



Published in final edited form as:

*J Med Primatol.* 2010 August ; 39(4): 252–265. doi:10.1111/j.1600-0684.2010.00430.x.

## Detecting signatures of inter-regional and inter-specific hybridization among the Chinese rhesus macaque Specific Pathogen-free (SPF) population using single nucleotide polymorphic (SNP) markers

Sree Kanthaswamy<sup>1,2,\*</sup>, Jessica Satkoski<sup>1</sup>, Alex Kou<sup>2</sup>, Venkat Malladi<sup>1</sup>, and David Glenn Smith<sup>1,2</sup>

<sup>1</sup>Department of Anthropology, University of California–Davis, CA 95616

<sup>2</sup>California National Primate Research Center (CNPRC), University of California–Davis, CA 95616

### Abstract

While rates of gene flow between rhesus and longtail macaque populations near their hybrid zone in Indochina have been quantified elsewhere, this study demonstrates that the inter-specific introgression is not limited to the Indochinese hybrid zone but is more geographically widespread. SNP analysis across the rhesus and longtail genomes shows evidence for inter-specific admixture between Chinese rhesus and mainland longtails, with implications for genetic diversity both in the Chinese super-SPF population at the California National Primate Research Center and other primate facilities. Eastern Chinese rhesus appeared more highly derived than western Chinese rhesus, and allele sharing between longtails and Chinese rhesus was not random with regard to geographic distance, but no significant nuclear genetic differences between eastern and western Chinese rhesus was detected among the 245 genic SNPs assayed. The implications of this inter-specific admixture for the use of Chinese rhesus and mainland longtail in biomedical research should be considered.

### Keywords

Genetic admixture; Population genetics; Phylogenetics; *Macaca mulatta*; *M. fascicularis*

### Introduction

The level 2 Specific Pathogen-free (SPF), or super-SPF (SSPF), colony at the California National Primate Research Center (NPRC) consists of 95 Chinese and 302 Indian rhesus macaques (*Macaca mulatta*) that descend from non-SPF Chinese and Indian-origin founders, respectively. Independently derived at different times and never interbred, both the Chinese and Indian SSPF animals are free of B virus, rhesus cytomegalovirus (RhCMV), rhesus rhadinovirus (RRV), simian foamy virus (SFV), simian immunodeficiency virus (SIV), simian type D retrovirus (SRV 1 through 5) and simian T-lymphotropic virus (STLV).

\*Correspondence: Sree Kanthaswamy, PhD, Department of Anthropology, University of California–Davis, Davis, California, One Shields Avenue, Davis, CA95616. Phone (530) 219-2017; FAX (530) 752-8885 skanthaswamy@ucdavis.edu.

Longitudinal analyses of genetic variation in these two populations permitted the assessment of the direct derivation of Indian and Chinese SSPF animals from conventional stocks at the California NPRC, which is different from SSPF derivation strategies employed at the other NPRCs, where SSPF animals are derived from the level 1 SPF stock [animals free of herpes virus-1 (B virus), simian immunodeficiency virus (SIV), Type D simian retrovirus (SRV) and simian T-lymphotropic virus (STLV-1)]. From chronologically stratified data, Kanthaswamy et al. (17) showed that the California NPRC Indian SSPF population contains animals with various proportions of Chinese ancestry as a result of the introduction of Chinese founders into the conventional colony since the 1980s [29]. This deliberate hybridization between eastern (Indian) and western (Chinese) regional varieties of rhesus macaques is unique to the California NPRC.

Because the California NPRC's Chinese SSPF population descends from animals imported directly from China, the presence of Indian-Chinese admixture has not been detected [17]. However, Satkoski et al. [30] concluded that widespread human-mediated inter-regional admixture has occurred among captive rhesus in China prior to their export to the US because wild-caught Chinese rhesus show a discrete regional variation in their mitochondrial DNA (mtDNA) with one haplogroup being largely restricted to eastern China, one to western China, one to southern China and a mixture of all haplogroups in central China, while the mtDNA variation in captive animals in Chinese regional breeding facilities does not exhibit this strict geographic pattern in the distribution of mtDNA haplogroups. Li et al. [19] have argued that rhesus macaques first entered southern China from Indochina and rapidly expanded eastward before dispersing into the rugged provinces of western China. This is significant because Smith and McDonough [32] showed that the genetic difference between the mtDNA of Chinese rhesus macaques from the eastern and western parts of China is approximately 70% as large as that between Indian and Chinese rhesus macaques, hence, potentially of biomedical importance. Since mtDNA is only a single locus and subject to much stronger stochastic effects than nuclear loci due to its lower effective population size, it is not known whether or not this difference is representative of the entire genome.

Kanthaswamy et al. [14] estimated that the magnitude of gene flow between Chinese (and Burmese) rhesus macaques and Vietnamese longtail macaques (*M. fascicularis*) was at least twice that between the geographically contiguous rhesus macaque populations of India and Nepal, demonstrating what is probably a long history of natural hybridization between rhesus and longtail macaques in Indochina. Other researchers have claimed that regions north of the Isthmus of Kra, i.e., west-central region of Thailand, southern Laos and central Vietnam, where both species have been marginally sympatric throughout most, if not all, of the Pleistocene, represents a zone of inter-species hybridization [6,8,9,11,34,37-39].

Based on the evidence for inter-regional hybridization among captive rhesus in China and natural mating between rhesus and longtails in Indochina, human-mediated crossbreeding between these species of macaques in captivity as well as effects of genetic introgression that range beyond the purported hybrid zone are both reasonable hypotheses: Therefore, if allele sharing between Chinese rhesus and longtails is not affected by geography, it suggests that the shared alleles are ancestral, and alternatively, if more alleles are shared in geographically contiguous areas than in distant populations, then it supports introgression.

Results from Satkoski et al. [30] and Kanthaswamy et al. [14] suggest that residual effects of past and present introgression between rhesus and longtail may exist in the Chinese SSPF colony at the California NPRC. Because the accuracy of population genetic inference increases with the degree of genomic coverage [1,16,31], the intent of this research is to study a large number of SNPs (i.e., 830 orthologous loci across both the rhesus and longtail

genomes and 1307 loci across the rhesus genome) to determine if the signatures of inter-specific and inter-regional admixture exist in the ancestry of the California NPRC's "pure blood" Chinese SSPF population and test the two hypotheses described above.

## Methods

Twelve distinct rhesus and longtail macaque populations, originating across mainland Asia and insular Southeast Asia were sampled for this study (Table 1). The relative geographic distributions of the two species are shown in Figure 1. One thousand three hundred and seven SNP genotypes were used to compare 36 Indian-origin and 37 Chinese-origin rhesus macaques whose mtDNA haplogroups had been assigned by Smith and McDonough [32] to Ind1 (N = 32), Ind2 (N = 4), ChiE (N = 14), ChiW1 (N = 8), ChiW2 (N = 10), ChiW3 (N = 4) and ChiS (N = 1). Ind2, the rarer form of the two Indian haplogroups, represents 5 to 10% of Indian rhesus macaques and is a derived lineage of the Burmese mtDNA haplogroup BURM [32]. We have inferred that the presence of Ind2 in India resulted from gene flow from Burma. If so, we might expect its members to exhibit greater evidence of past introgression from inter-species hybridization with longtail macaques than members of Ind1 unless linkage disequilibrium (LD) between mtDNA and the nuclear genome, which should follow the same pattern of LD between two unlinked nuclear loci, has completely dissipated, but no significant genetic differences between eastern and western Chinese rhesus macaques was detected among the 245 "genic" (located within a gene) SNPs assayed. Nuclear genetic differences between eastern and western Chinese rhesus macaques were not detected but the results show that each of the Chinese mtDNA haplotypes reflect relatively strict geographic distributions among wild rhesus populations in China [32] with ChiW, ChiE and ChiS being largely restricted to the western (Yunnan, Sichuan and Guizhou provinces), southern (Guangxi and Guangdong provinces) and eastern (all other provinces of China south of the Yellow River) parts of China, respectively. It is noteworthy that mtDNA haplotypes belonging to haplogroup ChiS [31], which is more closely related to ChiE than to ChiW [30,32], are also found in rhesus macaques from Vietnam, which shares a long border with Quangzhi province and might represent a route through which rhesus macaques reached China from Indochina [19]. The origin of individuals of each Chinese haplogroup is shown in Table 1. The presence of ChiS in both Vietnam and southern China and its close relationship to ChiE, which exhibits the greatest geographic diversity of all mtDNA haplogroups in China suggest an eastward dispersal of rhesus macaques from southwestern China and the derivation of ChiW from ChiE. The very close similarity of rhesus macaques from Hainan Island to ChiE haplotypes suggests that this dispersal occurred as recently as early Holocene times, because Hainan Island was permanently separated from the mainland (Guangxi and Guangdong provinces) as recently as 8,500 years ago [41]. If inter-species admixture persisted after this eastward dispersal, then we would expect eastern Chinese rhesus macaques to exhibit less inter-species admixture than western Chinese rhesus macaques. Otherwise, such inter-species admixture must predate 8,500 years ago.

A subset of the 1,307 SNPs, numbering 830 loci, were used to compare 24 longtail macaques from Vietnam (N = 5), Peninsular Malaysia (N = 4), Indonesia (N = 3), Mauritius (N = 6) and the Philippines (N = 6). The individuals sampled for this study comprise a mixture of individuals purchased from commercial importers and breeders as well as animals bred and housed at primate centers, specifically California NPRC, Oregon NPRC, Wisconsin NPRC and the Caribbean Primate Research Center (CPRC) (see Table 1 for information on sample size, geographic origin, importer/breeder and current location of the animals included in this study).

The SNPs described here are part of the much larger group of markers discovered by Malhi and colleagues [21]. Genotypes were collected at the University of California, Davis

Genome Center using the Illumina (San Diego, CA) GoldenGate custom genotyping assay on a BeadXpress Reader. The collection of rhesus SNP genotypes is described in detail in Satkoski et al. [31]. To genotype the longtail macaques, raw fluorescence data was imported into the Illumina BeadStudio software and genotype calls were made using BeadStudio cluster files generated with rhesus macaque samples. Loci producing fluorescence either above or below the range generated by rhesus macaque DNA were eliminated from the analysis. Thus, although the longtail macaques were amplified at all 1307 loci, only 830 of these loci met the genotype standards set for rhesus macaques. Of these 830 SNPs, all polymorphisms were segregating in at least one rhesus population, while only 323 of them were segregating in at least one longtail population.

The 830 common SNPs were localized in the rhesus genome using the genome BLAST function of the National Center for Biotechnology Information website [www.ncbi.nlm.nih.gov]. The loci were classified as either genic or “nongenic” (located at least 1 nucleotide position outside of a gene). Of these, 245 loci were genic and 585 were nongenic (see Appendices 1 and 2). Two genic SNPs and six non-genic SNPs are located on the X chromosome.

We used genotypic data collected from 96 of the 97 rhesus and longtail macaques to compute observed and expected heterozygosities [24] for each of the 11 of the 12 geographic populations with Plink version 1.06 [27]. Pairwise  $F_{st}$  values [40] were estimated using the web version of GENEPOP [28] to assess population level differentiation based on the same sample set of 96 animals. Pairwise  $F_{st}$  values were calculated using the 1307 SNP dataset for comparisons of rhesus by geographic origin while pairwise  $F_{st}$  values were calculated with the 830 SNP dataset for comparisons of rhesus and longtail macaques by breeder/importer and by geographic origin. Heterozygosity and pairwise  $F_{st}$  could not be calculated for the twelfth population (from southern China) because it contained only one individual, but was used in all subsequent analyses.

The program STRUCTURE 2.3.1 [26] was used to infer the genetic structure among the populations. Two separate analyses were run, both conducted with  $20^3$  MCMC iterations after a  $20^3$  iteration burn in. *A priori* population information was disregarded, and correlation among allele frequencies under an admixture model was assumed. In the first analysis of only rhesus macaques, 1307 SNP loci were included and individuals were grouped with regard to geographic origin and the breeding center/importer from which they were acquired. Between 1 and 9 ( $0 < K < 10$ ) populations were assumed, and the simulation was repeated 10 times for each population value ( $K$ ). A log normal probability of fit between the model and the data was assessed for each simulation and the population value with the best fit was chosen for interpretation. In the second analysis, genotypes for only the 830 common loci were included, and both rhesus and longtail macaques were grouped according to their geographic origin. Between 1 and 12 populations were assumed, and the simulation was again repeated 10 times for each population value.

## Results

The estimates of average OH and EH are presented in Tables 2 and 3 for the 1307 SNPs in the rhesus macaque genome and the 830 SNPs in the rhesus and longtail macaque genomes, respectively. Based on the analysis of 1307 SNPs, the average OH and EH across the rhesus populations are 0.23 and 0.24, respectively. Somewhat surprisingly, average estimates of OH (0.25) and EH (0.21) among the Indian rhesus were identical to those of the Chinese populations. Based on the subset of 830 SNPs, estimates of average OH and EH for the Indian populations were 0.23 and 0.24, respectively, while estimates of OH and EH for the Chinese rhesus macaques were also nearly identical (0.21) (Figure 2). The mean estimates of

OH and EH across all populations of rhesus were equal at 0.22. Not surprisingly, estimates of average OH (0.067) and EH (0.065) across the longtail macaque populations were more than three-fold lower than those estimated for rhesus macaque populations, reflecting the ascertainment bias introduced by genotyping longtail macaques for SNPs detected in another species. Estimates of OH and EH for all three insular populations (Indonesia, Mauritius, and the Philippines) of longtail macaques combined were 0.053 and 0.052, respectively, or more than four times lower than those estimated for rhesus macaques. In sharp contrast, the Vietnamese longtail macaque population exhibited OH and EH values that were intermediate between those for the rhesus and the insular longtail macaque populations.

Estimates of OH and EH based on 323 markers polymorphic in rhesus macaques were comparable to those estimates based on all 830 SNPs. However, the 323-marker estimates in the longtail macaque populations were two- or three-times greater than their 830-marker estimates. This is not surprising, because this higher heterozygosity estimate could be free of the influence from the 507 rhesus SNPs that were monomorphic in longtail macaques. As for the 830 rhesus SNPs, the OH and EH estimates for the Vietnamese longtail macaques were the highest among all macaque population samples when only the 323 polymorphic SNPs were considered. There were no appreciable differences between OH and EH estimates of genic and non-genic SNPs in either species of macaques (see Table 3).

Estimates of population differentiation (pairwise  $F_{st}$ ) based on the 1307 rhesus macaque SNPs among the five mtDNA haplogroups (two of Indian origin and three from China) of rhesus macaques are presented in Table 4. The range of differentiation was from -0.001 (between two ChiW haplogroups: ChiW 1 and ChiW2 from western China) to 0.236 (between Ind1, the predominant Indian haplogroup and ChiE, the eastern Chinese haplogroup). Unlike mtDNA, which is relatively highly divergent between eastern and western China, the SNPs show little differentiation between eastern and western China. However, some estimates of differentiation correspond to the isolation-by-distance pattern exhibited by rhesus mtDNA haplogroups even though genetic drift impacts mtDNA variation four times greater than nuclear DNA variation [25]. Despite the presence of Ind2 individuals throughout India, these animals cluster more closely with animals belonging to western Chinese haplogroups than the other Indian haplogroup, consistent with the hypothesis that Ind2 originated in Burma, which shares a long border with Yunnan province of western China. According to Smith and McDonough [32] the presence of the Ind2 haplogroup in India is evidence for introgression between Burmese and Indian rhesus because it is closely related to the BURM haplogroup in Burmese rhesus, and the analysis presented here supports that assertion. The surprisingly low distance between members of Ind2 and ChiS is consistent with persistent linkage disequilibrium between mtDNA and nuclear genes in members of Ind2, which originated in Burma, suggesting fairly recent, rather than ancient, introgression. Also consistent with isolation by distance, Ind1 rhesus macaques are more divergent from Chinese rhesus macaques with ChiE haplotypes than from those with ChiW haplotypes. The sharing of ChiS between Indochinese (e.g., Vietnamese) and southern Chinese (those from Guangxi and Guangdong provinces) rhesus macaques [19,31] further supports mtDNA distribution patterns that follow isolation by distance.

Table 5 presents pairwise  $F_{st}$  values based on the 830 SNPs that are shared between rhesus and longtail macaques. Genetic distances ranged from -0.07 (between the Oregon NPRC Ind2 animals and the LABS Ind1 animals) to 0.79 (between the LABS Ind1 rhesus and the Philippine longtails). The levels of differentiation among the geographic samples of rhesus as measured with these 830 SNPs are consistent with those obtained with 1307 SNPs above and again conform to an isolation-by-distance model under commonly accepted assumptions regarding dispersal of macaques in southeast Asia. Levels of differentiation between



populations of longtail macaques ranged from 0.01 (between the Peninsular Malaysian longtails and the Indonesian longtails) to 0.31 (between Philippine and Vietnamese longtail macaques). It is noteworthy that some populations of longtail macaques (e.g., Mauritian from Philippine, Vietnamese and Malaysian and Mauritian from Vietnamese) are as divergent from each other as are Indian and Chinese rhesus macaques, confirming inferences based on mtDNA [33]. Both Philippine and Mauritian longtail macaques are more similar to Indonesian than to other populations of longtail macaques. Longtail macaques from Vietnam were by far the most divergent of the longtail macaque populations studied. Measurements of inter- and intra-population differentiation using SNPs were generally concordant with those based on mtDNA [33] and STRs [14], but estimates with SNPs reflected much greater differentiation among population pairs than either mtDNA or STRs. Table 6 presents pairwise differences between populations of rhesus and longtails based on 245 genic and 585 non-genic SNPs. The mosaic pattern of distance estimates among the genic and non-genic SNPs could possibly reflect the effects of differential selection on functional loci and selective sweeps on neutral markers [18].

Table 7 shows the average proportion of shared alleles between all population pairs across the 245 genic SNPs. For example, if both populations are polymorphic, the proportion of shared alleles would be 100%, while if one population was polymorphic and the other monomorphic, the proportion of shared alleles would drop to 50%. Populations monomorphic for different alleles would, of course, share 0% of alleles.

Table 1 displays the key to Figures 3 and 4, which illustrate the population structure of rhesus and longtail macaque SNPs with maximum log normal probabilities for K values of 4 and 5, respectively. Based on 1307 SNP genotype data, STRUCTURE separated the Indian-derived and Chinese origin animals discretely (Figure 3). Indian rhesus macaques exhibit structure that potentially represents introgression of Burmese nuclear genes into India. The Chinese animals, regardless of their mtDNA haplogroup affinities or source (California NPRC, Three Spring Scientific and Valley Biosystems), exhibited heterogeneous gene pools with similar levels of genetic substructure, suggesting that structure does not result from varying levels of inter-regional admixture between eastern and western Chinese rhesus macaques. The STRUCTURE analysis, however, does show evidence for contribution from a second population that is generally lower in eastern than western Chinese rhesus. If signals of the second population detected in the Chinese rhesus are actually longtail or Ind2 rhesus macaques in origin, this demonstrates higher longtail introgression in western versus eastern Chinese rhesus.

One ChiW sample from the California NPPRC was more similar to the ChiE samples and another ChiW sample from VBS exhibited some SNPs that are probably of Indian origin; both instances probably resulted from unrecognized admixture in the two facilities' colonies. The Cayo Santiago rhesus population from the CPRC, which descend from 41 original founders from Lucknow and has been closed to gene flow since 1936, forms a distinctively homogenous group compared to all other Indian-derived rhesus populations. Based on their SNP composition, animals of Indian ancestry, including those from the University of Miami, LABS and the Wisconsin NPRC, exhibited variable (although small) proportions of ancestry contributed by Chinese rhesus. While the Wisconsin Ind2 animal did not carry any Chinese SNPs, all of the Ind 2 animals from the Oregon NPRC exhibited small to moderate amounts of Chinese ancestry.

The STRUCTURE analysis of 12 geographically representative samples of rhesus and longtail macaques based on 830 SNPs is given in Figure 4. Figure 4 contains three large population clusters (longtail macaques, Chinese rhesus macaques and Indian rhesus macaques) with additional influence, which could reflect introgression, from two currently

unidentifiable populations in some individuals. Longtail macaques from Vietnam and Malaysia as well as a rhesus macaque from western China belonging to haplogroup ChiW1 were partially assigned to the same population. Some of the rhesus macaques belonging to this mtDNA haplogroup were also partially assigned to the same population as Indian rhesus macaques belonging to haplogroups Ind1 and Ind2. Several Ind2 animals, especially those from the Oregon and Wisconsin NPRCs, were partially assigned to the same population as were many Chinese rhesus macaques. The longtail macaques from Indonesia, Mauritius and the Philippines appeared more genetically homogenous than the other groups.

## Discussion

As SNPs are more mutationally stable than STRs, they represent more plesiomorphic (ancestral) than apomorphic (derived) polymorphisms, while STR alleles are more likely to be apomorphic. Thus, distance estimates based on STRs tend to asymptote over time due to their higher rates of homoplasies [character state reversals and parallel or convergent evolution, [4,10,20] causing STR-based inferences on ancient introgression and differentiation to be imprecise [15]. Conversely, ascertaining current population structure based on SNPs is more subjected to incomplete lineage sorting and balancing selection [3,34].

Some of the SNPs studied here are within coding regions (genic SNPs) of the macaque genome, and may have been influenced by selection, while other SNPs in the present study which are in neutral regions of the genome (or non-genic SNPs) may have experienced genetic hitchhiking, i.e., when neutral mutations spread through populations because they are linked to genes under positive selection [18]. While such selective pressures could potentially mask interactions between gene flow and selection, the density of genome-wide coverage by SNPs should provide a better resolution for detecting signals of introgression and population subdivision with regard to biomedically-relevant genes or gene clusters, notwithstanding the bias introduced by the SNPs used in this analysis having been ascertained first in rhesus macaques.

The estimates of genetic diversity (OH and EH) among the Indian and Chinese rhesus macaques were nearly identical. This was unexpected because previous assessment based on mtDNA [33] and STR loci [14,22,30] have revealed greater variation among Chinese rhesus macaques. While the results of the present study are in agreement with the lower estimates of allozyme diversity in Chinese rhesus [23], Morin et al. [22] attributed the discrepancies between distance estimates from coding and non-coding loci to ascertainment biases that were introduced during the screening of Indian rhesus for protein polymorphisms first identified in Indian rhesus macaques, but that is not the case with the rhesus SNPs used here. While studies of mtDNA reveal a much greater level of homogeneity in Indian than in Chinese rhesus macaques, the much lower effective population size of mtDNA renders it far more vulnerable than nuclear genes to stochastic effects leading to loss of genetic heterogeneity but does not explain the greater level of genetic homogeneity at STR loci found in Indian than Chinese rhesus macaques. As described in Malhi et al. [21], biases in ascertaining rhesus SNPs were minimized by identifying SNPs by comparing DNA of a rhesus sample from western China with the draft genome of the Indian rhesus macaque. It has been noted, however, [Satkoski et al. unpublished data], that this process biases towards the discovery of common polymorphism, leaving rare polymorphisms undiscovered. Thus, while estimates of heterozygosity of STRs of Chinese rhesus macaques are only slightly higher than those in Indian rhesus macaques, the average number of alleles per STR for Chinese rhesus macaques is nearly twice that for Indian rhesus macaque [31]. Were Chinese rhesus macaques to carry significantly more rare polymorphisms than Indian rhesus macaques, as these data suggest and as might result from the genetic bottleneck the Indian

rhesus macaques are hypothesized to have experienced [12], OH and EH could be underestimated in Chinese populations using the SNPs included in the present study which focused on more common SNPs. In fact, other strategies of SNP discovery have led to estimates of genetic diversity that show higher diversity in Chinese than in Indian rhesus macaques [7,12].

The identification of the SNPs reported here in the rhesus genome has certainly introduced a bias to the rhesus-longtail population genetic comparisons reported here. Thus, none of the SNPs we studied are unique to longtail macaques, and the presence of these SNPs in longtail macaque populations could be due either to shared ancestral variation (in the case of Insular longtail macaque populations) and/or introgression (in the case of mainland longtail macaque populations). These biases are reflected in the lower proportion (830 out of 1307) of orthologous loci shared by the two macaque species, only 323 (fewer than half) of which represent polymorphism present in both species, and the generally lower genetic diversity levels (with the exception of the Vietnamese longtails, which exhibited the highest genetic diversity estimates of any longtail population) relative to the rhesus populations. It is interesting that computations of genetic diversity among the rhesus and longtails based on STRs did not reflect any appreciable biases even though most of those loci were first developed in Indian and Chinese rhesus [14,33]. This could be due to the fact that STRs are non-coding neutral loci.

Pairwise  $F_{st}$  estimates based on 830 SNPs were far greater than those based on STRs. The downward bias in STR-based divergence estimates might result from homoplasy [20] rather than the homogenizing effects of ongoing gene flow. This is especially likely for the insular longtail macaque populations among which, despite significant natural barriers to gene flow, consistently lower levels of differentiation were estimated from STRs than from SNPs. In addition to genetic drift, local adaptation and/or dispersal patterns could have also caused the geographic patterns in the macaque population differentiation estimates that are generally consistent with isolation by distance as exemplified by the greater distances between Indian rhesus macaques and Chinese rhesus macaques from eastern than from western China and the greater similarity among rhesus macaque from the southern Chinese province of Guangdong, Vietnam and Indian rhesus macaques with Ind2 than any shared with Indian rhesus macaques with Ind1.

The lower divergence of both Mauritian and Philippine longtail macaques from Indonesian longtail macaques is consistent with their hypothesized Indonesian origins via ship-borne transportation and rafting from NE Borneo via either Palawan and/or the Sulu Archipelago, respectively. The markedly high level of divergence of Vietnamese longtail macaques from all insular populations of longtail macaques probably results both their length of isolation from insular longtail populations as well as introgression from rhesus macaques. The supposition that Vietnamese longtail macaques are more similar to Indonesian than to any other population of longtail macaques is consistent with the hypothesized Indonesian origin of the species.

With 1307 SNP markers, Indian rhesus belonging to the Ind2 haplogroup could be discriminated from those belonging to Ind1 and given the dispersion of Ind2 animals throughout India, probably suggests a separate origin in Burma with later migration into India probably causing SNPs that originated in Burma to exhibit residual LD with mtDNA. Satkoski et al. [31] also successfully differentiated Indian rhesus macaques with the Ind1 and Ind2 mtDNA haplogroups using 23 highly polymorphic STRs. Chinese rhesus macaques are more closely related to Indian rhesus macaques belonging to Ind2 than to those belonging to Ind1, probably because rhesus macaques with Ind2, which are more closely related to the mtDNA of Burmese rhesus macaques, share more SNPs originating in



mainland southeast Asian origin with Chinese rhesus macaques than with most Indian rhesus macaques.

The Indonesian and Peninsular Malaysian longtail macaques combined exhibit no structure based on the 830 SNPs. These populations may have become geographically partitioned as recently as the end of the last ice age (ca. 10 kYA). This estimated timeline for genetic divergence between these two populations is concomitant with Kanthaswamy et al.'s [13] timescale on the emergence of geographic subdivisions between Sumatran and Bornean orangutans that were probably due to the same changes in sea levels.

Despite the similarity between Indonesian and Malaysian longtail macaques, also reported from their mtDNA [33], Mauritian longtail macaques resembled the former much more closely than any other longtail macaque population. This observation substantiates the long-held notion that the Mauritian animals were founded by animals originating in Indonesia approximately one to two centuries ago [2,35,36] and is consistent with the conclusion of Tosi and Coke [39] that Mauritian longtail macaques are of Sumatran origin. The mainland longtail population is the most distant of all longtail macaque populations, as previously reported by several authors [3,14]. Like the Mauritian longtail macaques, the Philippine longtail macaques also exhibited strong founder effects leading to loss of genetic diversity, but their genetic relationships with the other insular and mainland longtails based on pairwise  $F_{st}$  estimates shows a strong relationship to geographic distance and suggests dispersion from Indonesia, rather than the mainland. These data conflict with the inferences from STR data reported by Kanthaswamy et al [14] that suggested that the Philippine longtails are more closely related to the mainland longtail macaques and underscores the danger in basing inferences on just one kind of data.

Genetic differentiation among some populations of longtail macaques equaled that between Indian and Chinese rhesus macaques, confirming what we have reported based on mtDNA and STRs [14,32]. Thus, care should be taken to insure that these genetic differences do not obscure the treatment effects in biomedical research experiments in which both regional varieties of longtail macaques are included. The assessments of genetic differentiation among the different rhesus and longtail populations in Table 7 reveals a prevalence of shared alleles between mainland macaque populations of the two species and an absence of these same alleles among island populations of longtails, suggesting that introgression of rhesus alleles into longtail populations occurs primarily in the wild hybrid zone, rather than through hybridization in captivity. Stevison and Kohn [34] argued that gene flow between the two species is asymmetric because signals of inter-specific gene flow could be eliminated by removing the Indochinese samples from their analysis. Tosi and Coke [39] showed that mainland longtail macaques carry rhesus-like Y-chromosome DNA, suggesting introgression of rhesus genes in mainland longtail macaques, but mtDNA resembling that of insular longtail macaque populations, suggesting the absence of introgression of longtail macaque DNA in rhesus macaques. In contrast, Kanthaswamy et al. [14] observed only a marginal drop in rates of gene flow when Vietnamese samples were removed. Dispersal rates of 3.4 to 6.9 migrants/generation were estimated between mainland rhesus and longtail populations and these rates are between two- and ten-times the migration rate among pairs of insular populations [14]. The somewhat lower level of structure for members of the ChiE mtDNA haplogroup would be consistent with the hypothesis that the source of that structure is introgression from longtail macaques.

The high migration rate between the mainland longtail and rhesus macaques however has not completely transformed the overall genetic structure of mainland rhesus macaques. The eastern rhesus (those from China) are more genetically similar to the western rhesus (those from the Indian subcontinent) than to Indochinese longtails, despite the fact that mtDNA of

Chinese rhesus is more closely related to that of *M. cyclopsis* and *M. fuscata* (both of which belong to the *fascicularis* group of macaque species together with rhesus and the longtail macaques) than to that of Indian rhesus macaques [5,32]. This discrepancy between phylogenetic inferences based on mtDNA and SNPs probably results from stochastic effects caused by the much lower effective population size of mtDNA than SNPs. On average, the genetic distances between each pair of rhesus populations estimated here are at least half the levels of differentiation between any rhesus macaque population and the Indochinese longtail population.

The past and ongoing natural gene flow between overlapping populations of rhesus and longtails and human-mediated crosses in captivity between eastern and western Chinese rhesus macaques have been postulated by Satkoski et al. [30] and Kanthaswamy et al. [14] as the source for increased genetic diversity among longtail and rhesus macaques currently supplied from Indochina and China, respectively, notwithstanding the genetic bottleneck hypothesized for Indian rhesus macaques [12]. In the present study, the higher SNP-based genetic diversity estimates in the Vietnamese longtail macaque population and the lower genetic diversity estimates in the Indonesian longtail macaque population further support introgression as the reason for inflated Indochinese genetic diversity. Moreover, when only segregating SNPs were considered, the Vietnamese longtail macaques exhibited the highest level of genetic diversity reflecting their hybrid nature and/or effects from disruptive selection.

The higher estimates of SNP diversity in the proposed hybrid zone and the clinal change in SNP frequencies along the geographic transect of the macaque range sampled here (from the Indian subcontinent, through Southeast Asia to southern, then eastern China) suggest that the effects of past and present introgression is not exclusive to sympatric macaque populations in the hybrid zone north of the Isthmus of Kra. Instead, as the cryptic population structures of the Ind2 rhesus, those of longtails from Vietnam and rhesus from western China suggest that the spatial pattern of gene flow extends beyond northern Indochina eastwards into the range of rhesus macaques in eastern China, westwards into the range of rhesus macaques in the Indian subcontinent, via Burma, and as far south into the range of longtail macaques as Peninsular Malaysia. The strong correlation between genetic distance based on pairwise *F<sub>st</sub>* estimates and geographic distance as evidenced in the lowest level of genetic differentiation between the ChiW1 and ChiW2 haplogroups to the highest level of differentiation between the Ind1 and ChiE haplogroups, the sharing of ChiS with rhesus from southern China (Gaungxi and Guangdong provinces) and the closer relationship of ChiS to ChiE and Ind2 than to ChiW or Ind1 suggest both a west-to-east dispersal of Chinese rhesus macaques from Indochina to eastern China and the derivation of ChiW from ChiE. The Malaysian longtail macaques exhibit evidence of the same introgression seen in the Vietnamese longtail macaque population, while this is lacking in all other populations of longtail macaques. According to Tosi et al. [37,38], male rhesus-mediated gene flow connected the eastern and western rhesus populations, and the longtail populations located further south of the present hybrid zone. These patterns of genetic structure and gene flow among mainland rhesus and longtail populations may not have evolved sympatrically but instead arose from extensive past and/or present admixture between the two species in that geographic region as the latitude of the hybrid zone fluctuated with climate changes throughout the Pleistocene.

While confirmation is required from wild caught samples of macaques of both species with well-documented geographic origins, this study demonstrates that traces of extensive introgression among the mainland populations of rhesus and longtails persist in the domestic supply of primate products and research facilities in the US. As all Chinese rhesus macaques housed at the California NPRC and other NPRCs share genes with conspecific and

congeneric macaque populations from other regions of China and Indochina, the potential influence of natural between-species hybridization on biomedical investigations including AIDs research in the US cannot be ignored.

Except for the pairwise  $F_{st}$  estimates among Chinese rhesus macaque populations, there was no systematic variability in the estimates of differentiation based on SNPs in the coding and non-coding genomic regions. However, differentiation between the eastern and western Chinese rhesus macaques based on 585 non-genic SNPs were 3 to 5-fold greater than estimates based on 245 genic SNPs. This is consistent with previous reports, based on mtDNA [31], of major genetic differences between rhesus macaques from eastern and western China.

The regions from which the founders are derived determine the MHC haplotype composition of the SPF colonies [17] with 23% of the Indian individuals, but only 1% of Chinese rhesus macaques, possessing at least one A01 allele. The proportion of SNPs shared between Chinese rhesus macaques from western China and Indochinese longtail macaques is much greater than that between Indochinese longtail macaques and eastern Chinese rhesus macaques and thus illustrates greater degree of introgression in the western Chinese rhesus population. If most of the inter-specific admixture is restricted to the western Chinese and Indochinese animals, then SPF stocks derived from eastern Chinese founders are freer of influences from introgression that can potentially impact experimental outcomes than those derived from western Chinese founders.

Data from this study and Kanthaswamy et al. [14] demonstrate that analyses of genetic structure based on a combination of STR- and SNP-based procedures are more reliable than those using one or the other class of markers. Because of their abundance throughout the macaque genomes including coding and non-coding regions, SNPs provide an extra level of information that can facilitate the identification of genes with signatures of adaptation and selection based on population genetic and demographic approaches that may be biomedically relevant. STRs define a greater range of rare alleles whose association with specific phenotypes can be readily detected in association studies.

## Acknowledgments

This study was supported by the California National Primate Research Center base grant (No. RR000169-48), by an ARRA supplement awarded to SK and Nick Lerche, CNPRC (No. RR018144-07) and NIH grants RR005090 and RR025871 to DGS. The collaboration of the Caribbean Primate Research Center (CPRC) and the California, Oregon and Wisconsin National Primate Research Centers faculty and staff members is also acknowledged and greatly appreciated. Animals used in this research were managed in compliance with Institutional Animal Care and Use Committee (IACUC) regulations or in accordance with the National Institutes of Health guidelines or the US Department of Agriculture regulations prescribing the humane care and use of laboratory animals.

## References

1. AKEY JM, ZHANG G, ZHANG K, JIN L, SHRIVER MD. Interrogating a High-Density SNP Map for Signatures of Natural Selection. *Genome Res.* 2002; 12:1805–1814. 10.1101/gr.631202. [PubMed: 12466284]
2. BONHOMME M, BLANCHERA A, CUARTEROS S, CHIKI L, CROUAU-ROY B. Origin and number of founders in an introduced insular primate: estimation from nuclear genetic data. *Mol Ecol.* 2008; 17:1009–1019. [PubMed: 18261045]
3. BONHOMME M, BLANCHER A, CROUAU-ROY B, CUARTERO S. Assessing introgression in the primate biomedical models *Macaca mulatta* and *Macaca fascicularis*. *J Hered.* 2009; 100(2): 158–169. doi:10.1093/jhered/esn093. [PubMed: 18974398]

4. BOWCOCK AM, RUIZ-LINARES A, TOMFOHRDE J, MINCH E, KIDD JR, CAVALLI-SFORZA LL. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*. 1994; 368:455–457. [PubMed: 7510853]
5. CHU JH, LIN YS, WU HY. Evolution and dispersal of three closely related macaque species, *Macaca mulatta*, *M. cyclopis*, and *M. fuscata*, in the eastern Asia. *Mol Phylogenet Evol*. 2007; 43:418–429. [PubMed: 17321761]
6. DELSON, E. Fossil macaques, phyletic relationships and a scenario of deployment.. In: Lindburg, DG., editor. *The Macaques: Studies in Ecology, Behavior, and Evolution*. Van Nostrand–Reinhold; New York: 1980. p. 10-30.
7. FERGUSON B, STREET SL, WRIGHT H, PEARSON C, JIA Y, THOMPSON SL, ALLIBONE P, DUBAY CJ, SPINDEL E, NORGREN RB JR. Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). *BMC Genet*. 2007; 8:43. [PubMed: 17601350]
8. FOODEN J. Systematic review of Southeast Asian longtail macaques, *Macaca fascicularis* (Raffles 1821). *Fieldiana Zool*. 1995; 81:1–206.
9. FOODEN J. Systematic review of the rhesus macaque, *Macaca mulatta* (Zimmermann, 1780). *Fieldiana Zool*. 2000; 96:1–180.
10. GOLDSTEIN DB, LINARES AR, CAVALLI-SFORZA LL, FELDMAN MW. An evaluation of genetic distances for use with microsatellite loci. *Genetics*. 1995; 139:463–471. [PubMed: 7705647]
11. HAMADA Y, URASOPON N, HADI I, MALAIWIJITNOND S. Body size, proportions and pelage color of free-ranging rhesus macaques (*Macaca mulatta*) from a zone of hybridization in northern Thailand. *Int J Primatol*. 2006; 27:497–513.
12. HERNANDEZ RD, HUBISZ MJ, WHEELER D, SMITH DG, FERGUSON B, ROGERS J, NAZARETH L, BOURQUIN T, MCPHERSON J, MUZNY D, GIBBS R, NIELSEN R, BUSTAMANTE CD. Genetic Variation Reveals Diametric Demographic Histories and Patterns of Linkage Disequilibrium for Chinese and Indian Rhesus Macaques. *Science*. 2007; 316:240–243. [PubMed: 17431170]
13. KANTHASWAMY S, KURUSHIMA J, SMITH DG. Population subdivision and gene flow among wild orangutans. *Primates*. 2002; 43(4):315–327. [PubMed: 12426465]
14. KANTHASWAMY SATKOSKIJ, GEORGE D, KOU A, ERICKSON BJ-A, SMITH DG. Interspecies hybridization and the stratification of nuclear genetic variation of rhesus (*Macaca mulatta*) and long-tailed macaques (*Macaca fascicularis*). *Int J Primatol*. 2008; 29:1295–1311. DOI 10.1007/s10764-008-9295-0. [PubMed: 19122840]
15. KANTHASWAMY GILL, SATKOSKI J, GOYAL V, MALLADI V, KOU A, BASUTA K, SARKISYAN L, GEORGE D, SMITH DG. The Development of a Chinese-Indian Hybrid (Chindian) Rhesus Macaque Colony at the California National Primate Research Center (CNPRC) by Introgression. *J Med Primatol*. 2009a; 38(2):86–96. [PubMed: 18715266]
16. KANTHASWAMY S, FERGUSON B, DUBAY C, ET AL. Resources for genetic management and genomics research on non-human primates at the National Primate Research Centers (NPRCs). *J Med Primatol*. 2009b; 38(s1):17–23. [PubMed: 19863674]
17. KANTHASWAMY K, KOU A, SATKOSKI J, NG J, GILL L, LERCHE NW, SMITH DG. Genetic characterization of specific pathogen-free (SPF) rhesus macaque (*Macaca mulatta*) populations at the California National Primate Research Center (CNPRC). *Am J Primatol*. In Press.
18. KAPLAN NL, HUDSON RR, LANGLEY CH. The “hitchhiking effect” revisited. *Genetics*. 1989; 123:887–889. [PubMed: 2612899]
19. LI D, XU H, CHENG A, SMITH DG, YAO Y, DU D, NI Q. Subspecific taxonomy and phylogenetic analysis of Chinese rhesus macaques (*Macaca mulatta*) based on mitochondrial control region sequences. *Am J of Primatol*. In Press.
20. LISTMAN JB, MALISON RT, SUGHONDHABIROM A, YANG BZ, RAAUM RL, THAVICHACHART N, SANICHWANKUL K, KRANZLER HR, TANGWONCHAI S, MUTIRANGURA A, DISOTELL TR, GELERNTER J. Demographic changes and marker

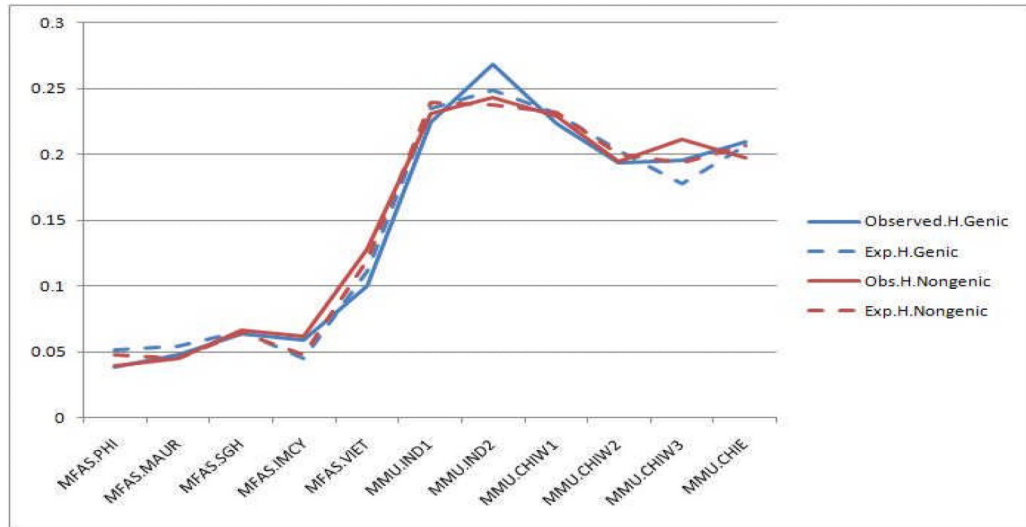
- properties affect detection of human population differentiation. *BMC Genet.* 2007; 8:21. doi: 10.1186/1471-2156-8-21. [PubMed: 17498298]
21. MALHI RS, SICKLER B, LIN D, SATKOSKI J, GEORGE D, KANTHASWAMY S, SMITH DG. MamuSNP: A SNP resource for Rhesus macaques (*Macaca mulatta*). *PLoS ONE.* 2007; 2(5):e438. doi:10.1371/journal.pone.0000438. [PubMed: 17487284]
  22. MORIN PA, KANTHASWAMY S, SMITH DG. Simple sequence repeat, SSR, polymorphisms for colony management and population genetics in rhesus macaques (*Macaca mulatta*). *Am J Primatol.* 1997; 42:199–213. [PubMed: 9209585]
  23. MELNICK DJ, KIDD KK. Genetic and evolutionary relationships among Asian macaques. *Int J Primatol.* 1985; 6:123–160.
  24. NEI, M. *Molecular Evolutionary Genetics.* Columbia University Press; New York: 1987.
  25. NEIGEL, JE.; AVISE, JC. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation.. In: Nevo, E.; Karlin, S., editors. *Evolutionary processes and theory.* Academic Press; NY: 1986. p. 513-534.
  26. PRITCHARD JK, STEPHENS M, DONNELLY P. Inference of population structure using multilocus genotype data. *Genetics.* 2000; 155:945–959. [PubMed: 10835412]
  27. PURCELL S, NEALE B, TODD-BROWN K, THOMAS L, FERREIRA MAR, BENDER D, MALLER J, SKLAR P, DE BAKKER PIW, DALY MJ, SHAM PC. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet.* 2007:81.
  28. RAYMOND M, ROUSSET F. GENEPOP: Population genetics software for exact tests and ecumenicism. *J Hered.* 1995; 86(3):248–249.
  29. ROBERTS JA, SMITH DG, HENDRICKX A. Managing the rhesus supply. *Science.* 2000; 287:1591. [PubMed: 10733423]
  30. SATKOSKI, J.; GEORGE, D.; SMITH, DG.; KANTHASWAMY, S. Genetic Characterization of Wild and Captive Rhesus Macaques in China.. *J Med Primatol.* 2007. [www.blackwell-synergy.com/doi/abs/10.1111/j.1600-0684.2007.00228.x](http://www.blackwell-synergy.com/doi/abs/10.1111/j.1600-0684.2007.00228.x)
  31. SATKOSKI JA, MALHI RS, KANTHASWAMY S, TITO RY, MALLADI VS, SMITH DG. Pyrosequencing as a method for SNP identification in the rhesus macaque (*Macaca mulatta*). *BMC Genomics.* 2008; 9:256. [PubMed: 18510772]
  32. SMITH DG, MCDONOUGH J. Mitochondrial DNA variation in Chinese and Indian rhesus macaques (*Macaca mulatta*). *Am J Primatol.* 2005; 65(1):1–25. [PubMed: 15645455]
  33. SMITH DG, MCDONOUGH JW, GEORGE DA. Mitochondrial DNA variation within and among regional populations of longtail macaques (*Macaca fascicularis*) in relation to other species of the *fascicularis* group of macaques. *Am J Primatol.* 2007; 69(2):182–189. [PubMed: 17177314]
  34. STEVISON LS, KOHN HM. Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. *Mol Ecol.* 2009; 18(11):2457–2475. [PubMed: 19457196]
  35. SUSSMAN RW, TATTERSALL I. Behavior and ecology of *Macaca fascicularis* in Mauritius: a preliminary study. *Primates.* 1981; 22:192–205.
  36. SUSSMAN RW, TATTERSALL I. Distribution, abundance, and putative ecological strategy of *Macaca fascicularis* on the island of Mauritius, southwestern Indian Ocean. *Folia Primatol.* 1986; 46:28–43.
  37. TOSI AJ, MORALES JC, MELNICK DJ. Y-chromosome and mitochondrial markers in *Macaca fascicularis* indicate introgression with Indochinese *M. mulatta* and a biogeographic barrier in the Isthmus of Kra. *Int J Primatol.* 2002; 23(1):161–178.
  38. TOSI AJ, MORALES JC, MELNICK DJ. Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution.* 2003; 57(6):1419–1435. [PubMed: 12894949]
  39. TOSI AJ, COKE CS. Comparative phylogenetics offer new insights into the biogeographic history of *Macaca fascicularis* and the origin of the Mauritian macaques. *Mol Phylogenet Evol.* 2007; 42:498–504. [PubMed: 16979912]
  40. WEIR BS, COCKERHAM CC. Estimating F-statistics for the analysis of population structure. *Evolution.* 1984; 38:1358–1370.



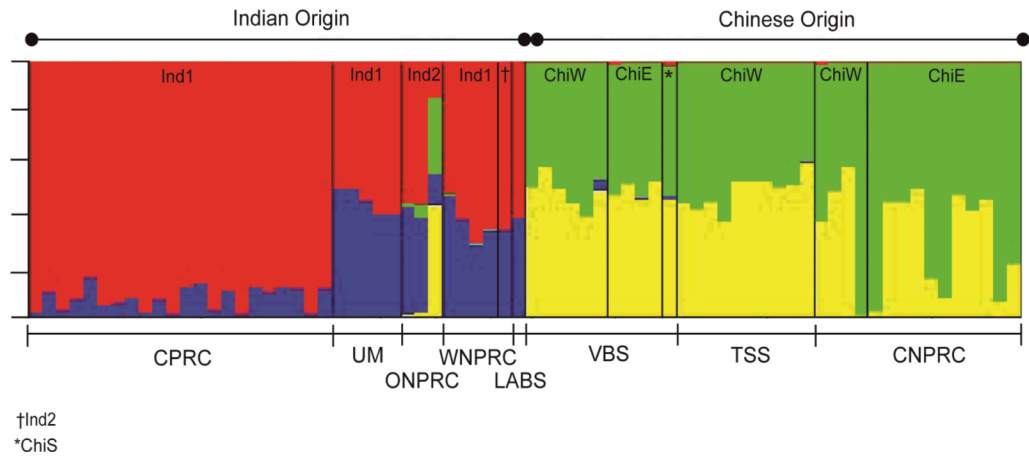
41. YAO YT, HARFF J, MEYER M, ZHAN WH. Reconstruction of paleocoastlines for the northwestern South China Sea since the last glacial maximum. *Sci. China Ser. D Earth Sci.* 2009; 52(8):1127–1136.



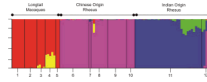
**Figure 1.** Geographic range of rhesus (light grey) and longtail (dark grey) macaques (Adapted from Kanthaswamy et al. [15]).



**Figure 2.**  
Estimates of OH and EH of the 830 SNPs.



**Figure 3.** Bayesian clustering based on 1307 SNPs across the rhesus genome.



**Figure 4.**  
Bayesian clustering based on 830 SNPs across the rhesus and longtail genomes.



**Table 1**

Geographic localities and breeding and research facilities represented by the rhesus and longtail macaque samples.

Population	Sample size (N)	Species	Origin	Breeder/Importer or Center
1	6	<i>M. fascicularis</i>	The Philippines	New Iberia
2	6	<i>M. fascicularis</i>	Mauritius	COVANCE
3	4	<i>M. fascicularis</i>	Malaysia	Wild/Singapore General Hospital
4	5	<i>M. fascicularis</i>	Vietnam	Primate Products
5	3	<i>M. fascicularis</i>	Indonesia	CNPRC
6	14	<i>M. mulatta</i>	Eastern China	Guangdong/VBS; Suzhou/CNPRC
7	1	<i>M. mulatta</i>	Southern China	Guangdong/VBS
8	8	<i>M. mulatta</i>	Western China (ChiW1)	Guangdong/VBS Sichuan/TSS, GPC and VBS
9	10	<i>M. mulatta</i>	Western China (ChiW2)	Sichuan/TSS and VBS Suzhou/CNPRC
10	4	<i>M. mulatta</i>	Western China (ChiW3)	Sichuan/TSS and VBS
11	32	<i>M. mulatta</i>	Kashmir, New Delhi and Uttar Pradesh, India (Ind1)	CPRC, UM, WNPRC and LABS
12	4	<i>M. mulatta</i>	New Delhi, India (Ind2)	ONPRC and WNPRC

Guide to names and abbreviations. New Iberia: New Iberia Research Center, New Iberia LA. COVANCE: Covance Research Products, Alice Texas. Primate Products: Primate Products, Inc. Immokalee FL. CNPRC: California National Primate Research Center, Davis CA. VBS: Valley Biosystems, Sacramento CA. TSS: Three Springs Scientific Inc., Perkaskie PA. CPRC: Caribbean Primate Research Center, Cayo Santiago PR. UM: University of Miami, Miami FL. WNPRC: Wisconsin National Primate Research Center, Madison WI. LABS: Alpha Genesis (formerly Laboratory Animal Breeders and Services of Virginia), Yemassee SC. ONPRC: Oregon National Primate Center, Beaverton OR.

**Table 2**

Estimates of mean observed (OH) and expected (EH) heterozygosity values across 1307 SNPs in the rhesus genome. The ChiW haplogroup consists of three subhaplogroups: ChiW1, ChiW2 and ChiW3 [32].

	OH	EH
Ind1 (CPRC, LABS, UM, WNPRC)	0.237	0.247
Ind2 (ONPRC and WNPRC)	0.256	0.244
ChiE (CNPRC)	0.206	0.212
ChiW1 (CNPRC, TSS)	0.238	0.238
ChiW2 (CNPRC, TSS)	0.201	0.208
ChiW3 (CNPRC, TSS)	0.212	0.196

**Table 3**

Estimates of mean observed (OH) and expected (EH) heterozygosities at the 830 SNPs in the rhesus and longtail genomes including the 323 segregating SNPs in the longtail genome, and genic and non-genic SNPs shared by both spp.

	OH (830 SNPs)	OH (323 SNPs)	EH (830 SNPs)	EH (323 SNPs)	OH (245 Genic SNPs)	EH (245 Genic SNPs)	OH (585 Non-genic SNPs)	EH (585 Non-genic SNPs)
Philippine longtail (New Iberia)	0.039	0.098	0.0486	0.123	0.038	0.052	0.039	0.047
Mauritius longtail (COVANCE)	0.046	0.117	0.0476	0.122	0.048	0.054	0.045	0.045
Peninsular Malaysia longtail (Singapore General Hospital)	0.066	0.164	0.0648	0.163	0.064	0.065	0.066	0.065
Indonesian longtail	0.061	0.156	0.0469	0.120	0.059	0.045	0.062	0.048
Vietnamese longtail (Primate Products)	0.120	0.310	0.117	0.300	0.100	0.112	0.128	0.119
Ind1 rhesus	0.229	0.253	0.238	0.261	0.225	0.234	0.231	0.239
Ind2 rhesus	0.251	0.279	0.241	0.275	0.268	0.249	0.243	0.238
ChiW1 rhesus	0.228	0.261	0.232	0.254	0.224	0.231	0.229	0.232
ChiW2 rhesus	0.194	0.229	0.201	0.238	0.194	0.203	0.194	0.200
ChiW3 rhesus	0.207	0.251	0.189	0.229	0.196	0.178	0.212	0.194
ChiE rhesus	0.201	0.233	0.206	0.238	0.209	0.206	0.198	0.206

**Table 4**

Pairwise  $F_{st}$  values among the different geographic populations of rhesus as labeled according to their mtDNA haplogroups [32].

	China East	China South	ChiW1	ChiW2	Ind 1
ChiW1 rhesus	0.010	-0.010			
ChiW2 rhesus	0.003	0.030	-0.001		
Ind1 rhesus	0.236	0.200	0.230	0.214	
Ind2 rhesus	0.148	0.069	0.136	0.116	0.029

Table 5

Pairwise differentiation among the various rhesus and longtail populations based on 830 shared SNPs. The ChiW haplogroup consists of three haplotypes: ChiW1, ChiW2 and ChiW3 [32].

	CPRC	LABS	UM	ONPRC	WNPRC	Guangdong	Sichuan	Suzchou	Indon.	Maur.	Philip.	P. Malay.
CPRC rhesus (Ind1)												
LABS rhesus (Ind1)	0.05											
UM rhesus (Ind1)	0.06	-0.01										
ONPRC (Ind2)	0.08	-0.07	0.01									
WNPRC (Ind1, Ind2)	0.05	-0.01	0.02	0.02								
Guangdong, China rhesus (ChiW, CNPRC, TSS)	0.25	0.24	0.21	0.13	0.22							
Sichuan, China rhesus (ChiW, CNPRC, TSS)	0.25	0.2	0.21	0.12	0.21	0.01						
Suzchou, China rhesus (ChiE, CNPRC, TSS)	0.25	0.22	0.21	0.12	0.21	0.01	0.02					
Indonesian longtail (CNPRC)	0.54	0.73	0.57	0.57	0.57	0.63	0.58	0.6				
Mauritius longtail (COVANCE)	0.57	0.79	0.65	0.67	0.65	0.7	0.62	0.65	0.07			
Philippine longtail (New Iberia)	0.58	0.79	0.65	0.67	0.65	0.7	0.63	0.65	0.13	0.25		
Peninsular Malaysia longtail (National University of Singapore)	0.54	0.73	0.59	0.6	0.59	0.65	0.59	0.61	0.01	0.13	0.24	
Vietnamese longtail (Primate Products)	0.51	0.6	0.54	0.52	0.53	0.58	0.55	0.57	0.15	0.27	0.31	0.15



**Table 6**

Pairwise differentiation among the various rhesus and longtail populations based on 245 genic (above diagonal and 585 non-genic (below diagonal) SNPs.

	Philippine	Mauritius	P. Malaysia	Indonesia	Vietnam	Ind1	Ind2	ChiW1	ChiW2	ChiW3	ChiE
Philippine longtail		0.20	0.22	0.11	0.36	0.60	0.68	0.65	0.68	0.74	0.67
Mauritius longtail	0.16		0.10	0.02	0.31	0.59	0.67	0.63	0.66	0.73	0.65
Peninsular Malaysia longtail	0.25	0.24		0.11	0.26	0.58	0.61	0.59	0.63	0.68	0.63
Indonesian longtail	0.11	0.10	0.04		0.23	0.57	0.58	0.58	0.62	0.67	0.63
Vietnamese longtail	0.29	0.31	0.19	0.19		0.54	0.56	0.54	0.58	0.61	0.58
Ind1 rhesus	0.57	0.57	0.55	0.54	0.52		0.03	0.21	0.22	0.19	0.22
Ind2 rhesus	0.67	0.67	0.60	0.57	0.55	0.02		0.12	0.13	0.12	0.15
ChiW1 rhesus	0.64	0.65	0.59	0.57	0.55	0.23	0.12		0.02	0.03	0.02
ChiW2 rhesus	0.68	0.68	0.64	0.62	0.60	0.23	0.13	0.03		-0.01	0.00
ChiW3 rhesus	0.73	0.74	0.67	0.65	0.62	0.22	0.10	0.00	0.01		0.02
ChiE rhesus	0.66	0.66	0.64	0.62	0.59	0.27	0.18	0.11	0.09	0.07	

Table 7

The average proportion of shared alleles between all population pairs across the 245 genic SNPs. If both populations are polymorphic, the proportion of shared alleles would be 100%, while if one population was polymorphic and the other monomorphic, the proportion of shared alleles would drop to 50%. Populations monomorphic for different alleles would, of course, share 0% of alleles.

	Philippine	Mauritius	P. Malaysia	Indonesia	Vietnam	Ind1	Ind2	ChiW1	ChiW2	ChiW3
Philippine longtail										
Mauritius longtail	0.94									
Peninsular Malaysia longtail	0.92	0.93								
Indonesian longtail	0.94	0.95	0.94							
Vietnamese longtail	0.87	0.88	0.88	0.87						
Ind1 rhesus	0.63	0.62	0.63	0.61	0.68					
Ind2 rhesus	0.62	0.62	0.62	0.61	0.67	0.92				
ChiW1 rhesus	0.62	0.64	0.63	0.61	0.69	0.79	0.82			
ChiW2 rhesus	0.63	0.64	0.64	0.62	0.70	0.85	0.85	0.88		
ChiW3 rhesus	0.67	0.67	0.68	0.66	0.70	0.82	0.83	0.84	0.91	
ChiE rhesus	0.63	0.63	0.63	0.61	0.69	0.86	0.86	0.89	0.95	0.90