the STR markers in *Macaca mulatta* was determined by e-PCR in the UniSTS database from NCBI.

doi:10.1371/journal.pone.0055315.t002

populations. Tajima's D [55] and Fu's F_s statistics [56], calculated using ARLEQUIN v.3.11, were also used to seek evidence of demographic expansions within individual haplogroups and subhaplogroups. Statistical significance was tested with 1000 permutations. The null hypothesis was that of a stable population. Additionally, we calculated the statistics of genetic diversity, neutrality tests, and mismatch distributions for the main populations of *Macaca mulatta* in China.

Estimation of divergence time. Divergence times were estimated using a Bayesian MCMC method implemented in BEAST v.1.5.3 [57], which employed a relaxed molecular clock approach [58]. A relaxed lognormal clock model of haplogroup variation and a Yule prior for branching rates was assumed. Divergence time estimations were based on the second dataset only. After removing indels in the outgroup taxa, the final alignment for divergence age estimations comprised 2,260 nucleotide positions. The dataset was partitioned and each optimal nucleotide substitution model was selected by using the Akaike Information Criterion as implemented in MODELTEST v.3.7 [50].

As a calibration point, we used the divergence time of 5.5 Ma (95% confidence interval = 4.68–6.32 Ma) [10] for matrilineal relationships between *M. sylvanus* and the other Asian macaques. Instead of using hardbounded calibration points, we relied on published dates and specified a normal distribution prior, with a mean of 5.5 Ma and a standard deviation of 0.5 Ma. Two replicates were run for 100 million generations while sampling trees and parameters every 1,000 generations. The adequacy of using a 50% burn-in to generate MCMC trees and for convergence of all parameters was assessed visually using TRACER v.1.4.1 [59]. Subsequently, the sampling distributions of two independent replicates were combined using the software LOGCOMBINER v.1.5.2 [54], and the resulting 100,002 samples were summarized and visualized using TREEANNOTATOR v.1.5.2 [54] and FIG TREE v.1.2 [60].

STR Analyses

Statistical analyses. All loci were collected as an EXCEL file, and then converted into the computable files using CONVERT [61]. We used ARLEQUIN v.3.11 to check if the microsatellite data departed from Hardy-Weinberg expectations. Tests for linkage disequilibrium were checked by using the GENEPOP v.4.0 [62]. The total number of alleles (N_A), allelic richness (R), gene diversity (GD) and allele frequencies were calculated in FSTAT v.2.9.3.2 [63].

The presence of genetic bottlenecks was tested by using the heterozygosity excess method [64] implemented in BOTTLENECK v.1.2.02 [65]. Following recommendations of Piry et al., we used the two-phase model (TPM) with 95% stepwise mutations and a variance of 12 [65,66]. We used the Wilcoxon signed-rank test because this test was proven to be more powerful when less than 20 loci are used [65]. The mode-shift test implemented in BOTTLENECK [65] was also run.

Population structure and migration. First, we assessed population structure by using pairwise F_{ST} values calculated in ARLEQUIN v.3.11 and the Bayesian clustering approach implemented in STRUCTURE v.2.3.3 [66,67,68]. For analyses using STRUCTURE, the admixture model with correlated allele frequencies was chosen and the number of genetic groups K was set from 1 to 8. Each run used five replicates, which consisted of a burn-in period of 1 000 000 MCMC steps followed by 1 000 000 MCMC steps. Plots of the mean value of the log posterior probability of the data [mean of $\ln P(D)$] were examined to check the maximum $\ln P(D)$ value [66]. Because spatially explicit Bayesian clustering methods can be powerful when inferring genetic structure [69,70], particularly at low levels of differentiation [71,72], we used

GENELAND v.4.0.3 [72,73,74] to search for structure in the rhesus macaque. We specified a K_{max} of 8, used the correlated allele frequency model, and allowed for the presence of null alleles. This analysis included samples that had specific locations or provincial localities. The analysis was run for 1,000,000 generations and a thinning of 1,000. To make sure the MCMC was converging and mixing properly, we used 10 independent runs and compared parameter estimates. For parameter inference we selected the run with the highest log posterior probability and specified a burnin of 250.

We also used the maximum likelihood and Bayesian inference [75,76,77] algorithms in MIGRATE v.3.3.2 [78] to estimate test the null hypothesis of unbiased, equal rates of migration rates and equal effective population sizes between the two largest haplogroups resolved in the haplotype network. We chose the single-step model in the analysis. For the maximum likelihood analysis, we specified three long chains while discarding the first 100,000 trees per chain. For the Bayesian inference, we employed 10 replicates while also discarding the first 100,000 trees per chain as burn-in.

Results

Sequence Data

The D-loop-only dataset included 168 variable (polymorphic) sites of which 26 were autapomorphic and 142 potentially parsimony informative. Inclusion of 16 indels resulted in 214 haplotypes, 92 of which we generated and 122 downloaded from GenBank. The haplotypic composition of each sampled population was summarized in Table S1. The concatenated dataset contained 2,260 nucleotide positions in 67 ingroup haplotypes plus 12 in the outgroup. About 29% of the positions (660 bp) were variable and 462 were potentially parsimony-informative. No evidence of NUMTs was detected.

Genealogical Relationships

Analysis of the D-loop data did not yield a well-resolved MP tree (Figure S1) whereas the haplotype network exhibited considerable substructure (Figure 2). The two major haplogroups of Chinese rhesus macaques identified in previous studies [9,24]—WEST and EAST—were recovered in the network. Sixteen mutational steps separated these two clusters. The network further delineated two discrete subhaplogroups within WEST: WEST A and WEST B. WEST A, included 45 haplotypes from Sichuan, two from Yunnan, and two from Guizhou, and corresponded to subhaplogroups ChiW2 (100% support) and ChiW3 (65% support) reported by Smith and McDonough [9], and Haplogroup A (100% support) reported by Li et al. [24]. Subhaplogroup WEST B, which mainly corresponded to subhaplogroup ChiW1 and Ind2 [9] and haplogroup B [29], was comprised of 57 haplotypes, including 35 from Yunnan, 12 from Sichuan, five from India B, and eight from Myanmar.

Haplogroup EAST mainly corresponded to haplogroup ChiE and Ind1, defined by Smith and McDonough [9], and to haplogroups C, D, and E, reported by Li et al. [24]. Among all haplogroups from Viet Nam, and East, West, and South China, EAST was most widespread geographically and contained the greatest amount of genetic diversity. It was comprised of 11 haplotypes from Viet Nam, 11 from Fujian, eight from Zhejiang, 21 from Guangxi, five from Anhui, three from Henan, one from Hainan, one from Hunan (one sequence), two from Yunnan, and six from Sichuan provinces. EAST did not have distinct subgroups and most haplotypes were restricted to one locality. Populations in both Hainan and Hunan had multiple samples yet only one

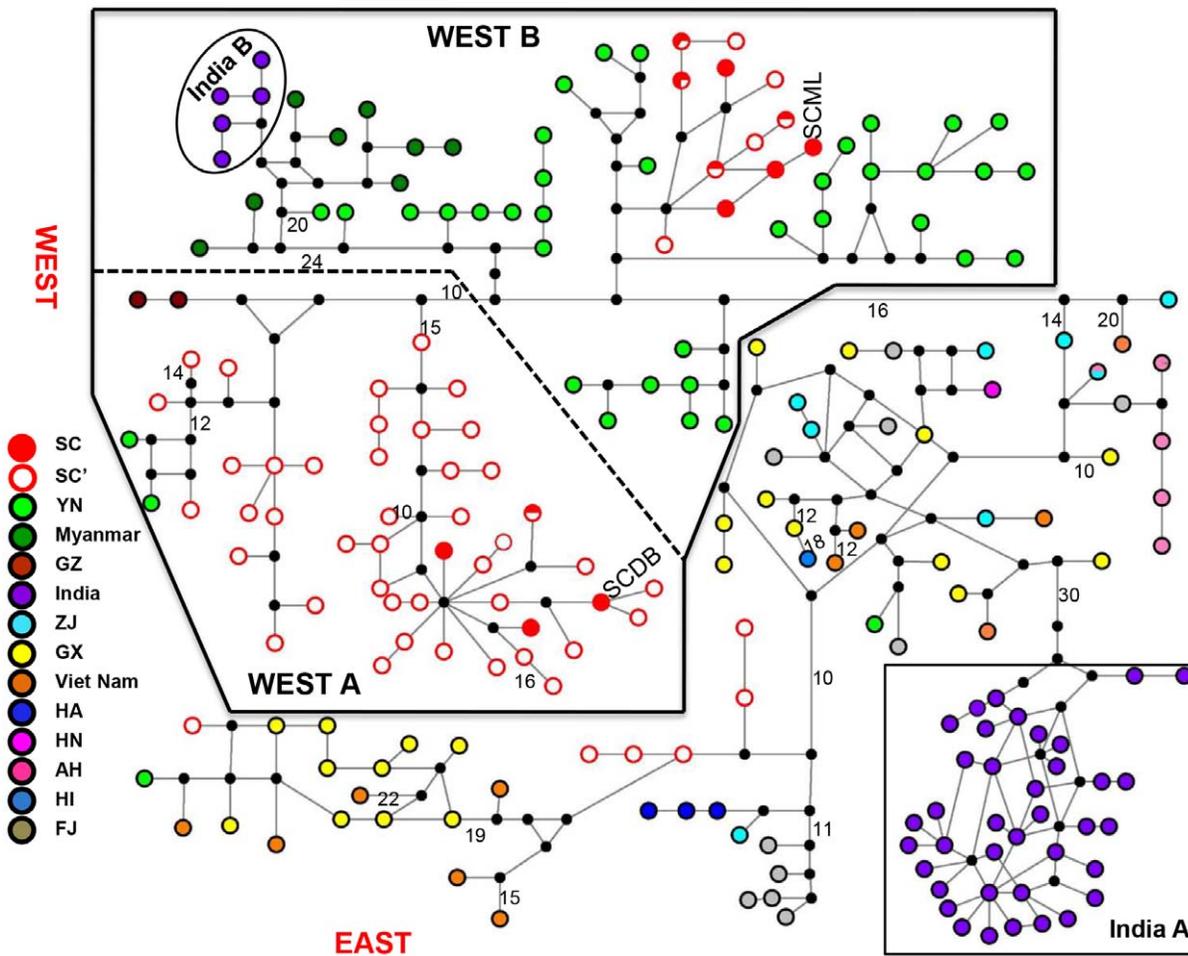


Figure 2. Median-joining network for 214 haplotypes from 735 rhesus macaques, *Macaca mulatta*, based on 489 bp of mtDNA D-loop. The size of each circle is not proportional to the observed frequency of a given haplotype. Closed black circles are inferred intermediate haplotypes not sampled in our study. The largest possible number of mutational steps between two haplotypes is shown. doi:10.1371/journal.pone.0055315.g002

haplotype. The 38 haplotypes from India comprised a unique haplogroup, which was separated by 30 mutational steps from nearest haplogroup EAST. Because of lack of specific locality information for the Indian populations, we could not conduct further analyses. Regardless, evidence suggested that the more widespread Indian and Burmese rhesus macaques contained greater genetic variability than the more geographically restricted populations from western China, as reported by Smith and McDonough [9].

For the concatenated data, the Bayesian inference and maximum parsimony genealogies had identical topologies and similar support values (Figure 3). The hypothesized genealogy clearly resolved well-supported WEST and EAST haplogroups of Chinese *M. mulatta*. The WEST haplogroup was further divided into two subhaplogroups. Subhaplogroup WEST B was represented by 21 haplotypes from Yunnan and Sichuan (YNSM, YNZY, YNJG, YNMJ, and SCML) and four others were from Sichuan (SCJY). Subhaplogroup WEST A, from Sichuan and Yunnan, had two haplotypes from SCDB, two from SCJY, and two from Yunnan. Unfortunately, the two samples from Yunnan in WEST A and all the samples from SCJY were missing specific locality data. EAST was comprised 35 haplotypes, which included all sampling locations other than Yunnan and Sichuan. One haplotype from Yunnan (YN) (Hap_89YNSM) clustered in EAST.

Isolation-by-distance and Historical Population Analysis

Although we had relatively few populations with specific sampling information, a positive correlation between geographical distance and genetic differentiation was obtained. The average pairwise F_{ST} estimates based on the D-loop sequence data from the ten populations of *Macaca mulatta* that had specific sampling locality information was 0.769. The two populations from Sichuan had a far larger F_{ST} value (0.984) compared with the four populations from Yunnan ($F_{ST}=0.0784$). IBD was observed for both all the ten populations and the six western populations that had sampling data (Figure 4A, B). The correlation between geographical distances and genetic differentiation was statistically significant for both comparisons (Mantel $r = 0.8137$; $P < 0.001$ with 10,000 permutations; Mantel $r = 0.8527$; $P = 0.0188$ with 10,000 permutations, respectively). We did not calculate IBD using STRs because of the results from STRUCTURE (below).

Historical demographics for WEST A, WEST B, India A, and EAST provided important insights into the history of *M. mulatta* (Figure 5; Table 3). WEST A and WEST B had multimodal distributions with steep curves and a high frequency of diverged haplotypes, indicating that mismatch distributions deviated from expectations of expansion. Population expansion was indicated both within EAST and India A. Significant population growth for

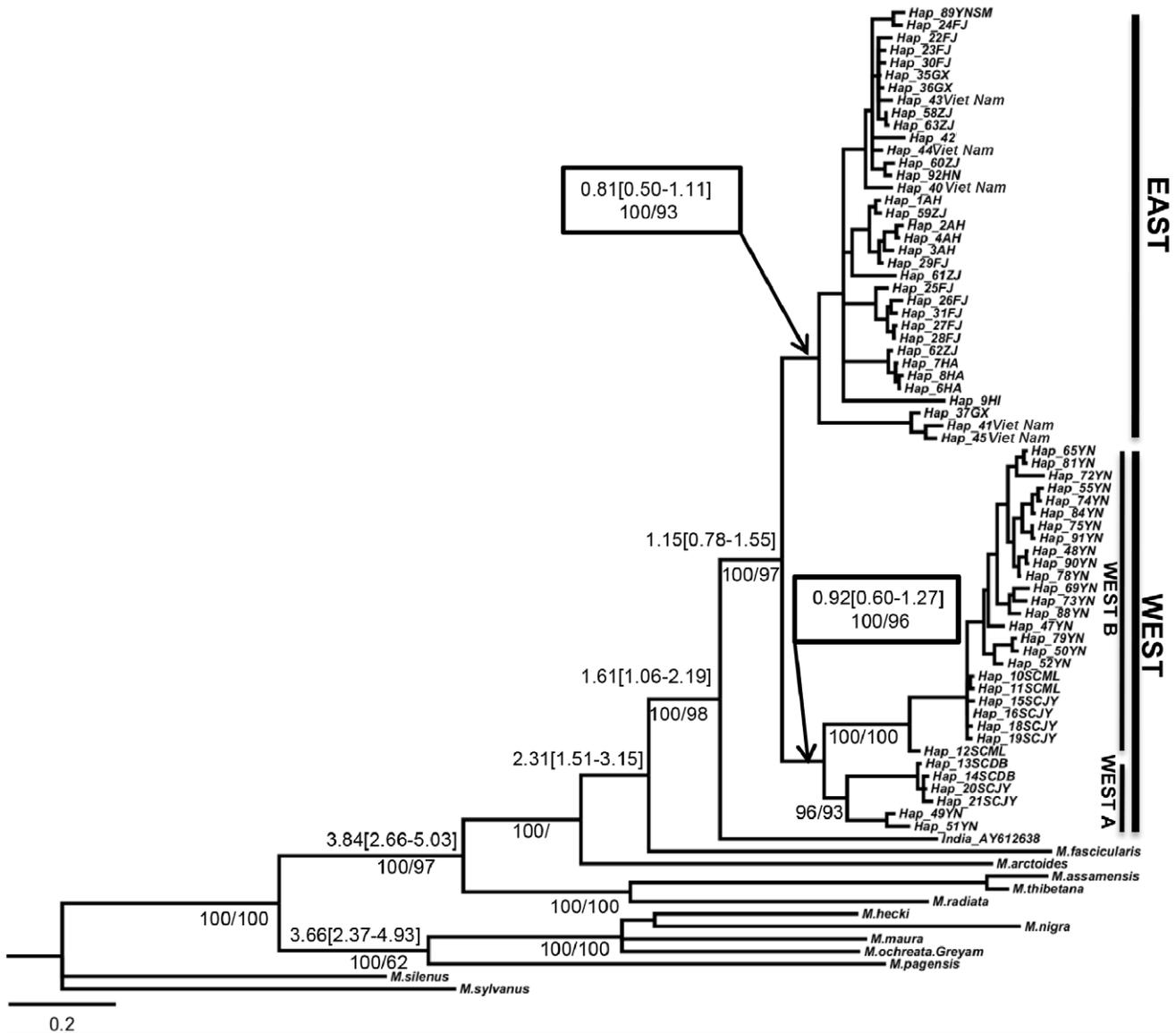


Figure 3. Bayesian inference matrilineal genealogy for the rhesus macaque, *Macaca mulatta*. The genealogy is based on the concatenated sequences of three mtDNA gene sequences (16S+cytb+D-loop) consisting of 2260 bp. For the main nodes, Bayesian posterior probabilities (BPP \geq 90 retained) and bootstrap support from maximum parsimony (MP; 1000 replicates; MP \geq 50 retained) are shown above the branch, respectively. And the main nodes were dated by Bayesian inference using a Yule prior and the estimates are given below the branch, with 95% confidence intervals (CI). doi:10.1371/journal.pone.0055315.g003

the four haplogroups was not inferred by neutrality tests, except for India A (Tajima’s $D = -1.6218$; P value = 0.0330). For the main populations of Chinese *M. mulatta*, there was considerable variation in haplotype diversity (h) and it ranged from 0 to 1.0000 ± 0.063 (Table S2). Similarly, nucleotide diversity (π) ranged from 0 to 0.06052 ± 0.00521 . These values were highest in ZJ, FJ, GX, and Viet Nam, a reflection of the sympatric occurrences of multiple haplotypes. In addition to the haplotypes from above, four provinces occupied central locations in EAST suggesting that these haplotypes had a greater heterogeneity and an earlier history of divergence than other groups in EAST. The very low haplotype and nucleotide diversity in HA and AH corresponded to their peripheral locations in the median-joining network (Figure 2).

Estimated Divergence Times

The main nodes (Figure 3) were dated with 95% confidence intervals. The split between *M. mulatta* and *M. fascicularis* was dated at 2.31 Ma (1.51–3.15), and the separation of the western and eastern haplogroups was dated to 1.15 Ma (0.78–1.55). The age of the most recent common ancestor (MRCA) of the western haplogroup was estimated at 0.92 Ma (0.6–1.27) while that of the eastern haplogroup was ca. 0.81 Ma (0.5–1.11).

STR Data

A total of 265 contemporary samples were genotyped at six autosomal microsatellite loci (Table S3). All microsatellite loci were located on different chromosomes and the chromosome numbers of the STRs in *M. mulatta* were found by e-PCR in the UniSTS database of NCBI. All loci were unlinked and showed no

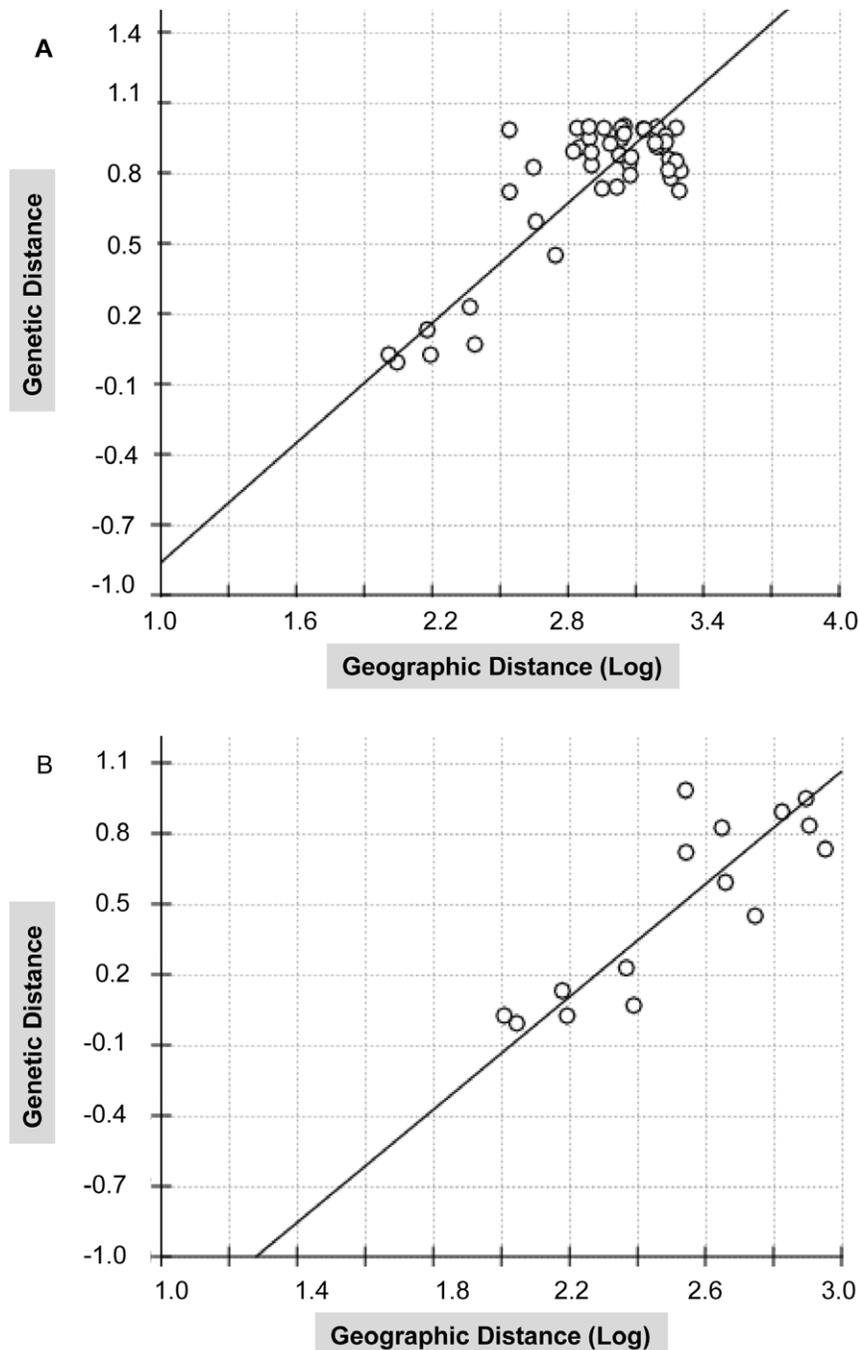


Figure 4. An isolation-by-distance analysis for Chinese rhesus macaques, *Macaca mulatta*. Plots of genetic distance (F_{ST}) and log10 transformed geographical distance between the ten populations that had specific sampling information. A denotes the ten populations that displayed a significant relationship ($P < 0.001$; $R = 0.8137$ with 10,000 permutations); B indicates the six western populations that displayed a significant relationship ($P = 0.0188$; $R = 0.8527$ with 10,000 permutations).
doi:10.1371/journal.pone.0055315.g004

departures from Hardy-Weinberg equilibrium. The total number of alleles (N_A), number of private alleles (N_{PA}), allelic richness (R), and gene diversity (GD) were summarized in Table 4. The highest N_A /locus was found in population YNJG (N_A /locus = 9.5), and the highest allelic richness and gene diversity were found in population SCJY ($R = 5.078$; $GD = 0.822$). In contrast, populations HI and SCDB had the lowest values [N_A /locus (HI) = 3.7; R (HI) = 3.667; GD (SCDB) = 0.677]. Using the Wilcoxon signed-rank test, we

detected an excess of heterozygotes in three populations [$P_{(YNMJ)} = 0.039$; $P_{(SCJY)} = 0.039$; $P_{(HI)} = 0.016$]. Except for four populations (YNZY, FJ, HI, and SCJY), the allelic frequencies of nearly all the populations fit the normal L-shaped distribution in the mode-shift test. These results indicated that most populations had not experienced bottlenecks. Pairwise F_{ST} values (Table 5) indicated no significant genetic differentiation between all areas sampled (mean pairwise $F_{ST} = 0.033$). STRUCTURE resolved one

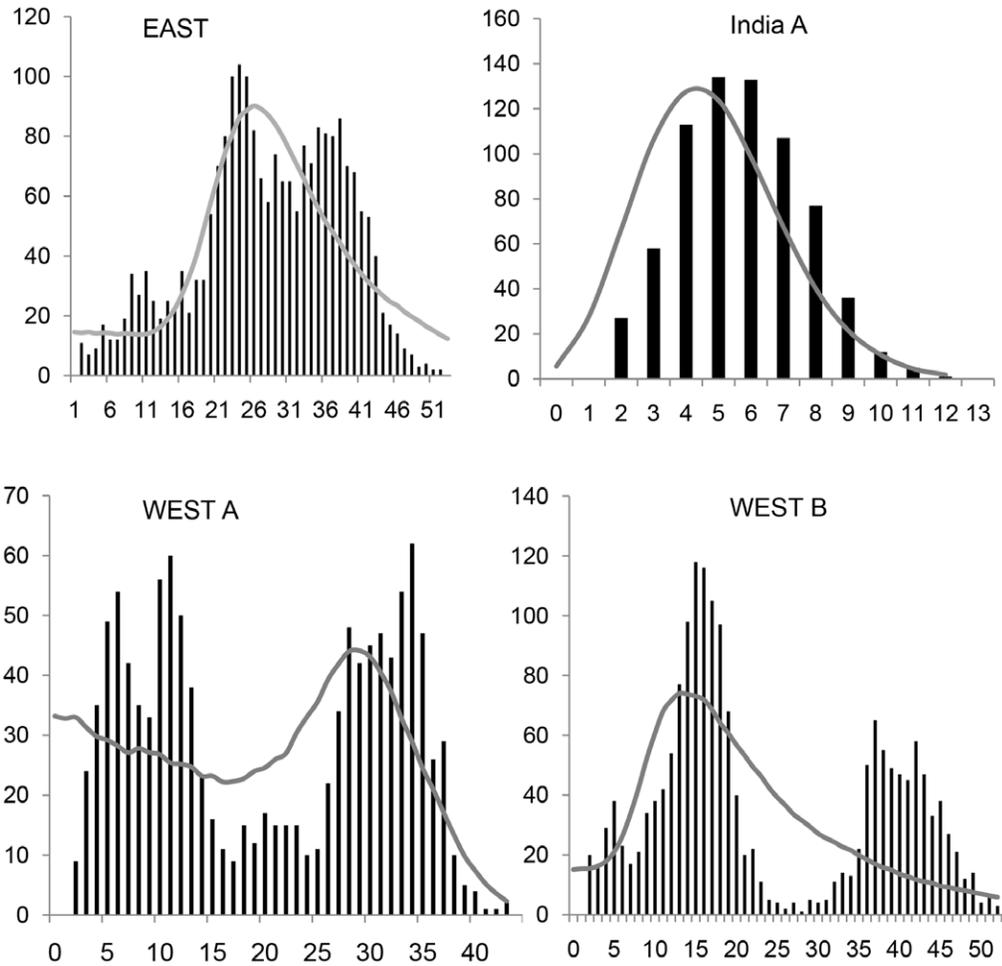


Figure 5. Mismatch distributions of Chinese rhesus macaques, *Macaca mulatta*, for haplogroups EAST, India A, WEST A, and WEST B depicted in Figure 2. The abscissa shows the number of pairwise differences between compared haplotypes. The ordinate shows the proportion for each value. Black bars show observed frequency distributions while the curves show the distribution expected under the sudden-expansion model.
doi:10.1371/journal.pone.0055315.g005

unique genetic group for *M. mulatta* ($K = 1$; Figure 6). Conversely, all ten independent runs of GENELAND based on all samples obtained a K -value of 7, suggesting that the MCMC was mixing well (Figure 7). Although $K = 7$ was the most likely number of clusters based on the run with the highest probability, $K = 6$ was also supported by the data and no individual was placed into cluster 7 with high probability. However, in many cases the posterior probability (PP) value for membership in a given population was not high, indicating the absence of strong support

for inclusion in one population over another. All 102 samples from Yunnan (49 from YNJG, 36 from YNSM, 7 from YNMJ, and 10 from YMZY) composed one population and with the largest posterior probability (PP = 0.332) among all groups. The 29 individuals from SCDB formed one population (PP = 0.219) as did the 30 animals from AH (PP = 0.185). Another population was comprised of 30 individuals from SCML, 8 from ZJ, 2 from Viet Nam, plus 8 from FJ (PP = 0.2112). The fifth population contained seven macaques from GX, 7 from Viet Nam, and seven from FJ

Table 3. Statistics of genetic diversity, neutrality test, and the mismatch distribution analysis.

Haplogroup/Subhaplogroup	Tajima's <i>D</i> (<i>P</i> value)	Fu's <i>F_s</i> (<i>P</i> value)	SSD (<i>P</i> value)	Raggedness index (<i>P</i> value)
WEST A	-0.350(0.394)	-24.175(0.000)	0.007(0.803)	0.003(0.992)
WEST B	0.290(0.695)	-24.100(0.000)	0.011(0.248)	0.002(0.855)
EAST	-0.167(0.513)	-24.076(0.001)	0.002 (0.520)	0.001(0.986)
India A	-1.622(0.033)	-25.612(0.000)	0.001(0.628)	0.018(0.457)

These analysis involved the four haplogroups of *Macaca mulatta* in China. SSD = Sum of Squares Deviations.
doi:10.1371/journal.pone.0055315.t003

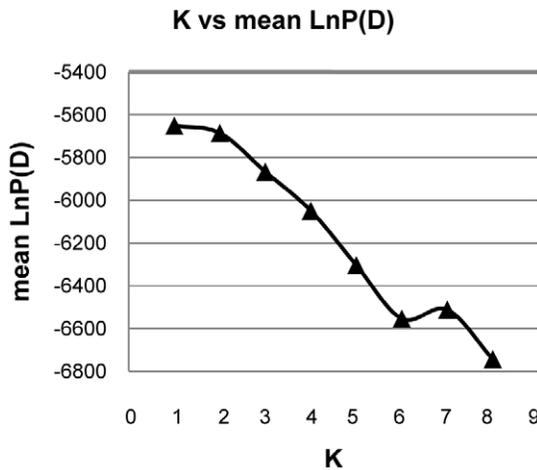


Figure 6. Summary diagram of STRUCTURE analyses for the rhesus macaque, *Macaca mulatta*. The analysis is based on six variable microsatellite loci for animals sampled across China. doi:10.1371/journal.pone.0055315.g006

(PP=0.210). Finally, the five samples from HI composed one population (PP=0.189). Effective population sizes were nearly identical for WEST and EAST (Table 6). However, the migration rate from WEST to EAST was greater than from EAST to WEST in both maximum likelihood predictions and Bayesian inference.

Discussion

Genetic Structure of Chinese *Macaca mulatta*

Using a large dataset of maternally inherited mtDNA gene sequences and nuclear DNA microsatellite data [15,16,17,20,23,79], we discover different genetic patterns for *M. mulatta*. The large F_{ST} value from the mtDNA data indicates a high level of genetic differentiation among populations of *M. mulatta*. Two maternal super-haplogroups exist, one in western China (Yunnan, Sichuan) and the other in eastern China (e.g., Hainan and Henan). The lineages have high levels of genetic divergence

Table 4. Statistical analyses for the microsatellite data.

Locality (N)	GD	NA/locus	R
WEST (192)	0.784	10.7	9.762
YNJG (49)	0.770	9.5	4.793
YNSM (36)	0.794	8.8	5.001
YNMJ (7)	0.766	5.2	4.567
YNZY (10)	0.731	5.8	4.629
YN (20)	0.775	8.0	4.866
SCML (30)	0.799	8.3	5.017
SCDB (29)	0.677	6.3	4.040
SCJY (10)	0.822	6.3	5.078
EAST (74)	0.75683	9.3	9.255
AH (30)	0.680	5.7	3.863
HI (5)	0.754	3.7	3.667
ZJ (8)	0.786	6.0	4.899
GX (7)	0.732	4.5	4.112
Viet Nam (9)	0.795	5.7	4.667
FJ (15)	0.792	6.7	4.803

Localities with sample sizes (N), gene diversity (GD), mean number of alleles (N_A) per locus and allelic richness (R). Locality codes are defined in Figure 1. doi:10.1371/journal.pone.0055315.t004

between them, as previously reported in molecular and morphological studies. For the first time and in contrast to haplogroups, our analyses of the microsatellite data do not detect significant population differences (mean pairwise F_{ST} =0.033; K from STRUCTURE=1; Figures 6, 7), which may suggest that the dispersal history of *M. mulatta* is too recent for the accumulation of genetic difference among the populations. This corresponds to the date of divergence between Indian and Chinese rhesus macaques estimated from SNPs at 162,000 YBP [33]. However, the conflict between the matrilineal history and the absence of nuclear DNA structure may also be due to sex-biased behavior. Male rhesus macaques are far more mobile than females and they often leave

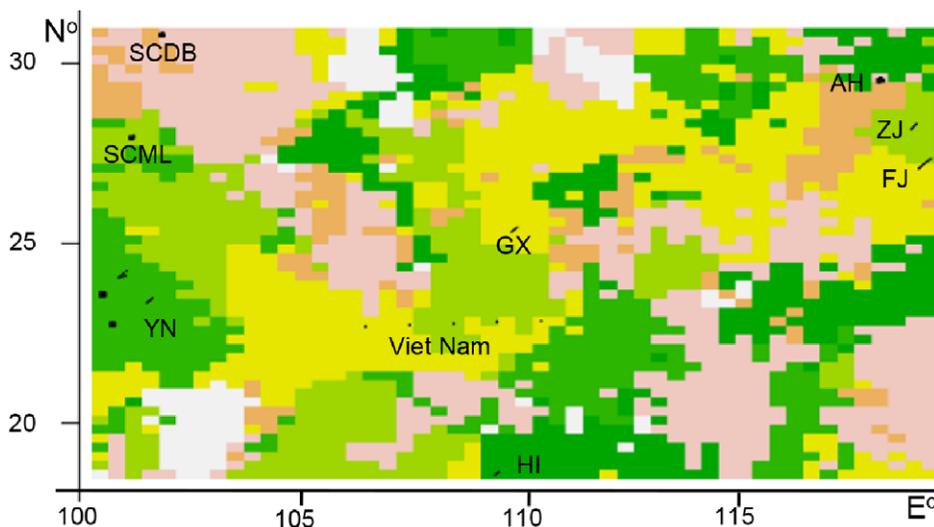


Figure 7. Population structure inferred in GENELAND at K=7. The black dots or lines indicate the sampling locations. The abscissa and ordinate show the coordinates of sampling locations. doi:10.1371/journal.pone.0055315.g007

Table 5. Pairwise F_{ST} estimates based on STR data from populations of *Macaca mulatta*.

	YNSM	YNMJ	YNZY	YN	SCML	SCDB	SCJY	AH	HI	ZJ	GX	Viet Nam	FJ
YNJG	0.008*												
YNSM	-0.002	0.026*											
YNMJ	0.005	0.012	-0.001										
YNZY	-0.002	0.004	-0.002	-0.002									
YN	0.008*	0.023*	0.004	0.026*	0.012								
SCML	0.053*	0.070*	0.045*	0.049*	0.050*	0.079*							
SCDB	0.026*	0.025*	0.041*	0.042*	0.021	0.013	0.077*						
SCJY	0.080*	0.084*	0.123*	0.116*	0.082*	0.099*	0.117*	0.123*					
AH	0.010	0.013	0.037	0.015	0.011	0.032	0.050*	0.053	0.073*				
HI	0.025*	0.036*	0.026	0.041*	0.022	0.018	0.065*	0.035	0.015	0.017			
ZJ	0.041*	0.038*	0.018	0.045*	0.002	0.033*	0.072*	0.067*	0.129*	0.070*	0.061*		
GX	-0.008	0.001	0.008	0.014	-0.013	0.006	0.063*	0.023	0.067*	-0.005	0.010	0.034	
Viet Nam	0.009	0.020*	0.026	0.035*	0.006	0.018*	0.075*	0.038*	0.029*	0.006	0.005	0.022	-0.012

Asterisks indicate statistical significance ($\alpha = 0.05$) based on sequential Bonferroni correction for multiple tests [92].
doi:10.1371/journal.pone.0055315.t005

one troop to join another [37]. Because genetic recombination and sorting may obscure dispersal patterns, we mainly discuss the history of dispersal in Chinese *M. mulatta* from the perspective of females.

For the wild population of the Chinese rhesus macaque, we expect IBD because distances between the sampled populations exceed 2000 km. The significant overall IBD suggests that geographic structuring owes to geographic proximity of the samples alone, i.e. that the populations have been present and stable for a long period of time. However, EAST and WEST exhibit different structural patterns. Specific locality data are available for six western populations and an IBD analysis of these data also detects significant IBD. Thus, we cannot reject the null hypothesis of IBD.

Two relatively independent associations of matriline occur in Yunnan and Sichuan (Figure 3): WEST B and WEST A. Only ca. 350 km separates the two populations from Sichuan (SCML and SCDB) and geographically they are closer to one another than either is to samples from Yunnan. However, these two populations exhibit the greatest amount of genetic divergence among all western localities and they do not appear to be part of the same gene pool ($F_{ST} = 0.9837$). The haplotype network (Figure 2) is congruent with this discovery.

Geographical effects other than IBD are also clear. Locality SCML contains three haplotypes which includes an endemic haplotype and two haplotypes shared with other areas. All of these

haplotypes are associated with a more diverse suite of sub-haplogroup WEST B in Yunnan. In contrast, SCDB contains only two haplotypes, one of which consists of a single individual. Further, the haplotypes of SCDB are located in the subhaplogroup WEST A, which comprise the main haplotypes of Sichuan (mainly from Heishui, Xiaojin, and Jiulong counties) as reported by Li et al. [24]. Thus, the current distribution of *M. mulatta* in Sichuan may be a consequence of two dispersal events, which is analogous to the dispersal history of India rhesus macaques [9]. SCDB, which is one of the populations in northwestern Sichuan, represents the first dispersal. The second dispersal involves south-central and eastern Sichuan (e.g. SCML). The topography of western Sichuan involves high mountains and large rivers, and this has probably played an important role in the two dispersal events via limiting gene flow and fostering genetic drift, both of which increase the pairwise genetic distances between populations. In this case, IBD based on straight-line distances fails to detect the pattern. For WEST, some geographic structuring must occur apart from that associated with haplogroup IBD. A rapid population expansion might indicate a historical localized extirpation. However, for WEST, we do not detect significant population growth from either the mismatch distribution analysis or the neutrality test.

The matrilineal pattern in EAST differs from that of WEST. A meaningful analysis of IBD is not possible because of only four specific sampling sites. Notwithstanding, the 35 haplotypes have a different pattern and this rejects the null hypothesis of all populations being equal. Samples from Zhejiang, Fujian, and Guangxi (China) and Viet Nam contain a large mix of haplotypes (Figure 2). Nearby provinces do not possess exclusive haplogroups. Here, haplotypic diversity is extreme. For example, 21 of 26 sampled individuals in Guangxi have unique haplotypes, as do 11 of 12 individuals from Viet Nam. This pattern differs drastically from that observed in WEST, where multiple individuals have the same haplotype. In addition, haplotypes from these four provinces occupy a central location in the EAST haplogroup suggesting greater heterogeneity and an earlier history of divergence (Figure 2) than for the WEST. Hunan (one haplotype), Hainan (one haplotype), Anhui (four haplotypes), and Henan (three haplotypes) occupy peripheral locations in the haplotype network, suggesting

Table 6. Summary of the effective population sizes and migrate rates for Chinese rhesus macaques.

Parameter	Bayesian Analysis				Maximum likelihood		
	2.50%	Mode	97.50%	Mean	2.50%	MLE	97.50%
Θ_{west}	0.090	0.098	0.100	0.094	1.130	1.182	1.237
Θ_{east}	0.073	0.098	0.100	0.090	1.115	1.201	1.294
$M_{east>west}$	71.333	115.000	194.000	128.052	1.617	1.981	2.396
$M_{west>east}$	207.333	341.000	516.667	349.862	1.919	2.374	2.902

doi:10.1371/journal.pone.0055315.t006

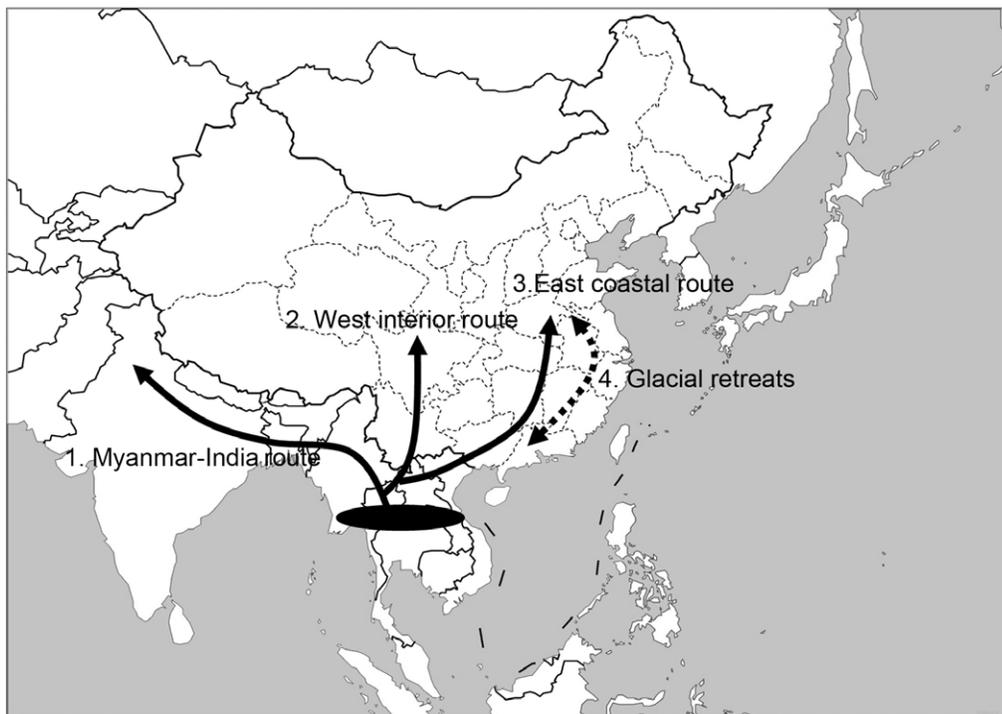


Figure 8. Biogeographic scenario for *Macaca mulatta*. The species originated in or near the northern part of the Indochinese peninsula, probably around 2.31 Ma. Two different dispersal routes occurred during a rapid northward expansion: the Myanmar-India route and Viet Nam-China route. Pleistocene glaciation may have affected the eastern haplogroup more than the western haplogroup.
doi:10.1371/journal.pone.0055315.g008

they could be founded by recent dispersal events following the last glacial maximum (LGM). However, the former two samples are from closed, protected populations, and the latter includes three localities. Thus, our sampling strategy is likely responsible for the absence of mixed haplotypic groups on a province-wide basis, with the possible exception of the highly diverged haplogroup on Hainan Island. Highly diverged haplogroups of treefrogs also occur on Hainan Island [80].

Our STRUCTURE analysis of microsatellite data does not detect significant population structure (Figure 6). This result is congruent to the spatially explicit clustering model in GENELAND. Although GENELAND analyses resolve a K of 7, the posterior probability values for different population memberships are not high enough strongly support inclusion of an individual in one population over another. Table 6 yields two points: the effective population sizes of WEST and EAST are nearly identical and thus we cannot reject the null hypothesis. This suggests that colonization of the two regions may have occurred at the same time scale, which corresponds to the time estimates (Figure 3). Further, analyses using MIGRATE reject the null hypothesis of unbiased migration. Migration and historical gene flow occurred predominately from WEST to EAST versus EAST to WEST, which suggests dispersal is more common in lowland areas than highlands. Pleistocene climatic cycling may drive the latter case.

Estimated Divergence Times and Dispersal

We estimate the matrilineal split between *M. mulatta* and its sister species *M. fascicularis* at 2.31 Ma (1.51–3.15, 95%). This date is similar to prior estimates of 2.6 Ma [20] and 2.5 Ma [17]. The MRCA of *M. mulatta* dates to 1.61 Ma (1.06–2.19, 95%), which is congruent with the 1.2 Ma in a previous study [17]. We estimate the age of the MRCA of WEST at 0.92 Ma (0.60–1.27, 95%) and

for EAST to be 0.81 Ma (0.50–1.11, 95%). This time estimate differs substantially from the date of divergence between Indian and Chinese rhesus macaques estimated from SNPs as being 0.16 Ma [28]. The comparisons of splits within *M. mulatta* are not equivalent for three reasons: they occur for different events; one is based on a single locus and the other on multiple loci; and our estimate is for the historical split of the matriline only and we note that our study demonstrates different sex-biased patterns. Perhaps most important, the SNP data estimate the cessation of biparental gene flow between the two groups as mediated by male dispersal. Thus, it is likely the difference reflects inheritance mechanisms of the two genomes, i.e. the maternally inherited mtDNA genome versus nuclear biparentally inherited SNPs and male-mediated gene flow.

Our time estimates yield a possible scenario of the dispersal history for *M. mulatta*. *Macaca mulatta* likely diverged from its ancestor shared with *fascicularis* about 2.31 Ma in or near northern Indochina. About 1.61 Ma (1.06–2.19, 95%), *M. mulatta* appears to have become widely distributed via a northward expansion using two different dispersal routes: Myanmar–India and Viet Nam–China (Figure 8). Dispersal into Myanmar and the Indian subcontinent is likely to have occurred along the Bay of Bengal. The second route appears to follow the eastern coast, arriving first in northern Viet Nam followed by China. Subsequent divisions are responsible for the formation of WEST and EAST.

Haplogroups WEST and EAST appear to have diverged about 1.15 Ma (0.78–1.55, 95%), a time of great change in Chinese and Southeast Asian environments. The last orogenic push of the QTP is associated with the formation of Asian monsoons, the beginning of Chinese loess (aeolian sedimentary deposits), and the appearance of the Yellow River [1]. These changes are likely to have

influenced the distribution and, thus, the genetic structure of Chinese rhesus macaques.

During the Quaternary Period, when climatic deterioration compressed suitable habitats southwards, *M. mulatta* seems to have dispersed southwards and repeatedly supplanted local matriline. This scenario best explains the great extent of mitochondrial DNA sequence differentiation between WEST and EAST. During the Pleistocene, two refugia might have existed: the western montane areas of southwestern China, and montane eastern China [81,82]. These two refugia may have promoted differentiation between matriline WEST and EAST.

Glaciation appears to have more severely affected EAST than WEST. The environment of East China experienced dramatic changes during the Quaternary [83]. The permafrost of North China expanded southwards by about 10° N latitude and reached the location of the Great Wall. The mean annual temperature lowered about 10°C to 12°C. The ice-age “mammoth fauna” roamed northern China southwards to the estuary of the Yangtze River. As sea-levels lowered about 140 m at times of maximum glaciation, the shoreline expanded eastwardly about 600 km [83].

We detect multiple mtDNA haplogroups in Zhejiang, Fujian, and Guangxi, China and in Viet Nam. The absence of overall geographic structuring and the apparent absence of IBD may have resulted from either the southward retreat of multiple haplogroups to a refuge during glacial periods, or a population expansion during interglacial periods [18,29]. However, it is also possible that our sampling strategy and human translocations influenced the patterns. Explanations involving natural environmental changes and anthropogenic influences have advantages. The former may more reasonably explain population expansions, as indicated by our mismatch distribution analysis. And the latter may provide a better explanation given the long, traditional trade in rhesus macaques in coastal regions of southeastern China, especially in Guangdong.

Implications for the Conservation of *M. mulatta* in China

The most serious threats to *M. mulatta* come from human habitat destruction and illegal poaching. These activities are responsible for the protection of the species as listed in Category II of the Chinese Wildlife Protection Act (1989) and in Appendix II of the Convention on the International Trade in Endangered Species of Fauna and Flora (CITES) [84]. For conservation, it is imperative that taxonomy reflect genetic and historical diversity. After all, the goal of conservation is to protect the evolutionary potential of a species, i.e. diversity, and bad taxonomy kills [85,86,87]. Consequently, future initiatives should determine whether the genetic history of *M. mulatta* corresponds to the traditional subspecific taxonomy or not, and this initiative is underway.

References

- Li JJ, Fang XM, Pan BT (2001) Late Cenozoic intensive uplift of Qinghai-Xizang plateau and its impacts on environments in surrounding area. *Quaternary Sciences* 21: 381–391.
- Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247–276.
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.
- Frenzel B (1968) Pleistocene vegetation of northern Eurasia—recent vegetation of northern Eurasia resulted from a relentless contest between steppe and forest. *Science* 161: 637–649.
- Milne RI, Abbott RJ (2002) The origin and evolution of Tertiary relict floras. *Advances in Botanical Research* 38: 281–314.
- Wen J (1999) Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. *Annual Review of Ecology and Systematics* 30: 421–455.
- Wolfheim J (1983) *Primates of the world: distribution, abundance and conservation*. Seattle, WA: University of Washington Press.
- Zhang Y, Quan G, Zhao T, Southwick C (1991) Distribution of macaques (*Macaca*) in China. *Acta Theriologica Sinica* 11: 171–185.
- Smith DG, McDonough J (2005) Mitochondrial DNA variation in Chinese and Indian rhesus macaques (*Macaca mulatta*). *American Journal of Primatology* 65: 1–25.
- Delson E (1980) Fossil macaques, phyletic relationships and a scenario of deployment. In: Lindburg D, editor. *The Macaques: Studies in Ecology, Behavior, and Evolution*. New York: Van Nostrand Reinhold Co. 10–30.
- Abegg C, Thierry B (2002) Macaque evolution and dispersal in insular Southeast Asia. *Biological Journal of the Linnean Society* 75: 555–576.
- Zhang YQ, Jin CZ, Masanaru T (2010) A partial skeleton of *Macaca* (Mammalia, Primates) from the early Pleistocene Queque cave site, Chongzuo, Guangxi, South China. *Veterebrata Palasiatica* 48: 275–280.

The western and eastern haplogroups indicate an important level of genetic structuring. Thus, WEST and EAST might be regarded as separate evolutionarily significant units (ESUs). These groups are experiencing shrinking distributions, increasing genetic isolation, and decreasing haplotype diversity. Attention should be paid to conserving populations in Henan and Hainan because both have very low genetic diversity, small residual areas of habitat, and significant morphological differences [8,88,89]. Further, the population in Henan occurs at the highest latitude of this species.

Supporting Information

Figure S1 A maximum parsimony tree for wild Chinese rhesus macaques, *Macaca mulatta*, derived from 214 D-loop haplotypes. Four outgroup haplotypes were used to root the tree. Bootstrap support from maximum parsimony (1000 replicates) ($\geq 50\%$ retained) were shown at the nodes. (TIF)

Table S1 The distribution of 214 D-loop haplotypes in each sampling location of wild Chinese rhesus macaques, *Macaca mulatta*. The first 92 haplotypes were sequenced in this study and the last 122 haplotypes were downloaded from GenBank. The 67 haplotypes colored in red were selected for concatenation with other mtDNA genes. (XLS)

Table S2 Statistics of genetic diversity, neutrality test and the mismatch distribution analysis. These analyses were performed for main populations of *Macaca mulatta* in China. SSD = Sum of Squares Deviations. (DOCX)

Table S3 The STR data for the wild Chinese rhesus macaque, *Macaca mulatta*. (XLSX)

Acknowledgments

We are very much appreciative for the insightful and helpful comments and suggestions from this manuscript’s two anonymous reviewers. Guo Cai and Hao-feng Wang provided assistance in sample collection. Technical support from colleagues in the lab of YPZ is gratefully acknowledged.

Author Contributions

Conceived and designed the experiments: YPZ BSY SJW JL. Performed the experiments: SJW JL YQW QQL. Analyzed the data: SJW JL QQL. Contributed reagents/materials/analysis tools: SJW JL QQL SFW CB. Wrote the paper: SJW JL RWM CB.

13. Gu Y (1980) A Pliocene macaque's tooth from Hubei. *Vertebrata Palasiatica* 18: 324–326.
14. Melnick DJ, Hoelzer GA (1996) The population genetic consequences of macaque social organisation and behaviour. In: Fa JE, Lindburg DG, editors. *The Evolution and Ecology of Macaque Societies*. Cambridge: Cambridge University Press. 413–443.
15. Morales JC, Melnick DJ (1998) Phylogenetic relationships of the macaques (Cercopithecidae: *Macaca*), as revealed by high resolution restriction site mapping of mitochondrial ribosomal genes. *Journal of Human Evolution* 34: 1–23.
16. Tosi AJ, Morales JC, Melnick DJ (2000) Comparison of Y chromosome and mtDNA phylogenies leads to unique inferences of macaque evolutionary history. *Molecular Phylogenetics and Evolution* 17: 133–144.
17. Tosi AJ, Morales JC, Melnick DJ (2003) Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution* 57: 1419–1435.
18. Chu JH, Lin YS, Wu HY (2007) Evolution and dispersal of three closely related macaque species, *Macaca mulatta*, *M. cyclopsis*, and *M. fuscata*, in the eastern Asia. *Molecular Phylogenetics and Evolution* 43: 418–429.
19. Fooden J, Albrecht GH (1999) Tail-length evolution in *fascicularis*-group macaques (Cercopithecidae: *Macaca*). *International Journal of Primatology* 20: 431–440.
20. Zhang YP, Shi LM (1993) Phylogeny of rhesus monkeys (*Macaca mulatta*) as revealed by mitochondrial DNA restriction enzyme analysis. *International Journal of Primatology* 14: 587–605.
21. Deinard A, Smith DG (2001) Phylogenetic relationships among the macaques: evidence from the nuclear locus NRAMP1. *Journal of Human Evolution* 41: 45–59.
22. Doxiadis GGM, Otting N, de Groot NG, de Groot N, Rouweler AJM, et al. (2003) Evolutionary stability of MHC class II haplotypes in diverse rhesus macaque populations. *Immunogenetics* 55: 540–551.
23. Hayasaka K, Fujii K, Horai S (1996) Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. *Molecular Biology and Evolution* 13: 1044–1053.
24. Kanthaswamy S, Smith DG (2004) Effects of geographic origin on captive *Macaca mulatta* mitochondrial DNA variation. *Comparative Medicine* 54: 193–201.
25. Kyes RC, Jones-Engel L, Chalise MK, Engel G, Heidrich J, et al. (2006) Genetic characterization of rhesus macaques (*Macaca mulatta*) in Nepal. *American Journal of Primatology* 68: 445–455.
26. Smith DG (2005) Genetic characterization of Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). *Comparative Medicine* 55: 227–230.
27. Smith DG, George D, Kanthaswamy S, McDonough J (2006) Identification of country of origin and admixture between Indian and Chinese rhesus macaques. *International Journal of Primatology* 27: 881–898.
28. Satkoski J, George D, Smith DG, Kanthaswamy S (2008) Genetic characterization of wild and captive rhesus macaques in China. *Journal of Medical Primatology* 37: 67–80.
29. Li DY, Xu HL, Smith DG, Cheng AC, Trask JS, et al. (2011) Phylogenetic analysis of Chinese rhesus macaques (*Macaca mulatta*) based on mitochondrial control region sequences. *American Journal of Primatology* 73: 883–895.
30. Smith DG, Kanthaswamy S, Viray J, Cody L (2000) Additional highly polymorphic microsatellite (STR) loci for estimating kinship in rhesus macaques (*Macaca mulatta*). *American Journal of Primatology* 50: 1–7.
31. Nurnberg P, Saueremann U, Kayser M, Lanfer C, Manz E, et al. (1998) Paternity assessment in rhesus macaques (*Macaca mulatta*): multilocus DNA fingerprinting and PCR marker typing. *American Journal of Primatology* 44: 1–18.
32. Kayser M, Ritter H, Bercovich F, Mrug M, Roewer L, et al. (1996) Identification of highly polymorphic microsatellites in the rhesus macaque *Macaca mulatta* by cross-species amplification. *Molecular Ecology* 5: 157–159.
33. Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, et al. (2007) Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. *Science* 316: 240–243.
34. Ferguson B, Street SL, Wright H, Pearson C, Jia Y, et al. (2007) Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). *BMC Genomics* 8: 43.
35. Street SL, Kyes RC, Grant R, Ferguson B (2007) Single nucleotide polymorphisms (SNPs) are highly conserved in rhesus (*Macaca mulatta*) and cynomolgus (*Macaca fascicularis*) macaques. *BMC Genomics* 8: 480.
36. Satkoski JA, Malhi R, Kanthaswamy S, Tio R, Malladi V, et al. (2008) Pyrosequencing as a method for SNP identification in the rhesus macaque (*Macaca mulatta*). *BMC Genomics* 9: 256.
37. Sade D (1972) A longitudinal study of the social behavior of rhesus monkeys. In: Tuttle RH, editor. *The Functional and Evolutionary Biology of Primates*. Chicago: Aldine-Atherton 378–398.
38. Hayaishi S, Kawamoto Y (2006) Low genetic diversity and biased distribution of mitochondrial DNA haplotypes in the Japanese macaque (*Macaca fuscata yakui*) on Yakushima Island. *Primates* 47: 158–164.
39. Vallet D, Petit EJ, Gatti S, Levvero F, Menard N (2008) A new 2C/TAB/PCI method improves DNA amplification success from faeces of Mediterranean (Barbary macaques) and tropical (lowland gorillas) primates. *Conservation Genetics* 9: 677–680.
40. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389–3402.
41. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596–1599.
42. Chatterji S, Pachter L (2006) Reference based annotation with GeneMapper. *Genome Biology* 7: R29.
43. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876–4882.
44. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
45. Swofford DL (2002) PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods). 4.0 Beta ed. Sunderland, MA: Sinauer Associates.
46. Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48.
47. Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* 141: 743–753.
48. Polzin T, Daneshmand SV (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* 31: 12–20.
49. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
50. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
51. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 1: 47–50.
52. Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics* 6: 13.
53. Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9: 552–569.
54. Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129: 555–562.
55. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
56. Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.
57. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.
58. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: 699–710.
59. Rambaut A, Drummond A (2005) Tracer v1.3: MCMC trace analysis tool. Institute of Evolutionary Biology, University of Edinburgh. Available: <http://tree.bio.ed.ac.uk/software/tracer/>. Accessed 2013 Jan 14.
60. Rambaut A (2006) FigTree: tree figure drawing tool, version 1.0. Institute of Evolutionary Biology, University of Edinburgh. Available: <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 2013 Jan 14.
61. Glaubitz JC (2004) CONVERT: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes* 4: 309–310.
62. Rousset F (2008) GENEPOP007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.
63. Goudet J (2002) FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices. 2.9.3.2. ed. Computer program and documentation distributed by the author. Available: <http://www2.unil.ch/popgen/softwares/fstat.htm>. Accessed 2013 Jan 14.
64. Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001–2014.
65. Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* 90: 502–503.
66. Rollins LA, Woolnough AP, Wilton AN, Sinclair R, Sherwin WB (2009) Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. *Molecular Ecology* 18: 1560–1573.
67. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
68. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
69. Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology* 18: 4734–4756.
70. Blair C, Weigel DE, Balazik M, Keeley AT, Walker FM, et al. (2012) A simulation-based evaluation of methods for inferring linear barriers to gene flow. *Molecular Ecology Resources* 12: 822–833.
71. Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics* 7: 295–302.

72. Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation—the correlated allele frequencies model revisited. *Bioinformatics* 24: 2222–2228.
73. Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics* 170: 1261–1280.
74. Guillot G, Mortier F, Estoup A (2005) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes* 5: 712–715.
75. Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22: 341–345.
76. Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152: 763–773.
77. Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the USA* 98: 4563–4568.
78. Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology* 13: 827–836.
79. Deinard AS, Lerche NW, Smith DG (2002) Polymorphism in the rhesus macaque (*Macaca mulatta*) NRAMP1 gene: lack of an allelic association to tuberculosis susceptibility. *Journal of Medical Primatology* 31: 8–16.
80. Blair C, Davy CM, Ngo A, Orlov NL, Shi HT, et al. (2012) Genealogy and demographic history of a widespread amphibian throughout Indochina. *Journal of Heredity*. doi:10.1093/jhered/ess079.
81. Harrison T, Ji XP, Su D (2002) On the systematic status of the late Neogene hominoids from Yunnan Province, China. *Journal of Human Evolution* 43: 207–227.
82. Zhang H, Yan J, Zhang GQ, Zhou KY (2008) Phylogeography and demographic history of Chinese black-spotted frog populations (*Pelophylax nigromaculata*): evidence for independent refugia expansion and secondary contact. *BMC Evolutionary Biology* 8: 21.
83. Li J, Shu Q, Zhou S, Zhao Z, Zhang J (2004) Review and prospects of Quaternary glaciation research in China. *Journal of Glaciology and Geocryology* 26: 235–243.
84. Timmins RJ, Richardson M, Chhangani A, Yongcheng L (2008) *Macaca mulatta*. In: IUCN 2010. IUCN Red List of Threatened Species. Version 2010.1. Available: <http://www.iucnredlist.org/details/12554/0>. Accessed 2013 Jan 14.
85. Daugherty CH, Cree A, Hay JM, Thompson MB (1990) Neglected taxonomy and continuing extinctions of tuatara (*Sphenodon*). *Nature* 347: 177–179.
86. May RM (1990) Taxonomy as destiny. *Nature* 347: 129–130.
87. Vieites DR, Wollenberg KC, Andreone F, Kohler J, Glaw F, et al. (2009) Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences of the USA* 106: 8267–8272.
88. Fooden J (2000) Systematic review of the rhesus macaque, *Macaca mulatta* (Zimmermann, 1780). *Fieldiana Zoology, New Ser. No. 96*, 1509: v, 1–180.
89. Groves C (2001) *Primate Taxonomy*. Washington, DC: Smithsonian Institution Press.
90. Li QQ, Zhang YP (2004) A molecular phylogeny of *Macaca* based on mitochondrial control region sequences. *Zoological Research* 5: 385–390.
91. Morin PA, Kanthaswamy S, Smith DG (1997) Simple sequence repeat (SSR) polymorphisms for colony management and population genetics in rhesus macaques (*Macaca mulatta*). *American Journal of Primatology* 42: 199–213.
92. Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43: 223–225.