Phylogenomics reveals pervasive ancient introgression in the radiation of macaques

Xinxin Tan[1,2#], Jiwei Qi[1#], Zhijin Liu[3#], Pengfei Fan[4], Liye Zhang[5], Jing Li[6], Christian Roos[5,7*], Xuming Zhou[1*], Ming Li[1,8*]

[1]CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

[2] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

[3] College of Life Sciences, Capital Normal University, Beijing 100049, China

[4] School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China

[5] Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077, Göttingen, Germany

[6]Key Laboratory of Bio-resources and Eco-environment of Ministry of Education, College of 9 Life Sciences, Sichuan University, Chengdu 610064, China

[7]Gene Bank of Primates, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077, Göttingen, Germany

[8]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China

# Contributed equally to this work.

*Correspondence and requests for materials should be addressed to M.L. (lim@ioz.ac.cn), X.Z. (E-mail: zhouxuming@ioz.ac.cn) and C.R. (croos@dpz.eu).

Running heads: pervasive introgression in *Macaca*

Abstract

Introgression is common in speciation and demographic history of animals. The genus *Macaca*, with 23 species assigned into four to seven species groups, exhibits the largest geographic range and represents the most successful adaptive radiation of nonhuman primates. Here, we conducted phylogenomic analyses of macaques using 16 newly generated and eight published macaque genomes. Strong support for the division of the genus into seven species groups was found. Though both ancient introgression and incomplete lineage sorting have contributed to the radiation and evolution of macaques, extensive ancient and recent introgression events affect almost all extant Asian macaque species. Particularly, the contradicting phylogenetic position of the *silenus/nigra* lineage in the macaque phylogeny has likely resulted from introgression between the ancestors of the *fascicularis/mulatta* and *silenus/nigra* lineages. Furthermore, using the dated phylogenetic tree combined with documented records, an integrated scenario for the radiation of macaques is reconstructed. This study provides novel insights into ancient introgression involved in the radiation of macaques that may help to understand the rapid speciation of nonhuman primates.

Keywords: *Macaca*, species group, phylogenomics, introgression, incomplete lineage sorting, radiation.

Introduction

Introgression and hybridization are important mechanisms in speciation and evolution, and over the past decade increasing attention has been paid to their role in evolutionary diversification [1, 2, 3, 4, 5, 6, 7]. Hybridization has been recognized as one of the most interesting topics in evolutionary biology, and the genetic introgression, caused by hybridization, is not always maladaptive and could be a potent evolutionary force [3, 8, 9, 10]. With the development of high-throughput sequencing technology and new computational tools, phylogenomics emerged and allows researchers to explicitly account for hybridization when constructing species trees [11, 12]. However, incomplete lineage sorting (ILS), an alternative explanation for contradicting gene trees, needs to be considered as well when reconstructing a species tree [13]. ILS occurs mainly during rapid speciation events, when new lineages descend within a short time period from ancestors, particularly when these have large effective population sizes (*Ne*) [14]. As a result, ILS will cause ancestral polymorphisms to persist even when descendant lineages diverged [15]. Therefore, both ILS and gene flow can lead to topological discordance between gene trees and species trees [16, 17]. For example, comparative genomics found the speciation of rorqual evolution occurred under strong gene flow, which lead to the morphological and genetic analyses produced different scenarios [12]. Edelman et al. [18] used 20 *de novo* genome assemblies to explore the speciation history and gene flow in rapidly radiating *Heliconius* butterflies and distinguish ILS from introgression. They discovered that introgressed loci are underrepresented in low-recombination and gene-rich regions [18]. For primates, information about a genome-wide view on hybridization is still limited but genome sequencing efforts are on the way [19]. For example, Rogers, et al. [20] confirmed reticulation in baboons and documented multiple episodes of admixture and introgression during the radiation of *Papio* using a comparative genomic approach.

The genus *Macaca* (Primates: Cercopithecidae) exhibits the largest geographic range and represents the most successful adaptive radiation of nonhuman primates [21] (**Fig. 1A**). The 23 currently recognized macaque species are subdivided into four to seven species groups according to similarities in morphology, ecology, behavior, distribution and genetics [21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33] (**supplementary Fig. S1**). However, the number and composition of the

species groups is disputed. Before the whole-genome age, phylogenetic analyses were mainly based on a few mitochondrial and/or nuclear markers resulting in contradicting phylogenetic relationships [26, 34, 35, 36], which are partially caused by introgression and hybridization events among various macaque species.

Introgression and hybridization between macaque species are common, for instance, among Sulawesi macaques, between *M. mulatta* and *M. fascicularis*, and between *M. mulatta* and *M. thibetana* [37, 38, 39, 40, 41, 42, 43, 44, 45]. Moreover, for *M. arctoides* a hybrid origin was suggested as different molecular markers place the species into different clades [33, 34, 35, 36, 46]. However, most analyses have focused on gene flow within macaque species groups, or between species of the *fascicularis*/*mulatta* and *arctoides*/*sinica* species groups. Few studies included samples from all species groups when identifying genus-wide hybridization, and few attempted to detect ancient hybridization events in macaques [36, 47]. A recent study reported the history of the genus *Macaca* using 14 individuals from nine species and detected extensive gene flow signals, with the strongest signals between the *fascicularis/mulatta* and *silenus* groups [48]. Introgression signals among island species, such as Sulawesi macaques and other species, were also observed. This put forward a variety of possible reasons for this phenomenon including the genomic similarity of closely related species or ancestral introgression. Therefore, whether hybridization occurred among all macaques, especially among species that are geographically distant still needs to be investigated. Furthermore, it remains unclear how much ancient hybridization or ILS have contributed to the diversity of the macaque genome.

In this study, we aimed to trace the complex, likely reticulated evolutionary history of macaques and newly sequenced whole genomes of 16 individuals representing 13 Asian (*M. mulatta, M. cyclopis, M. fuscata, M. arctoides, M. assamensis, M. leucogenys, M. sinica, M. leonina, M. silenus, M. nemestrina, M. maura, M. tonkeana, M. nigra*) and the only African macaque species (*M. sylvanus*). The sequences were mapped to the latest *Papio anubis* reference genome (Panu_3.0) to obtain single nucleotide variants (SNVs). Combined with eight published macaque genomes, we performed a comprehensive genomics analysis among 24 macaque individuals, representing 16 species and all species groups, aiming to reconstruct the phylogeny of the genus *Macaca,* investigate ILS and introgression among macaque species, and trace the

evolutionary history and radiation of the genus *Macaca*.

Results

Genome Sequencing, Variant Discovery and Diversity

Whole genome sequencing was performed with the Illumina HiSeq 2500 and 4000 platforms. In total, more than 1,364 Gbp clean data (quality-controlled reads) were obtained and the statistics results of the clean data for each species are shown in **supplementary table S1**. Published genome data of eight macaques (*M. mulatta* (Chinese rhesus macaque, CR), *M. mulatta* (Indian rhesus macaque, IR), *M. fascicularis*, *M. arctoides*, *M. assamensis, M. thibetana, M. nemestrina* and *M. tonkeana*) were downloaded from NCBI (**supplementary table S1**) and two of them overlapped with [48]. Clean data for each species were mapped to the *Papio anubis* reference genome (Panu_3.0) and yielded genome-wide coverage from 16.85 to 42.34× with an average effective depth at 28.24× (**supplementary table S1**). We picked a more distant reference genome (*Papio anubis*) rather than the more recently evolved macaque (*Macaca mulatta*) to avoid biases that would arise from mapping to an in-group reference genome. Even though this decision likely led to some degree of lost data (in regions that lacked ortholog to *Papio anubis*), however, the mapping coverage (4×) for macaques range from 93.55% to 97.59%, which is very similar to mapping coverage (4×) for *Papio* (98.75%), suggesting a potential bias to be limited. The mapping rate for *M. nemestrina* (2) is lower than 90%, which could be caused by low DNA quality.

The number of high-quality SNVs for each individual ranged from 15,943,244 to 21,778,812, counting both homozygous and heterozygous alternatives (**supplementary table S2**). Genome-wide heterozygosity varied considerably among macaques (**supplementary Fig. S2** and **supplementary table S2**) which may partially reflect different demographic histories. Based on gene annotations from the reference genome, the exonic SNVs in macaques ranged from 225,336 to 257,712 (**supplementary tables S3** and **S4**). The species-specific SNVs for each macaque ranged from 1,279,649 to 8,688,344 (**supplementary table S2**).

Phylogeny and divergence among macaques

In order to reconstruct the phylogeny of macaques, we generated consensus sequences of all genomes for each macaque individual. We split the aligned chromosomes into 6,351 non-overlapping genome fragments (GFs), each 200-kilo basepair (kbp) long, which representing 46% of the genome sequence. The Densitree [49] representation of the 6,351 GF trees, reconstructed using maximum likelihood (ML) methods, suggests a division of macaques into seven clusters, which is consistent with seven species groups, but the relationships among them remain poorly resolved [21, 29] (**Supplementary Fig. S3**). The multispecies coalescent (MSC) species tree based on the 6,351 GFs was supported with posterior probabilities (PPs) of 1.0 for all branches (**Fig. 1B** and **supplementary Fig. S4**). This species tree, consistent with neighbor joining (NJ) tree and maximum-likelihood (ML) tree based on autosomal SNVs (**supplementary Fig. S5** and **supplementary Fig. S6**), supported the segregation of macaques into seven well-supported clades or lineages that correspond to the classification of the genus into seven species groups [21, 29]. Among macaques, the African *sylvanus* lineage diverged first, followed by an initial separation of extant Asian macaques into a clade with the *silenus* and *nigra* lineages and a clade including all other Asian lineages. The latter segregated into a clade including the *arctoides* and *sinica* lineages and another clade with the *fascicularis* and *mulatta* lineages (**Fig. 1B** and **supplementary Figs. S4-S6**). The ML tree based on mitochondrial genomes (mitogenomes), consistent with previous results [32, 33], showed that among Asian macaques, the *silenus/nigra* clade separated first, followed by the *sinica* lineage, while among the remaining lineages, *fascicularis* split first and *mulatta* and *arctoides* afterwards (**supplementary Fig. S7** and **supplementary table S5**). However, the quartet score 'q', which is the support value for possible phylogenetic arrangements, identified a conflict in resolving the branch leading to the ancestor of the *mulatta/fascicularis* and *arctoides/sinica* groups (**Fig. 1C-D**, branch no. 2 in **Fig. 1B** and **supplementary Fig. S4**). These two alternative topologies were also reported by Song et al. [48], however, the reason for this conflict is unclear. Some additional branches with a low q1 ($q1 < 0.8$, branch no. 1-6 in **Fig. 1B** and **supplementary Fig. S4**) show similar pattern.

A consensus network analysis [40] of the GF trees recovered a network with bifurcating branches that emerged from a central cycle in the center of the network that indicates conflicting signals for the position of the *silenus/nigra* clade among Asian macaques (**Fig. 1E** and

supplementary Fig. S8). At a threshold for conflicting edges of 30%, the position of the *silenus/nigra*, *arctoides/sinica* and *fascicularis/mulatta* clades were unascertainable (**supplementary Fig. S8**). When the threshold for conflicting edges was reduced, the phylogenetic signal became even more complex, indicating additional phylogenetic conflict and extensive gene flow in the data set (**supplementary Fig. S8**). F_{ST} analyses suggested significant differences within (medians for F_{ST} are 0.199) and between (medians for F_{ST} are 0.467) different macaque groups ($P < 0.0001$) (**supplementary table S6** and **Fig. 1F**). A clear division of macaques into different clusters was also obtained by a principal component analysis (PCA) (**supplementary Fig. S9** and **supplementary table S7**).

The genus *Macaca* was estimated to have originated ca. 7.76 million years ago (Mya) during the Late Miocene using MCMCTREE in PAML v4.9 [50] (**Fig. 2** and **supplementary Fig. S10**). We revealed a divergence time for the split of *M. sylvanus* from the Asian macaques at 5.29 Mya (95% highest posterior density, HPD, 3.65–6.70), close to the Messinian Salinity Crisis [51]. In Asia, divergence of species groups occurred in the Pliocene, starting with the split between the *silenus/nigra* clade and the other four species groups 4.01 Mya (95% HPD 5.59–2.61), followed shortly afterwards by the separation of the ancestors of the *mulatta/fascicularis* and *sinica/arctoides* clades (3.79 Mya, 95% HPD 5.28–2.50). These three clades finally diverged into species groups 3.20 Mya (95% HPD 4.79–2.07; *silenus* and *nigra* groups), 3.01 Mya (95% HPD 4.52–1.84; *sinica* and *arctoides* groups) and 2.94 Mya (95% HPD 4.25–1.52; *mulatta* and *fascicularis* groups). Speciation within species groups took place in the Pleistocene, 2.56–1.39 Mya.

Introgression in macaques

Previous studies suggested a vital role of hybridization in shaping the evolutionary history of macaques [35, 36, 37, 38, 41, 42, 43, 44, 45, 46, 52]. We further explored gene flow among macaque species using the *D* statistic [53] and inspected various cases of gene flow events among different macaque groups and species (**Fig. 3A-B, supplementary Fig. S11-S13** and **supplementary tables S8-S10**).

We found statistically significant evidence for genetic exchange between the

fascicularis/mulatta and *silenus/nigra* clades (labeled as I in **Fig. 3A** and **Fig. 3B**), which likely led to the inconsistent phylogenetic position of the *silenus/nigra* clade (**Fig. 1C**). *D* statistic results concordant with the topologies (((sin,mul),nig),Out) and (((sin,fas),nig),Out), and (((arc,mul),nig),Out) and (((arc,fas),nig),Out) suggested that the *nigra* group shares more derived alleles with *fascicularis/mulatta* than with *sinica/arctoides* (absolute value of *D*, D_{abs} = 0.1867~0.3190; absolute value of Z score, $Z_{abs} \geq 3$; **Figs. 3B** and **supplementary table S8**). With similar frequency, the *silenus* group shared significantly more derived alleles with *fascicularis/mulatta* than with *sinica/arctoides* (D_{abs} = 0.2203~0.3325, $Z_{abs} \geq 3$, **Figs. 3B** and **supplementary table S8**). Given that a similar number of derived alleles are shared between the *fascicularis/mulatta* clade and both the *nigra* and *silenus* groups, and considering that the species of the *nigra* group are isolated on the Indonesian island of Sulawesi, gene flow between *fascicularis/mulatta* and *silenus/nigra* most likely occurred before the divergence of those species groups (labeled as I in **Fig. 3A**), which is supported by the results of qpgraph (**supplementary Fig. S12**). Introgression can also be studied by topology-based ML analysis using PhyloNet [54], which identifies a statistically significant signal for gene flow between the ancestor of *fascicularis* and *mulatta* groups and the ancestor of the s*ilenus and nigra* groups (**Fig. 3C and supplementary Fig. S13**). The reticulations show inheritance probabilities of ~33%, resembling the quartet-score distribution of the coalescent tree analyses (**Fig. 1C** and **Supplementary Fig. S4**).

Likewise, based on *D* statistic of (((sin,arc),fas),Out) and (((sin,arc),mul),Out), the *arctoides* group shared significantly more derived alleles with *fascicularis/mulatta* than the *sinica* group shared with *fascicularis/mulatta*, indicating more gene flow between *arctoides* and *fascicularis/mulatta* (labeled as II in **Fig. 3A,** and **supplementary table S9**), which is consistent with earlier studies [34, 35, 36, 46]. However, PhyloNet did not detect evidence for gene flow between *arctoides* and *sinica* ancestors after *arctoides* and *sinica* diverged (**Fig. 3C** and **supplementary Fig. S13**).

In addition, we also found evidence for gene flow between *silenus* and *fascicularis/mulatta* and *arctoides/sinica* groups, respectively (labeled as III and IV in **Fig. 3A** and **supplementary tables S9** and **S10**). Given that both recent and ancient gene flow events may have occurred, their signatures could obfuscate each other. We integrated the observed gene flow signals identified

from *D* statistic into a single historical model using 'qpgraph' (**supplementary Fig. S12**). We found the likelihood score and worst residual (WR) for a topology without gene flow (admix=0) to be 15940.22 and 73.86, respectively (**supplementary Fig. S12A**). A model with four admixture events (gene flow I to IV, **supplementary Fig. S12D**) had a similar likelihood score (8819.78) and WR (73.86) as a model with two admixture events (gene flow I and II; likelihood score = 9005.03, WR= 73.86, **supplementary Fig. S12C**). Therefore, it remains difficult to determine if introgression events III and IV really occurred. Subsequently, to check whether the gene flow event between *fascicularis/mulatta* and *silenus/nigra* is ancient or recent, we added a recent introgression event through automated model search and repeated simulations between the *fascicularis/mulatta* and *silenus/nigra* groups. The results showed that this model (likelihood score = 15420.07, WR=73.86 **supplementary Fig. S12E**) is similar to that without any introgression event (**supplementary Fig. S12A**), indicating that gene flow between *fascicularis/mulatta* and *silenus/nigra* (labeled as I in **Fig. 3A**) is more likely to be ancient. In addition, introgression between *M. arctoides* and the ancestors of *M. assamensis* and *M. thibetana* was identified based on PhyloNet (**supplementary Fig. S13**), and with a higher *D* value (**supplementary table S10**). We added this hybridization event to a model with gene flow events I and II and obtained a global minimum likelihood score of 5653.37 and WR of 54.73 (**Fig. 3D**), indicating that this model is closest to the observed value.

Regions with local topologies discordant from the species tree may have arisen through introgression or ILS. QuIBL (quantifying introgression via branch lengths), which is a tree-based method [18] was then employed to distinguish ILS and introgression based on the distribution of internal branch lengths for a given three-taxon subtree (triplet) (18, 17). As for the internal branch of *fascicularis/mulatta* and *silenus/nigra* lineages, the Bayesian information criterion (BIC) test indicated that the phylogenetic discordances were caused by both ILS and introgression (**supplementary table S11**). The results showed the strongly reduced mean branch length when *M. arctoides* is the outgroup, which is inconsistent with species tree **(Fig. 3E-F**). In addition, relationship between internal branch length and non-ILS proportion of each topology showed minor topology also has a portion of the window tree nearly 100% non-ILS proportion, which suggest that speciation of several macaques influenced by a combination of ILS and introgression

(**Fig.3G**).

Detecting putatively introgressed windows

F_d statistic [55] was initially used to calculate the potential introgression in 10-kb windows firstly and genome regions with top 0.05 high F_d values were selected. Given it is of challenge to accurately identify the introgressed regions in macaque genomes, we applied a two-layer hidden Markov model [56] to infer local ancestry of admixed individuals of the *fascicularis/mulatta* lineage. We selected the top 1% windows with high sum of the proportion of source population (*silenus/nigra* lineage) ancestry for all SNPs in a 10 kb windows with a high F_d values as the final potential introgression regions (**Supplementary tables S12**). When examining local ancestral allele dosages, we found some functional genes (e.g., *CESR3*, *ACER3*, *GSK3B*) present in these potential introgressed regions (**Fig. 3H**, **supplementary Fig. S14** and **supplementary table S13**). The phylogenetic tree based on introgressed and non-introgressed regions resulted in different phylogenetic positions of the *fascicularis/mulatta* lineage. Nevertheless, although the introgressed regions were identified by strict approach, we should admit it is difficult to preclude ILS might have contributed to the complex radiation of macaques as gene flow occurred too early to be quantified. In addition, because we cannot locate the breakpoints of potential introgressed regions based on existing results and data, an in-depth scan is required in the future based on the potential introgression size reported here.

We then reconstructed a phylogenetic tree based on autosomal SNVs including only SNVs located in all putatively introgressed regions for *fascicularis/mulatta* lineage and there is no doubt that the *arctoides/sinica* lineages diverged first among Asian macaques (**supplementary Fig. S15**). In addition, we reconstructed a phylogenetic tree excluding SNVs located in all putatively introgressed regions for *fascicularis/mulatta* lineage (**supplementary Fig. S15**). The result upheld the topology displayed by the species tree (**Fig. 1B**). Numerous research reported that introgression among macaque species were/are ongoing even after speciation was completed [33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46]. For example, the *D* statistic of (((*M. assamensis*, *M. leucogenys*), *M. sinica*), Out) or (((*M. thibetana*, *M. leucogenys*), *M. sinica*), Out) suggested *M. leucogenys* shares more derived alleles with *M. sinica* (**supplementary table S10)**. Moreover, gene flow may have occurred between *M. nemestrina* and other species. The *D* statistic result

based on (((*M. leonina*, *M. nemestrina*), *nigra group*), Out) or (((*M. silenus*, *M. nemestrina*), *nigra group*), Out) suggested gene flow occurred between *M. nemestrina* and species from the *nigra* group (**Fig. 3B** and **supplementary table S8**). Overall, the results suggested that introgression in macaques are pervasive not only in the past but also more recently.

Population size history estimate

The history of the effective population size (N_e) for macaque species were modeled from the distribution of heterozygous sites across the genome using a pairwise sequentially Markovian coalescent (PSMC) model [57] (**Fig. 4B-4E** and **supplementary Fig. S16**). With the exception of *M. sylvanus* (**Fig. 4E**), which showed a truncated plot and most likely due to insufficient sequencing depth, the N_e of remaining macaque species are very similar from 5 Mya until ~ 2.5 Mya, indicating that all Asian macaque species shared the same demographic history during the Pliocene. The Pliocene was a period of global cooling after the warmer Miocene and the cooling and drying of the global environment may have contributed to the enormous spread of savannas during this time [58]. The change in vegetation undoubtedly makes a huge impact on arboreal animals, such as macaques. Subsequently, the ancestor of most macaque species showed an upward trend of N_e after ~2.5 Mya. However, the N_e of three species of the *nigra* group remained stable until 1.5 Mya with a subsequent upward trend at different rates, possibly corresponding to population growth and expansion once ecological conditions allowed. Macaques except of the *nigra* group experienced a population growth until the Pleistocene glaciations and the *Ne* of different macaque species declined during different glaciations indicating that they likely were restricted to smaller refugia during glacial periods. During the last interglacial period, N_e for *M. mulatta* (CR), *M. assamensis, M. nemestrina* and *M. tonkeana* increased rapidly but subsequently declined again.

Discussion

Phylogeny and classification of species groups

The present study enables us to verify the validity of species groups within the *Macaca* genus and to reconstruct their evolutionary history, which represents an exciting evolutionary scenery as a

primate taxon whose "Out of Africa" spread mirrors that of non-African humans. The species tree deduced by ASTRAL supports the division of macaques into seven well-supported clades or lineages that correspond to the classification of the genus into seven species groups [21, 29] (**Fig. 1B**). *M. sylvanus* diverged first and F_{ST} between *M. sylvanus* and the Asian macaques is higher than any pairs of Asian macaque groups (**supplementary table S6**). The results support the separation of *sylvanus* from the *silenus* group, which also confirmed the classification of Delson [23] based on morphology and geographic distribution. Asian macaques contained the *silenus* and *nigra* lineages and a clade including all other Asian lineages (**Fig. 1B**). The *silenus* and *nigra* lineages were grouped into a single species group (*silenus* group) by Fooden [22] and Delson [23], but recently separated into two distinct species groups: *silenus* and *nigra* or Sulawesi [21, 25, 29, 33]. Based on the species trees, the six investigated species of the *silenus* and *nigra* groups cluster into two clades: *silenus* groups (*M. leonina, M. silenus, M. nemestrina*) and *nigra* groups (*M. maura, M. tonkeana* and *M. nigra*) (**Fig. 1B**).

After the *silenus/nigra* clade diverged, the other Asian macaques split into two further clades. The first one includes *M. fascicularis, M. fuscata, M. cyclopis* and *M. mulatta* (**Fig. 1B**). Initially, all these four species were grouped into the *fascicularis* group [22], but Groves [25] separated *M. fuscata, M. cyclopis* and *M. mulatta* apart from *fascicularis* group into the *mulatta* group. The F_{ST} values vary widely between *M. fascicularis* vs *M. mulatta* (0.171) and *M. fascicularis* vs *M. fuscata* (0.415). Yan, et al. [41] indicated that about 30% genome of mainland *M. fascicularis* are of Chinese *M. mulatta* origin, which shows that the lower genetic distance between *fascicularis* and *mulatta* groups is reasonable and likely caused by hybridization events. Therefore, they should still be divided into two different groups, particularly as *M. fascicularis* from the Sundaland region might not contain any genomic contribution of *M. mulatta*. The second clade contains *M. arctoides, M. assamensis, M. thibetana, M. leucogenys* and *M. sinica* (**Fig. 1B**). *M. arctoides* was assigned to the *arctoides* group [22], but based on molecular data, later integrated into the *sinica* group [23, 34, 35]. In contrast, Groves [25] recognized *M. arctoides* as member of the *fascicularis* group. Due to all these uncertainties, *M. arctoides* was recently separated into its own *arctoides* group [21, 29, 33].

Complicated gene flow

A recent study has reported gene flow events between the *fascicularis* and *silenus* groups and concluded that introgression between both groups can occur mainly because of genome similarity in closely related species [48]. In this study, we found that speciation of several macaques involved a combination of ILS and introgression. We also discovered that *D* statistics reveals more asymmetric results for *fascicularis/mulatta* and *silenus/nigra* lineages when four new genomes (*M. cyclopis*, *M. leonine*, *M. silenus* and *M. maura*) were added. We extended the knowledge of this hybridization events as an ancient introgression event which were verified by PhyloNet and "qpgraph" analyses (**Fig. 3C-D** and **supplementary fig. S12-S13**).

The observed differences in the *D* statistic of different species combinations have multiple possible origins. First of all, apart from ancient and recent introgression between macaque groups and species, secondary introgression can also result in differences in *D* statistic. For instance, *M. arctoides* carries an excess of shared derived alleles with *nigra* and *silenus* groups (**Fig. 3B** and **supplementary table S8**). *M. arctoides* is believed to be of hybrid origin with a large proportion (30%) of its genome likely of *fascicularis/mulatta* origin [34, 35, 36, 46]. As *fascicularis/mulatta* groups carry many shared derived alleles with *silenus/nigra* groups, secondary introgression events likely led to the introduction of *silenus/nigra* alleles into *M. arctoides*, accounting for the high *D* statistic value in *M. arctoides*. Secondly, non-zero *D* values could also be caused by ancestral population structure. For example, considering the non-overlapping distribution of *M. sylvanus* and other species, we suspected that introgression signals between *M. sylvanus* and other species were introduced by the ancestral population structure, especially for insular species such as *M. cyclopis* and *M. fuscata* (**Fig. 3B**). Last but not the least, the observed small differences in the *D* statistics among combinations of species group (**Figs. 3B** and **supplementary tables S8-S10**) could be affected by genetic differences, drift and selection [6]. In addition, rapid diversification is often affected by incomplete lineage sorting and small sample size, which in turn limited our judgment on the origins of gene flow events. Thus, it is expected that inconsistent results were obtained when different methods were employed. Further examination of introgression in macaques should take those factors into account as well and include population-level samples for each species.

The radiation of macaques

Fossil data suggest that the genus *Macaca* arose about 7 Mya in North Africa during the Late Miocene [23, 59, 60, 61], which is consistent with our MCMCTree result (7.76 Mya, **Fig. 2**). The Messinian Salinity Crisis (5.9-5.3 Mya), in which the level of the Mediterranean Sea fluctuated and occasionally dried out, may have facilitated macaque dispersal from North Africa [51]. A large number of macaque fossils have been reported in Europe near the Mediterranean [33, 51]. Macaque ancestors likely invaded Europe via the arid Mediterranean during Messinian stage [26, 33, 46, 51]. The ancestor of *M. sylvanus* split from the main stem ~5.29 Mya and settled around the Mediterranean (**Fig. 2** and **Fig. 5**). Dental specimens of the oldest known Asian fossil referred to the genus *Macaca* are *M. palaeindicus* from the Late Pliocene Tatrot formation of the Siwalik Hills [23, 59, 62]. Hence, the ancestor of Asian macaques spread eastward step by step likely along the Siwalik Hills to the Hengduan Mountains. Asian macaques then diverged into three lineages. First, the ancestor of *silenus/nigra* split ~4.01 Mya, followed shortly afterwards (~3.79 Mya) by the ancestors of the *fascicularis/mulatta* and *arctoides/sinica* lineages. Subsequent hybridization events between *silenus/nigra* and *fascicularis/mulatta* might have occurred until the *silenus* and *nigra* groups finally diverged ~3.2 Mya. The Hengduan Mountains, which are a potential diversification hotspot for primates [1, 63, 64, 65, 66], limited movement to the north and east so that macaques spread southward and invaded the Indochinese Peninsula (**Fig. 5B-5D**) [62].

M. nemestrina is mitochondrially polyphyletic with populations from Borneo clustering with Sulawesi macaques [26, 61]. Therefore, macaques likely spread to the Sundaland and invaded Sulawesi via eastern Borneo [33, 67] (**Fig. 5B**). A large northern hemisphere ice sheets have existed ~3.3 Mya resulting in lowered sea-levels in the Sundaland region [68], facilitating the ancestor of the *nigra* group migrating southward and invading Sulawesi, while separating from the *silenus* group (~3.2 Mya; **Fig. 5B**). Isolated on Sulawesi, the *nigra* group finally diversified into different species during the Pleistocene (**Fig. 2** and **Fig. 5B**). The N_e of species of the *nigra* group underwent a bottleneck while the ancestor of the *silenus* group experienced an expansion after both groups have diverged (**supplementary Fig. S16**). Presumably, during the process of expansion, an ancestral population of the *silenus* group came into secondary contact with the ancestors of the *fascicularis/mulatta* and *arctoides/sinica* lineages and hybridized with them (**Fig.**

3). At that time, also the ancestors of the *fascicularis/mulatta* and *arctoides/sinica* lineages experienced population expansions (**supplementary Fig. S16**). Tosi, et al. hypothesized that extant *M. silenus* apparently derived from genetic stocks represented by modern *M. nemestrina* populations [26]. Our data show that *M. nemestrina* split from the ancestors of *M. leonina* and *M. silenus* ~2.39 Mya (**Fig. 2** and **Fig.5B**) and the latter invaded South India (**Fig. 5B**).

M. arctoides split from the ancestor of the *sinica* group ~3.01 Mya and spread into Indochina and the Malay Peninsula (**Fig. 5C**) [21,23, 62, 69, 70]. Soon thereafter, *M. arctoides* hybridized with the ancestor of the *fascicularis/mulatta* lineage [34] until both groups diverged. Fan, et al. [36] hypothesized a special hybrid origin of *M. arctoides* which was produced by female *M. mulatta* invading proto-*arctoides* populations and proposed that *M. fascicularis* from Vietnam exhibited significant gene flow with *M. arctoides* probably because ~30% of the *M. fascicularis* genome was of Chinese rhesus macaque origin [41]. Our *D* statistic results show that the signals for gene flow between *fascicularis* and *arctoides* are almost the same when compared with that between *arctoides* and the *mulatta* group (**supplementary table S9**), suggesting that gene flow occurred before the split of the *fascicularis* and *mulatta* groups. Whole genome data also suggest that, after the split from the ancestor of *M. arctoides*, populations of the *sinica* group spread into India and Sri Lanka ~2.56 Mya, while others expanded eastwards and spread into China, especially into the Tibet regions [70], and diversified into various macaque species (**Fig. 5C**). Due to the overlap in geographic distribution, gene flow among species of the *sinica* group was common, especially between *M. assamensis* and *M. thibetana* (**Fig. 1B-1C and supplementary Fig. S8**).

The ancestor of *M. fascicularis* spread from Indochina to the Malay Peninsula and Sundaland (**Fig. 5D**), migration events that might be closely linked to sea level changes [71]. Morales, et al. [24] suggests that *M. mulatta* likely originated from a *fascicularis*-like ancestor ~2.5 Mya in Indochina, from which it dispersed to its current range [24]. Our data show that differentiation of the *fascicularis/mulatta* clade into *fascicularis* and *mulatta* groups occurred ~2.94 Mya (**Fig. 5D** and **Fig. 2**). Subsequently, the ancestor of the *mulatta* group diversified into *M. fuscata, M. cyclopis* and *M. mulatta* during 2.34-1.86 Mya (**Fig. 2**). Both *M. fuscata* and *M. cyclopis* are thought to be derived from ancient *mulatta*-like populations in eastern Asia and the paraphyly observed from genetic data for *M. mulatta* suggests that *M. fuscata* and *M. cyclopis* likely

originated from different founder populations [72]. Ito, et al. found a maxillofacial fragment from Korea which was morphologically most similar to Japanese *M. fuscata* [73], suggesting a previously larger range for *M. fuscata* and that *M. fuscata* invaded Japan via northern China and Korea.

Conclusions

In this study, we conducted a phylogenetic analysis of 24 whole macaque genomes of which 16 are newly generated. We identified ~20 million high-quality SNVs in each macaque individual, which can be invoked as important baseline data in other studies. We support the division of macaques into seven species groups and found that there were visible differences in F_{ST} within and between species groups. Importantly, various ancient and recent introgression events between different species and species groups were identified. Most prominently, the contradicting phylogenetic position of the *silenus/nigra* lineage in the macaque phylogeny is likely the result of introgression between the ancestors of the *fascicularis/mulatta* and *silenus/nigra* lineages. In summary, this study suggests that, in parallel to ILS, introgression plays an important role in the radiation of macaques.

Materials and methods

Sample collection and library preparation

Samples of 16 macaque individuals were obtained from the Institute of Zoology, Chinese Academic of Sciences and the German Primate Center (**supplementary table S1**). In addition, genome data of eight other macaque individuals and one *Papio anubis* individual were obtained from the NCBI Short Read Archive (accession no. show in **supplementary table S1**). The present study was approved by the Animal Ethics Committee of the Institute of Zoology, Chinese Academy of Sciences. The procedure of blood and tissue collection was in strict accordance with the Animal Ethics Procedures and Guidelines of the People's Republic of China. DNA was isolated from blood or tissue samples with the Qiagen Blood and Tissue Kit. Sequencing libraries were prepared based on KAPA library preparation kit with insert sizes of 300-500 bp and sequenced using Illumina Hi-Seq 2500 and 4000 technologies. Quality control was performed

using FastQC with default parameters (www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Data Processing

Paired-end reads for all samples were mapped to the *Papio anubis* reference genome (Panu_3.0, GCA_000264685.2) with Burrows-Wheeler Aligner (BWA) mem v0.7.17 [74], except for *M. nemestrina* (2), which was trimmed by cutadapt.2.4 [75] firstly because of the poor sequencing quality due to DNA degradation. The output bam files were sorted using samtools [76] and duplicates were marked with GATK-v4.1.2.0 [77]. Single-nucleotide variants (SNVs) were called using GATK 4.1 [77] following best practice. We obtained the GVCF file for each individual using the "HaplotypeCaller" method in GATK and then all samples were merged based on "CombineGVCFs". We used the GenotypeGVCFs-based method with the "includeNonVariantSites" flag to get the vcf files. After that, we applied the "SelectVariants" to exclude indels and split the variant and nonvariant sites. All nonvariant sites were filtered in subsequent analyses. We applied the hard filter command "VariantFiltration" to exclude potential false-positive variant calls with the following criteria: "-filterExpression QD< 5.0 || FS > 60.0 || MQ < 40.0 || ReadPosRankSum< -8.0 || MQRankSum < -12.5" and "-genotypeFilterExpression DP < 4.0" [78]. All variant calls that failed the filters were removed. All SNVs were annotated by ANNOVAR v2013–06-21 [78]. (**supplementary tables S3** and **S4**) based on gff3 file from Ensembl (Papio_anubis.Panu_3.0.104.gff3). Heterozygosity is defined as the rate of heterozygous SNVs in the genome. We calculated the heterozygosity rate in a nonoverlapping window (window = 100 kb) for all 24 macaques. Genome-wide distribution of heterozygosity was shown in a box plot using boxplot in R package (**supplementary table S2** and **supplementary Fig. S2**).

Phylogenetic reconstruction using the window-based strategy

We constructed the consensus sequences for all individual genomes and aligned them per chromosome (sex chromosomes, mitogenomes and scaffolds were excluded) based on VCF, which did not include variants present in the Papio reference but were invariant among macaques. In addition, indels, repetitive regions and low-quality regions were removed from the consensus sequences. Per-chromosome alignments were split into nonoverlapping windows of 200 kbp. The maximum likelihood (ML) trees for each window were performed with RAxML v8 [80] with 100 bootstrap replicates and the GTR + CAT substitution model. From all 200 kbp window trees, we reconstructed a species tree with ASTRAL v4.10.5 [81] under the multispecies coalescent model

returning quartet scores and posterior probabilities. The species tree was rooted with *Papio anubis*. Consensus networks of the window trees were generated using SplitsTree5 [40] with different thresholds. Genealogical discordance across the genome of rhesus macaque was assessed by Twisst [82].

Phylogenetic analysis of concatenated whole-genome SNV data

SNVs were extracted from the filtered VCF file and converted into individual FASTA sequences using a python script (https://doi.org/10.5281/zenodo.2540861). Merging all FASTA sequences into a single file provided a multiple sequence alignment of all individuals and all concatenated SNVs. Phylogenetic trees were reconstructed with maximum-likelihood (ML) using RAxML v8 [79] with 100 bootstrap replicates and the GTR + CAT substitution model. In addition, a neighbor-joining (NJ) tree was constructed using the program TreeBeST V.1.9.2 (http://treesoft.sourceforge.net/treebest.shtml), which has a build-in algorithm to build the best tree that reconciled with the species tree and rooted with the minimized number of duplications and losses. To further visualize genetic relationships among macaques, we performed principal components analysis (PCA) on the filtered autosomal SNVs using the Eigensoft package, version 5.0 [83].

Phylogeny of whole mitochondrial genomes

We used NOVOPlasty [84] to *de novo* reconstruct the mitochondrial genomes (mitogenomes) for 18 macaques representing 15 species (**supplementary table S5**) and then annotated the genome sequences using MitoZ [85]. In addition, we downloaded 15 mitogenomes for the genus *Macaca* (**supplementary table S5**) from NCBI based on the filter criteria mentioned in Roos, et al. [33]. The 33 mitogenomes were aligned using MAFFT version 7 [86] and indels and poorly aligned positions were removed with Gblocks version 0.91b [87]. A ML tree was reconstructed using RAxML v 8 [79] with a GTR + CAT substitution model and applying 1000 bootstrap replicates to verify nodal support.

***D* statistic**

We applied the *D* statistic analysis [53] to evaluate the imbalanced frequency of alleles using the "qpDstat" subroutine in the AdmixTools v1.01 [88]. To examine possible gene flow among macaques, we used *Papio anubis* as the outgroup (Out) and three different macaque species with a phylogenetic topology following (((M1, M2), M3), Out), which were extracted from the species

501 tree through the ape package in R [89]. In total, 455 groupings were analyzed of which 424 had
502 $Z_{abs} > 3$ indicating statistically significant ($P < 0.001$) deviations from $D = 0$ and suggesting gene
503 flow between M3 and M2.

504 **Qpgraph**

505 Admixtools2 (v2.0.0) is a new and fast implementation of Admixtools package [90], which
506 preform unbiased comparison of any two qpgraph models using out-of-sample scores
507 (https://uqrmaie1.github.io/admixtools/index.html). We chose one SNP per 100 bp region from a
508 sample of genome-wide SNPs, yielding a dataset of 22,823,254 snps in total. We first created a
509 global estimation matrix for *f*2-statistics (Q matrix). This matrix was then used to fit the model
510 with different admixture edges. We first started with a species tree and fixed the outgroup as
511 baboons. Then, we added extra admixture edges one at a time based on evidence from *D* statistic.
512 The diagonal entries of the predicted covariance matrix of the f-statistics (Q matrix) were added
513 by a small number (0.0001) to stabilize the matrix inversion. Based on the score and worst
514 residual, different admixture edges and models are compared.

515 **Maximum likelihood inference for reticulation with PhyloNet**

516 We used the InferNetwork_MPL program in PhyloNet [54], which is based on Maximum Pseudo
517 Likelihood (MPL) algorithm, to analyze a set of every 3th GF ML tree, that is, 2150 trees in a
518 coalescent framework that accounts for incomplete lineage sorting (ILS) while allowing different
519 numbers of reticulation events. In addition, the individual "2" in all macaque species were pruned
520 from the input gene trees to reduce complexity and computational demand. We set the numbers of
521 reticulation events as 1 and 2 to conduct two independent analyses and each analysis was run with
522 100 iterations, yielding the five networks with the highest likelihood scores. We visualized the
523 optimal networks in Dendroscope v3 [91] and present it in **Fig. 3C** and **supplementary Fig. S13.**

524 **QuIBL**

525 We evaluated the likelihood of introgression and ILS for each locus in each species triplet using
526 QuIBL [18]. As the input data set for QuIBL analysis, we used 2,000 randomly chosen trees out of
527 the 6351 trees acquired from the genome-wide split window. To guarantee that the sampling was
528 biased-free, this sampling was carried out three times. In order to determine whether ILS (scenario
529 1) or a combination of ILS and introgression reflected the inner branch lengths, QuIBL evaluated
530 the probability values (Bayesian Information Criterion test, BIC) (scenario 2). The scenario of ILS

only with the lower BIC value is preferred when Delta. BIC ($BIC_{scenario2}$-$BIC_{scenario1}$) is larger than 10 and the BIC value is less than 0. The combination of ILS and introgression with the lower BIC value, however, is preferred when Delta.BIC is smaller than -10.

F_d

To identify putatively introgressed regions in macaques, we screened these segments according to the strategy of Martin, et al. [55] and van der Valk, et al. [6]. The method predicts that introgression between species at a specific genome region should reduce the between-species divergence in this region as compared to the rest of the genome [6, 55]. We computed the F_d value, which is sensitive to the number of informative sites within windows, for each 10-kb window across the whole genome of each macaque. Divergence (*Dxy*) was calculated for each window between M2 and M3 for each comparison in the *D* statistic. As *Dxy* is a measure of sequence divergence, introgressed windows are expected to have a relatively low *Dxy*, while the opposite is true for the background windows without introgression. We relied on the cutoff $P < 0.05$ of the F_d value to identify introgressed and background windows firstly. Additional, divergence (*Dxy*) was calculated for each window between M2 and M3 for each comparison in the *D* statistic. However, since the high divergence of *Dxy* could also be caused by incomplete lineage classification, or simply the characterization of a greater diversity of genomic regions, *Dxy* only use as an option value to confirm putatively introgressed regions.

Introgression analysis with efficient local ancestry inference

The two-layer admixture model implemented in efficient local ancestry inference [56] was used to infer introgressed segments in *fascicularis/mulatta* (the admixture population, -p 1) with the *silenus/nigra* as source population 2 (-p 11) and *arctoides*/*sinica* were used as source population 1 (-p 10). We run 30 iterations with mixture generation values of 2,910,000, 3,200,000 and 3,450,000, which was estimated based on MCMCTree result, and the averaged results over the three independent runs were used for detecting introgression tracts. Sites with a proportion of source population 2 ancestry greater than 1.5 were defined as introgression sites. We sum the proportion of source population 2 ancestry in each 10 kb windows, which is same to F_d and the top 1% windows were defined as putatively introgression tracts.

Analysis of demographic history using genome-wide data

We applied the pairwise sequentially Markovian coalescent (PSMC, v0.6.5-r67) analysis [57] to

estimate the demographic history of each macaque individual (**supplementary Fig. S2**). We use the Samtools mpileup [76] to generate the input files for PSMC with the parameter "-C" set to be 50 to improve mapping quality. In addition, only autosomal SNVs were reserved to generate the diploid sequence for PSMC. PSMC was run with 25 iterations (-N), a maximum 2N0 coalescent time (-t) of 15, an initial theta/rho ratio (-r) of 5, and the 64 time-intervals were parameterized as "4 + 25*2 + 4 + 6". PSMC plots were scaled with a mutation rate (μ) of 0.9×10^{-8} and a generation time (g) of 11 years for all macaque species [92]. Bootstrap analyses (100 bootstrap replications) were performed for each species (**supplementary Fig. S16**).

Divergence time calibration

Coding sequences (CDSs) and protein sequences for different representative outgroup species and three macaques (*M. mulatta, M. fascicularis, M. nemestrina, Papio anubis, Homo sapiens, Pan troglodytes* and *Pongo abelii*) were retrieved from Ensembl (release-98). The single-copy orthologue gene list shared by these species was obtained using OrthoFinder [93]. The macaques were mapped to the Panu_3.0 genome and thus their CDSs have the same genomic coordinates. Therefore, the CDSs of all macaques were extracted after ortholog detection. CDSs of these single-copy genes were aligned by PRANK (v.170427) [94] under codon model and concatenated into a single super-matrix. Fourfold degenerate (4D) sites were extracted from them using MEGA7 [95] and concatenated to calibrated the divergence time of the genus *Macaca* with MCMCTree program in PAML v4.9 (Phylogenetic Analysis of ML, v4.9) package [50]. Four calibration points were applied to constrain the divergence time of the nodes: Catarrhini 20–38 Ma, Hominidae 13–18 Ma, *Homo-Pan* 6–7 Ma and *Papio-Macaca* 5–8 Ma [33, 96]. To check for convergence of the stationary distribution, the analysis was run in duplicate, and the results were compared between runs and similar results for divergence time of macaques were obtained (**supplementary Fig. S10**).

Funding

The work was supported by the National Natural Science Foundation of China (31821001 and 31530068), the Strategic Priority Research Program of Chinese Academy of Sciences (XDB31000000, XDA23080000 and XDA19050000), and Sino-German Mobility Programme (M-0084).

Acknowledgements

The authers thank Prof. Luo Shujin and Dr. Sun Xin for their assistance in DNA isolation, library preparation and quality control for *M. nemestrina (2)*.

Data accessibility

Whole genome resequencing data produced in this study were deposited at the NCBI Sequence Read Archive (SRA) under Bioproject PRJNA732172.

Author contributions

ML conceived this study, and X.Z. and C.R. contributed to the idea for this study. X.T., J.Q. and Z.L. performed experiments, analyzed data and wrote the draft manuscript. L.Z. analyzed data. P.F. C.R. and J.L. provided samples. M.L., X.M. and C.R. revised the draft manuscript. All authors contributed to the final text.

Conflicts of interest

The authors declare that there is no conflict of interest.

References

1. Roos C, Zinner D, Kubatko LS, Schwarz C, Yang M, Meyer D, Nash SD, Xing J, Batzer MA, Brameier M, et al. 2011. Nuclear versus mitochondrial DNA: evidence for hybridization in colobine monkeys. *BMC Evol Biol.* 11:77.

2. Zinner D, Arnold ML, Roos C. 2011. The strange blood: natural hybridization in primates. *Evol Anthropol.* 20:96-103.

3. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, et al. 2013. Hybridization and speciation. *J Evol Biol.* 26:229-246.

4. Tung J, Barreiro LB. 2017. The contribution of admixture to primate evolution. *Curr Opin Genet Dev.* 47:61-68.

5. Fontsere C, de Manuel M, Marques-Bonet T, Kuhlwilm M. 2019. Admixture in Mammals and How to Understand Its Functional Implications: On the Abundance of Gene Flow in Mammalian Species, Its Impact on the Genome, and Roads into a Functional Understanding. Bioessays 41:e1900123.

6. van der Valk T, Gonda CM, Silegowa H, Almanza S, Sifuentes-Romero I, Hart TB, Hart JA, Detwiler KM, Guschanski K. 2020. The Genome of the Endangered Dryas Monkey Provides New Insights into the Evolutionary History of the Vervets. *Mol Biol Evol.* 37:183-194.

7. Vanderpool D, Minh BQ, Lanfear R, Hughes D, Murali S, Harris RA, Raveendran M, Muzny DM, Hibbins MS, Williamson RJ, et al. 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biol.* 18:e3000954.

8. Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, Babrzadeh F, Gharizadeh B, Luo M, Plummer FA, et al. 2011. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334:89-94.

9. Abbott RJ, Barton NH, Good JM. 2016. Genomics of hybridization and its evolutionary consequences. *Mol Ecol.* 25:2325-2332.

10. Zheng Z, Wang X, Li M, Li Y, Yang Z, Wang X, Pan X, Gong M, Zhang Y, Guo Y, et al. 2020. The origin of domestication genes in goats. *Sci Adv.* 6:eaaz5216.

11. Figueiró HV, Li G, Trindade FJ, Assis J, Pais F, Fernandes G, Santos SHD, Hughes GM, Komissarov A, Antunes A, et al. 2017. Genome-wide signatures of complex introgression and

adaptive evolution in the big cats. *Sci Adv.* 3:e1700299.

12. Árnason Ú, Lammers F, Kumar V, Nilsson MA, Janke A. 2018. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Sci Adv.* 4:eaap9873.

13. Avise JC, Robinson TJ. 2008. Hemiplasy: a new term in the lexicon of phylogenetics. *Syst Biol* 57:503-507.

14. Maddison WP, Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. *Syst Biol* 55:21-30.

15. Szöllősi GJ, Tannier E, Daubin V, Boussau B. 2015. The inference of gene trees with species trees. *Syst Biol* 64:e42-62.

16. Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Mol Biol Evol* 27:570-580.

17. Feng S, Bai M, Rivas-González I, Li C, Liu S, Tong Y, Yang H, Chen G, Xie D, Sears KE, et al. 2022. Incomplete lineage sorting and phenotypic evolution in marsupials. *Cell* 185:1646-1660.e1618.

18. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* 366:594-599.

19. Wu DD, Qi XG, Yu L, Li M, Liu ZJ, Yoder AD, Roos C, Hayakawa T, Rogers J, Marques-Bonet T, et al. 2022: Initiation of the Primate Genome Project. *Zool Res.* 43: 147-149.

20. Rogers J, Raveendran M, Harris RA, Mailund T, Leppälä K, Athanasiadis G, Schierup MH, Cheng J, Munch K, Walker JA, et al. 2019. The comparative genomics and complex population history of *Papio* baboons. *Sci Adv.* 5:eaau6947.

21. Zinner D, Fickenscher GH, Roos C. 2013. Family Cercopithecidae (Old World Monkeys). In: Mittermeier RA, Rylands AB, Wilson DE, editors. Handbook of the Mammals of the World. Volume 3, Primates. Barcelona, Lynx Edicions, p.550-627.

22. Fooden J. 1976. Provisional classifications and key to living species of macaques (Primates: *Macaca*). *Folia Primatol.* 25:225-236.

23. Delson E. 1980. Fossil macaques, phyletic relationships and a scenario of deployment. In: D. Lindburg D, editor. The Macaques. Studies in Ecology, Behavior, and Evolution. New York: van

Nostrand Rheinhold, p.10-30.

24. Morales JC, Melnick DJ. 1998. Phylogenetic relationships of the macaques (Cercopithecidae: *Macaca*), as revealed by high resolution restriction site mapping of mitochondrial ribosomal genes. *J Hum Evol*. 34:1-23.

25. Groves CP. 2001. Primate Taxonomy. Washington DC: Smithsonian Institution Press.

26. 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.

27. Sinha A, Datta A, Madhusudan MD, Mishra C. 2005. *Macaca munzala*: A new species from western Arunachal Pradesh, northeastern India. *Int J Primatol.* 26:977-989.

28. Chakraborty D, Ramakrishnan U, Panor J, Mishra C, Sinha A. 2007. Phylogenetic relationships and morphometric affinities of the Arunachal macaque *Macaca munzala*, a newly described primate from Arunachal Pradesh, northeastern India. *Mol Phylogenet Evol.* 44:838-849.

29. Roos C, Boonratana R, Supriatna J, Fellowes JR, Groves CP, Nash SD, Rylands AB, Mittermeier RA. 2014. An updated taxonomy and conservation status review of Asian primates. *Asian Primates J.* 4(2):1-38.

30. Li C, Zhao C, Fan PF. 2015. White-cheeked macaque (*Macaca leucogenys*): A new macaque species from Medog, southeastern Tibet. *Am J Primatol.* 77:753-766.

31. Hou W, Liu S, Jiang J, Fan Z, Fan P, Li J. 2016. The complete mitochondrial genome of White-cheeked macaque (*Macaca leucogenys*). *Mitochondrial DNA B Resour.* 1:374-375.

32. Fan P, Liu Y, Zhang Z, Zhao C, Li C, Liu W, Liu Z, Li M. 2017. Phylogenetic position of the white-cheeked macaque (*Macaca leucogenys*), a newly described primate from southeastern Tibet. *Mol Phylogenet Evol.* 107:80-89.

33. Roos C, Kothe M, Alba DM, Delson E, Zinner D. 2019. The radiation of macaques out of Africa: Evidence from mitogenome divergence times and the fossil record. *J Hum Evol.* 133:114-132.

34. Tosi AJ, Morales JC, Melnick DJ. 2000. Comparison of Y chromosome and mtDNA phylogenies leads to unique inferences of macaque evolutionary history. *Mol Phylogenet Evol.* 17:133-144.

35. Li J, Han K, Xing J, Kim HS, Rogers J, Ryder OA, Disotell T, Yue B, Batzer MA. 2009.

Phylogeny of the macaques (Cercopithecidae: *Macaca*) based on *Alu* elements. *Gene* 448:242-249.

36. Fan Z, Zhou A, Osada N, Yu J, Jiang J, Li P, Du L, Niu L, Deng J, Xu H, et al. 2018. Ancient hybridization and admixture in macaques (genus *Macaca*) inferred from whole genome sequences. *Mol Phylogenet Evol.* 127:376-386.

37. Ciani AC, Stanyon R, Scheffrahn W, Sampurno B. 1989. Evidence of gene flow between Sulawesi macaques. *Am J Primatol.* 17:257-270.

38. Evans BJ, Supriatna J, Melnick DJ. 2001. Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. *Evolution* 55:1686-1702.

39. Evans BJ, Supriatna J, Andayani N, Setiadi MI, Cannatella DC, Melnick DJ. 2003. Monkeys and toads define areas of endemism on Sulawesi. *Evolution* 57:1436-1443.

40. Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 23:254-267.

41. Yan G, Zhang G, Fang X, Zhang Y, Li C, Ling F, Cooper DN, Li Q, Li Y, van Gool AJ, et al. 2011. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nat Biotechnol.* 29:1019-1023.

42. Hamada Y, Yamamoto A, Kunimatsu Y, Tojima S, Mouri T, Kawamoto Y. 2012. Variability of tail length in hybrids of the Japanese macaque (*Macaca fuscata*) and the Taiwanese macaque (*Macaca cyclopis*). *Primates* 53:397-411.

43. Fan Z, Zhao G, Li P, Osada N, Xing J, Yi Y, Du L, Silva P, Wang H, Sakate R, et al. 2014. Whole-genome sequencing of tibetan macaque (*Macaca Thibetana*) provides new insight into the macaque evolutionary history. *Mol Biol Evol.* 31:1475-1489.

44. Evans BJ, Tosi AJ, Zeng K, Dushoff J, Corvelo A, Melnick DJ. 2017. Speciation over the edge: gene flow among non-human primate species across a formidable biogeographic barrier. *R Soc Open Sci.* 4:170351.

45. Ito T, Kanthaswamy S, Bunlungsup S, Oldt RF, Houghton P, Hamada Y, Malaivijitnond S. 2020. Secondary contact and genomic admixture between rhesus and long-tailed macaques in the Indochina Peninsula. *J Evol Biol.* 33:1164-1179.

46. Tosi AJ, Disotell TR, Morales JC, Melnick DJ. 2003. Cercopithecine Y-chromosome data provide a test of competing morphological evolutionary hypotheses. *Mol Phylogenet Evol.*

27:510-521.

47. Matsudaira K, Hamada Y, Bunlungsup S, Ishida T, San AM, Malaivijitnond S. 2018. Whole Mitochondrial Genomic and Y-Chromosomal Phylogenies of Burmese Long-Tailed Macaque (*Macaca fascicularis aurea*) Suggest Ancient Hybridization between *fascicularis* and *sinica* Species Groups. *J Hered.* 109:360-371.

48. Song Y, Jiang C, Li KH, Li J, Qiu H, Price M, Fan ZX, Li J. 2021. Genome-wide analysis reveals signatures of complex introgressive gene flow in macaques (genus Macaca). *Zool Res.* 42:433-449.

49. Bouckaert RR. 2010. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* 26:1372-1373.

50. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586-1591.

51. Elton S, O'Regan HJ. 2014. Macaques at the margins: the biogeography and extinction of *Macaca sylvanus* in Europe. *Quat Sci Rev.* 96:117-130.

52. Hamada Y, Urasopon N, Hadi I, Malaivijitnond S. 2006. Body size and proportions and pelage color of free-ranging *Macaca mulatta* from a zone of hybridization in northeastern Thailand. *Int J Primatol.* 27:497-513.

53. Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol.* 28:2239-2252.

54. Yu Y, Dong J, Liu KJ, Nakhleh L. 2014. Maximum likelihood inference of reticulate evolutionary histories. Proceedings of the National Academy of Sciences of the United States of America 111:16448-16453.

55. Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol Biol Evol.* 32:244-257.

56. Guan Y. 2014. Detecting structure of haplotypes and local ancestry. *Genetics* 196:625-642.

57. Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493-U484.

58. Salzmanna U, Williams M, Haywoodd AM, Johnsone A, Kenderc S, Zalasiewicz J. 2011. Climate and environment of a Pliocene warm world. *Palaeogeogr Palaeocl.* 309:1-8

59. Delson, E, Andrews P. 1975. Evolution and interrelationships of the catarrhine primates. In:

Luckett WP, Szalay FS, editors. Phylogeny of the Primates: A Multidisciplinary Approach. New York: Plenum Press. p.405–446.

60. Delson E. 1996. The oldest monkeys in Asia. In: Takenaka O, editors. International Symposium: Evolution of Asian Primates. Freude & Kyoto University, Primate Research Institute, Inuyama, Aichi, Japan, p.40.

61. Ziegler T, Abegg C, Meijaard E, Perwitasari-Farajallah D, Walter L, Hodges JK, Roos C. 2007. Molecular phylogeny and evolutionary history of Southeast Asian macaques forming the *M-silenus* group. *Mol Phylogenet Evol.* 42:807-816.

62. Fa JE. 1989. The genus *Macaca* – a review of taxonomy and evolution. *Mamm Rev.* 19:45-81.

63. Peng YZ, Pan RL, Jablonski NG. 1993. Classification and evolution of Asian colobines. Folia primatologica; *Int J Primatol.* 60:106-117.

64. Jablonski NG. 1998. The Natural History of the Doucs and Snub-Nosed Monkeys. Singapore: World Scientific Press.

65. Meijaard E, Groves CP. 2006. The geography of mammals and rivers in mainland Southeast Asia. In: Primate Biogeography: Progress and Prospects. Boston: Springer, p.305-329.

66. Thinh VN, Mootnick AR, Geissmann T, Li M, Ziegler T, Agil M, Moisson P, Nadler T, Walter L, Roos C. 2010. Mitochondrial evidence for multiple radiations in the evolutionary history of small apes. *BMC Evol Biol.* 10:74.

67. Meijaard E, van der Zon APM. 2003. Mammals of south-east Asian islands and their Late Pleistocene environments. *J Biogeogr.* 30:1245-1257.

68. Miller KG, Kominz MA, Browning JV, Wright JD, Mountain GS, Katz ME, Sugarman PJ, Cramer BS, Christie-Blick N, Pekar SF. 2005. The Phanerozoic record of global sea-level change. *Science* 310:1293-1298.

69. Fooden J. 1982. Ecogeographic segregation of macaque species. *Primates* 23:574-579.

70. Li B, He G, Guo S, Hou R, Huang K, Zhang P, Zhang H, Pan R, Chapman CA. 2020. Macaques in China: Evolutionary dispersion and subsequent development. *Am J Primatol.* 82:e23142.

71. Liedigk R, Kolleck J, Boker KO, Meijaard E, Md-Zain BM, Abdul-Latiff MAB, Ampeng A, Lakim M, Abdul-Patah P, Tosi AJ, et al. 2015. Mitogenomic phylogeny of the common long-tailed macaque (*Macaca fascicularis fascicularis*). *BMC Genomics* 16.

72. Chu JH, Lin YS, Wu HY. 2007. Evolution and dispersal of three closely related macaque species, *Macaca mulatta*, *M. cyclopis*, and *M. fuscata*, in the eastern Asia. *Mol Phylogenet Evol.* 43:418-429.

73. Ito T, Lee YJ, Nishimura TD, Tanaka M, Woo JY, Takai M. 2018. Phylogenetic relationship of a fossil macaque (*Macaca cf. robusta*) from the Korean Peninsula to extant species of macaques based on zygomaxillary morphology. *J Hum Evol.* 119:1-13.

74. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.

75. Kechin A, Boyarskikh U, Kel A, Filipenko M. 2017. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J Comput Biol* 24:1138-1143.

76. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.

77. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297-1303.

78. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43:11.10.11-11.10.33.

79. Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164-e164.

80. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313.

81. Sayyari E, Mirarab S. 2016. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Mol Biol Evol.* 33:1654-1668.

82. Martin SH, Van Belleghem SM. 2017. Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. *Genetics* 206:429-438.

83. Patterson N, Price AL, Reich D. 2006. Population structure and Eigenanalysis. *PLoS Genet.* 2:e190.

84. Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45:e18.

85. Meng G, Li Y, Yang C, Liu S. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res.* 47:e63.

86. Yamada KD, Tomii K, Katoh K. 2016. Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. *Bioinformatics* 32:3246-325

87. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17:540-552.

88. Zou Z, Zhang J. 2015. Are Convergent and Parallel Amino Acid Substitutions in Protein Evolution More Prevalent Than Neutral Expectations? *Mol Biol Evol.* 32:2085-2096.

89. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20:289-290.

90. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster, T, and Reich D. 2012. Ancient admixture in human history. *Genetics* 192:1065–1093.

91. Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol.* 61:1061-1067.

92. Xue C, Raveendran M, Harris RA, Fawcett GL, Liu X, White S, Dandouli M, Deiros DR, Below JE, Salerno W, et al. 2016. The population genomics of rhesus macaques (*Macaca mulatta*) based on whole-genome sequences. *Genome res.* 26:1651-1662.

93. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.

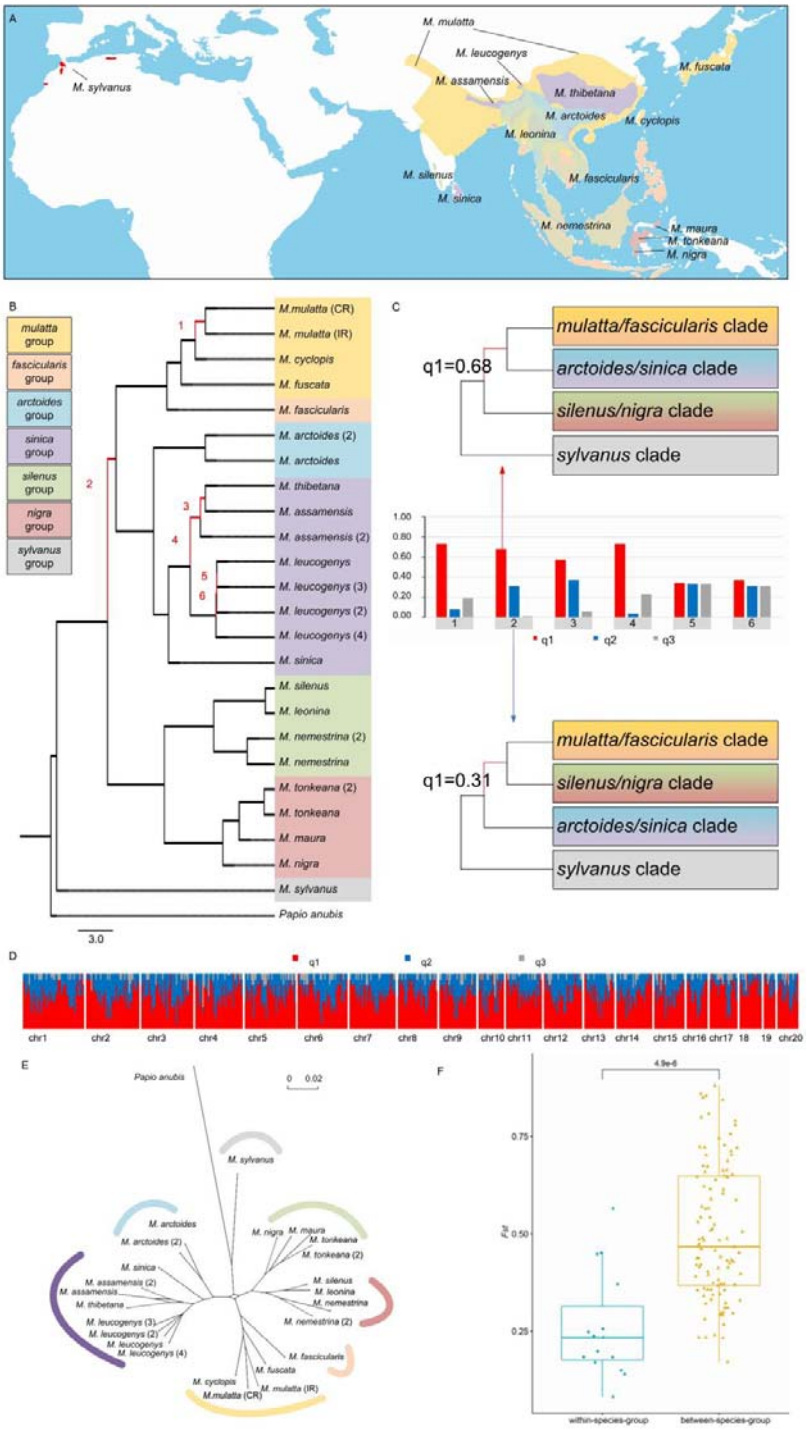
94. Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. *Methods Mol Biol.* 1079:155-170.

95. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 33:1870-1874.

96. Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira MAM, Kessing B, Pontius J, Roelke M, Rumpler Y, et al. 2011. A Molecular Phylogeny of Living Primates. *PLoS Genet.* 7:e1001342.

845 **Figures and Legends**

846



847

848 **Figure 1. Macaque phylogenetics and differentiation.**

849 (A) Distribution ranges of the 16 macaque species investigated in this study. The range overlaps

850 are shown in blended colors (orange, purple, green, blue and yellow). (B) A MSC species tree was

851 constructed from 6,351 individual genome fragments. Branches with low q1 (<0.8) are numbered
852 1 to 6. All branches receive maximal support (P = 1.0, ASTRAL analysis). Branch lengths were
853 calculated from a ML analysis. Different species groups are represented by different background
854 colors which are show to the left. (C) ASTRAL quartet-score analyses for branches 1 to 6 in Fig.
855 1B. Quartet scores were calculated for the three possible arrangements (q1 to q3) for the
856 respective branch. The principal quartet trees are depicted, with q1 representing the species tree.
857 Two alternative topologies for branch no. 2 as shown in Fig. 1C. (D) Genealogical discordance
858 across the genome of rhesus macaque, demonstrated by a window analysis (200-kb window size)
859 of a full-genome alignment mapped to Panu_3.0 (chromosomes showed at the bottom). The y axis
860 indicates the percentage of windows within a given interval (every 5 windows next to each other)
861 that conform to (red) or reject (blue and grey) the species tree. (E) Consensus networks for
862 macaques from 6,351 genome fragments with a minimum threshold of 30% to form an edge. (F)
863 Distributions of F_{ST} within (blue) and between (yellow) different macaque species groups. The
864 exact *P* value of above were computed based on Wilcoxon test.
865

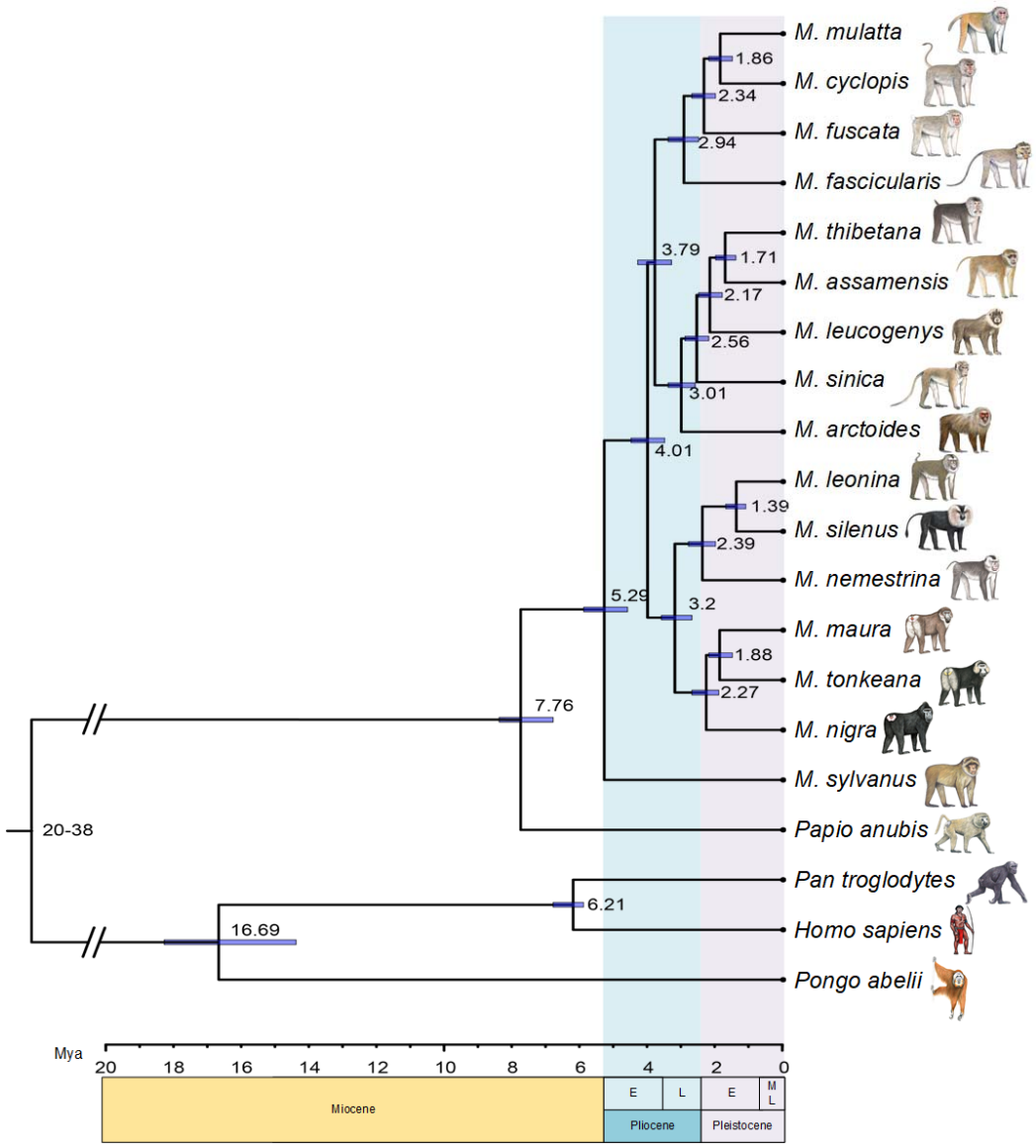


Figure 2. Divergence time tree of genus *Macaca*, estimated from 3,319,411 4d sites from 10,641 single-copy genes. Blue bars indicate 95% HPDs of divergence times and the time scale below shows million years ago (Mya). Abbreviations: L = late, M = middle, E = early. Four fossil-based calibration points are as follows: Catarrhini 20–38 Mya, Hominidae 13–18 Mya, *Homo-Pan* 6–7 Mya and *Papio-Macaca* 5–8 Mya.

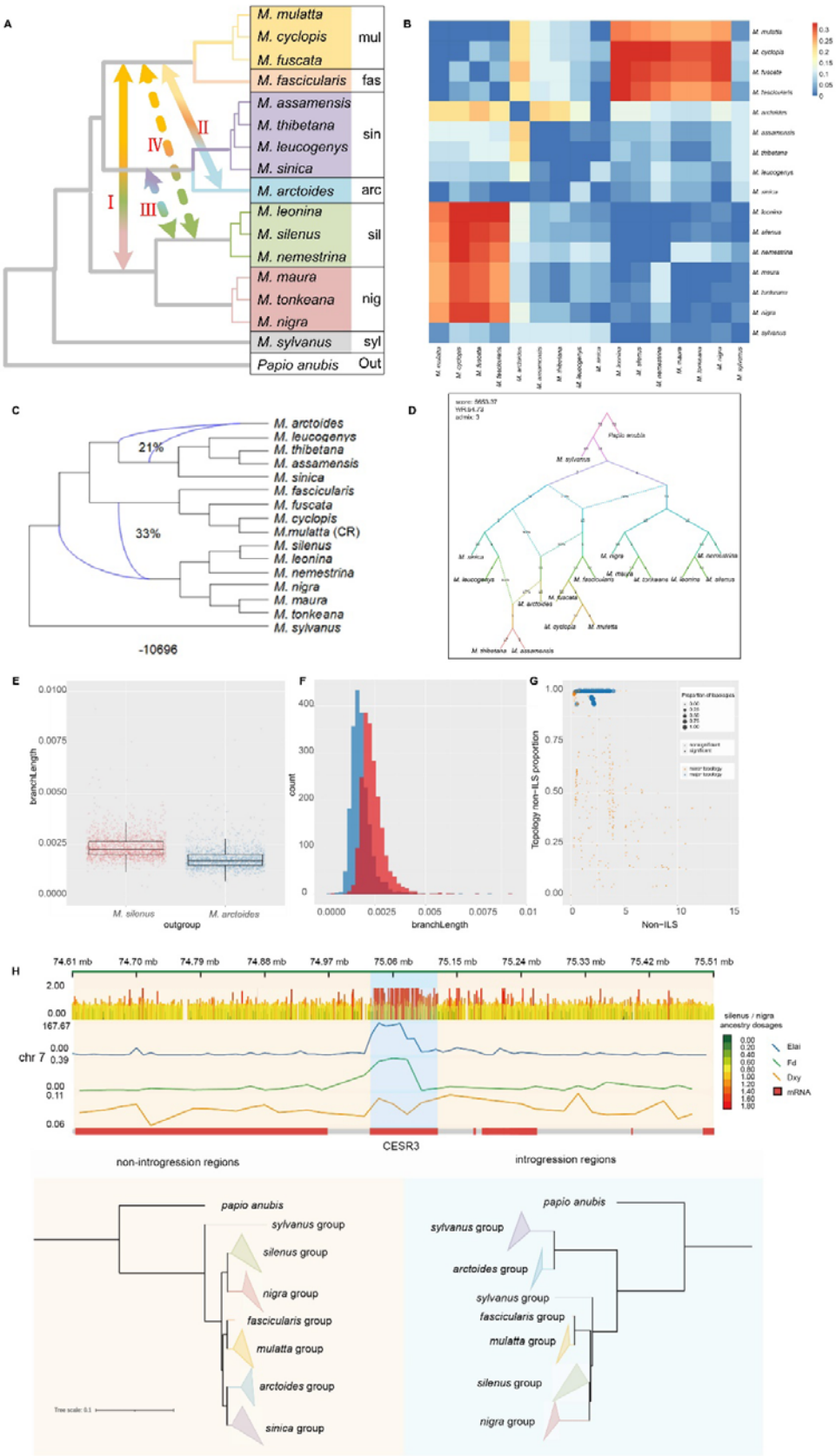


Figure 3 Gene flow among macaques.

(A) Overview on various gene flow events inferred from the samples included in this study based on *D* statistic. I, II, III and IV refers to major gene flow events and solid and dashed arrows represent results that could or could not be verified with qpgraph, respectively. (B) Highest *D* statistic for each species pairs. (C) Rooted networks for the *Macaca* phylogeny with reticulations inferred from PhyloNet. Reticulations are shown as blue arrows with inheritance probability denoted. Log-likelihood scores are shown below the networks. Notably, inheritance probability around 33% between *silenus*/*nigra* lineages and *fascicularis*/*mulatta* lineages resembles the distribution of quartet scores and the phylogenetic signals from genome fragments (Fig. 1). (D) The best qpgraph model with the smallest likelihood score and WR. (E-F) Distribution of internal branch lengths. Red represents the topology *M. silenus* is outgroup (concordant with species tree) and blue is *M. arctoides* as outgroup (consistent with introgression or ILS between *M. silenus* and *M. cyclopis*). (E) Boxplot of internal branch lengths, showing the strongly reduced mean branch length when *M. silenus* is the outgroup. (G) Relationship between internal branch length (Non-ILS) and non-ILS proportion of each topology. Size of each point corresponds to the proportion of trees in the respective triplet that recover the topology. Circles represent topologies significant for a non-ILS component, and diamonds represent those that are not significant. Blue triplets are concordant with the species tree. (H) The schematic for inferred of local ancestry for *fascicularis/mulatta* lineage. First line showed 1 Mb regions which contains introgressed windows. The second part, which is a heatmap show the ancestry dosages for *fascicularis/mulatta* source from *silenus/nigra* lineages. Sites with a proportion of source population (*silenus/nigra* lineage) ancestry greater than 1.5 were defined as introgression sites. The third line show the sum value of a proportion of source population (*silenus/nigra* lineage) ancestry in 10 kb windows. The fourth and fifth lines show the *Fd* and *Dxy* and the last part showed the genes in this region. The phylogenetic trees at the bottom of the figure showed the topologies based on the introgression windows and the non-introgression regions within this 1 Mb regions.

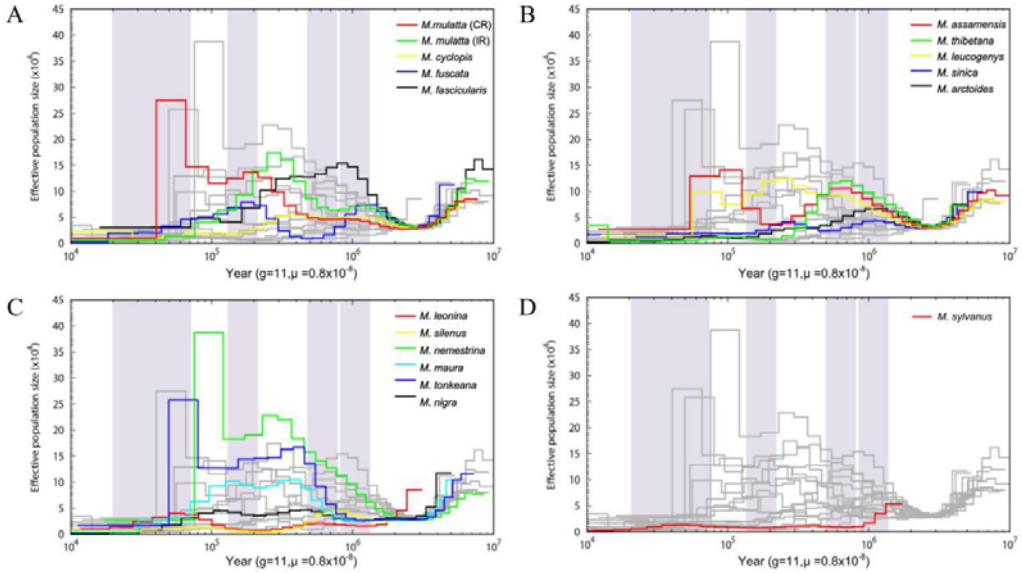


Figure 4 Historical effective population size.

(A-D) Historical effective population size using PSMC analyses for all macaque genomes. The x axis shows the time, and the y axis shows effective population size. Plots were scaled using a mutation rate (m) of 0.8×10^{-8} substitutions per nucleotide per generation and a generation time (g) of 11 years. Light purple shading indicates glacials in the Pleistocene and Holocene: Xixiabangma Glaciation (XG, 1.17~0.8 Mya), Naynayxungla Glaciation (NG, 0.78~0.50 Mya), Penultimate Glaciation (PG, 200–130 kya) and Last Glaciation (LG; 70–10 kya)

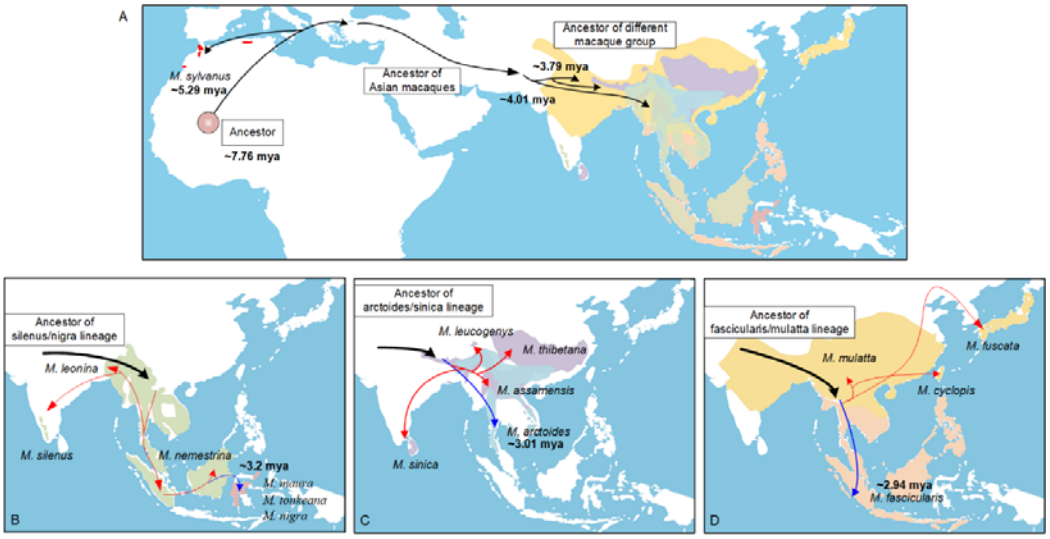


Figure 5 Scenario for the radiation of macaques.

(A) Macaques originated in Africa and after the split of *M. sylvanus* the ancestor of Asian macaques invaded Asia and initially split into three lineages, *silenus/nigra*, *arctoides/sinica* and *fascicularis/mulatta*. (B-D) Scenario for the further division of the *silenus/nigra* lineage (B), the *arctoides/sinica* lineage (C) and *fascicularis/mulatta* lineage into respective species groups and extant species.

Supplementary Figures

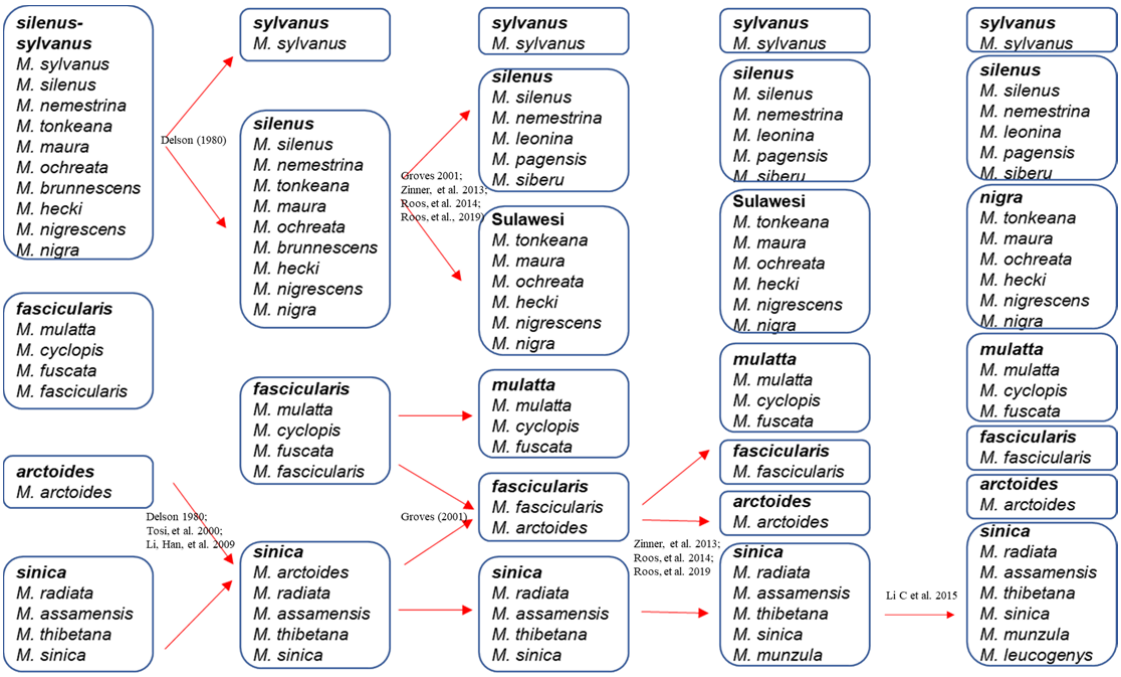


Figure S1. Changes in the classification of extant macaque species after Fooden (1976).

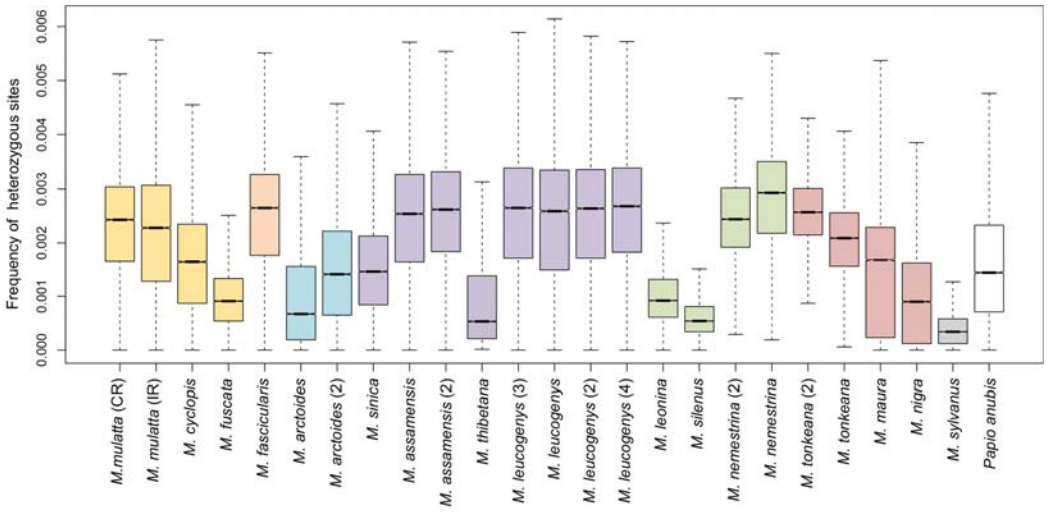


Figure S2. Genome-wide heterozygosity estimated from genomic 100-kbp windows.

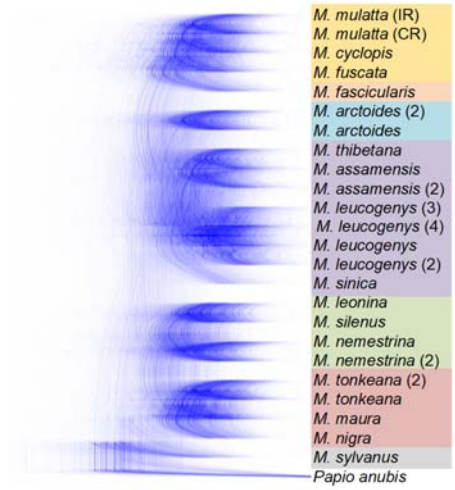


Figure S3. Densitree phylogenetic tree. Light blue lines represent phylogenetic trees produced from 6351 genome fragments.

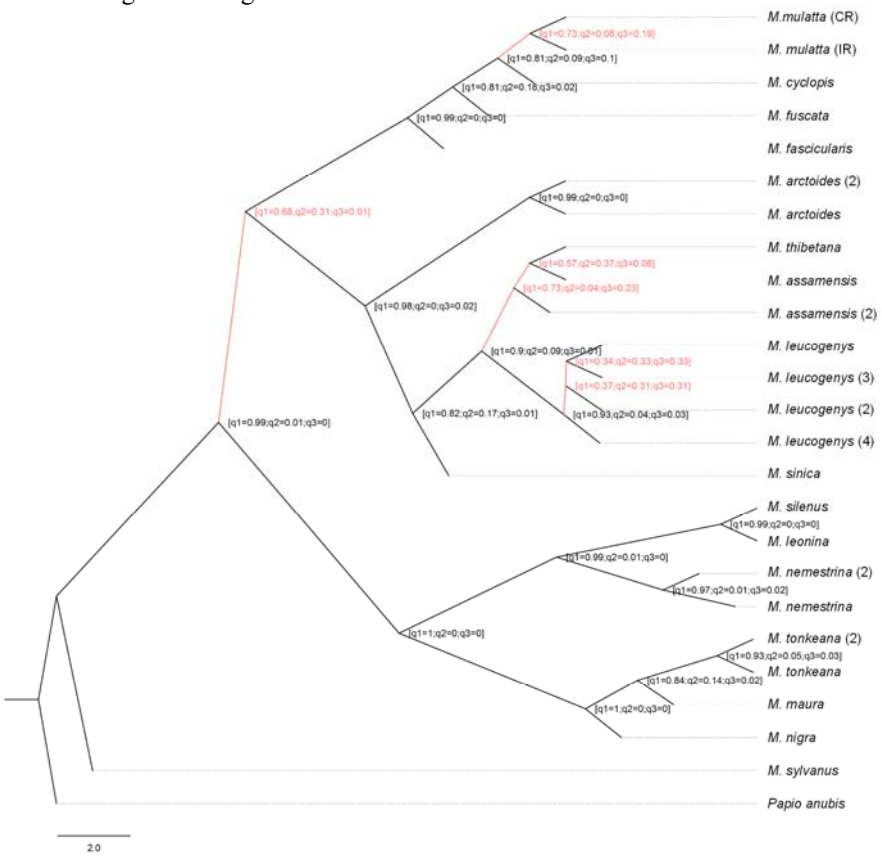


Figure S4. MSC-based species trees generated by ASTRAL using genome fragments. The tree was rooted with *Papio anubis*. ASTRAL quartet-score for all branches are shown and branches labelled in Figure 1B are showed in red.

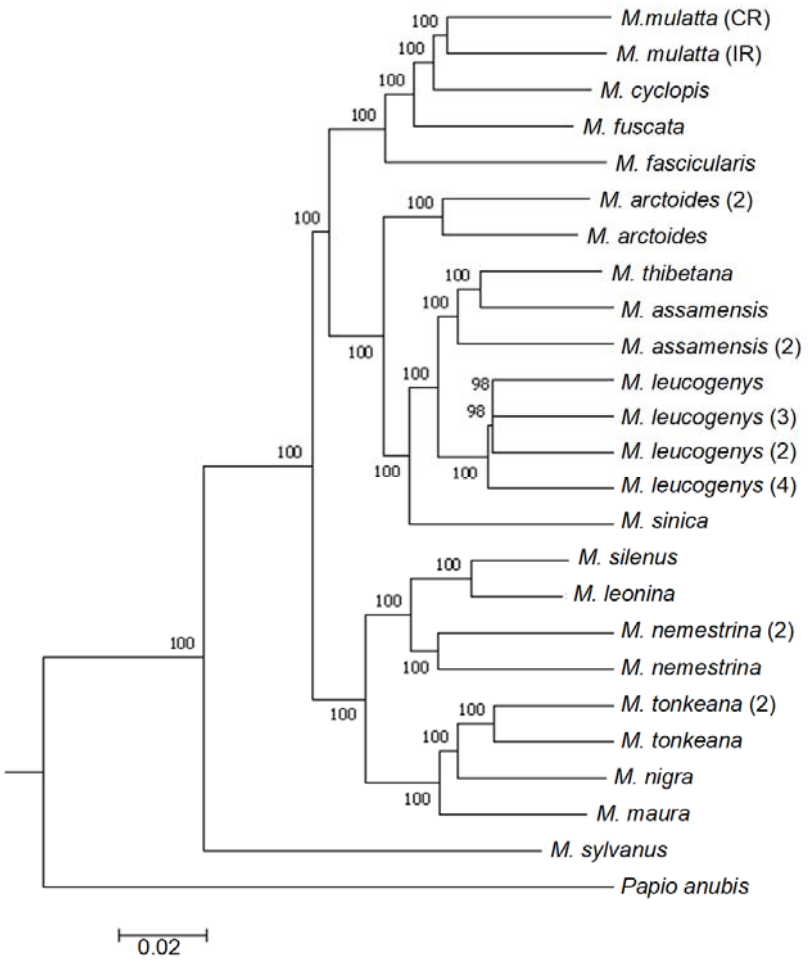


Figure S5. NJ tree based on Treebest using autosomal SNVs. Bootstrap values are given at nodes.

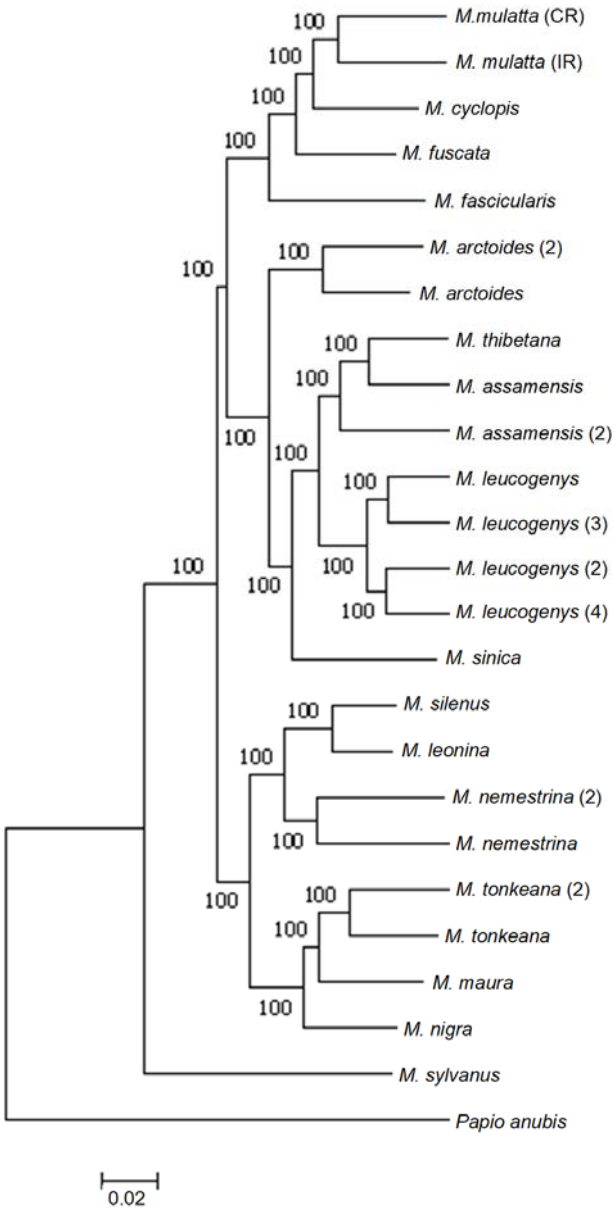


Figure S6. ML tree based on raxml using autosomal SNVs. Bootstrap values are given at nodes.

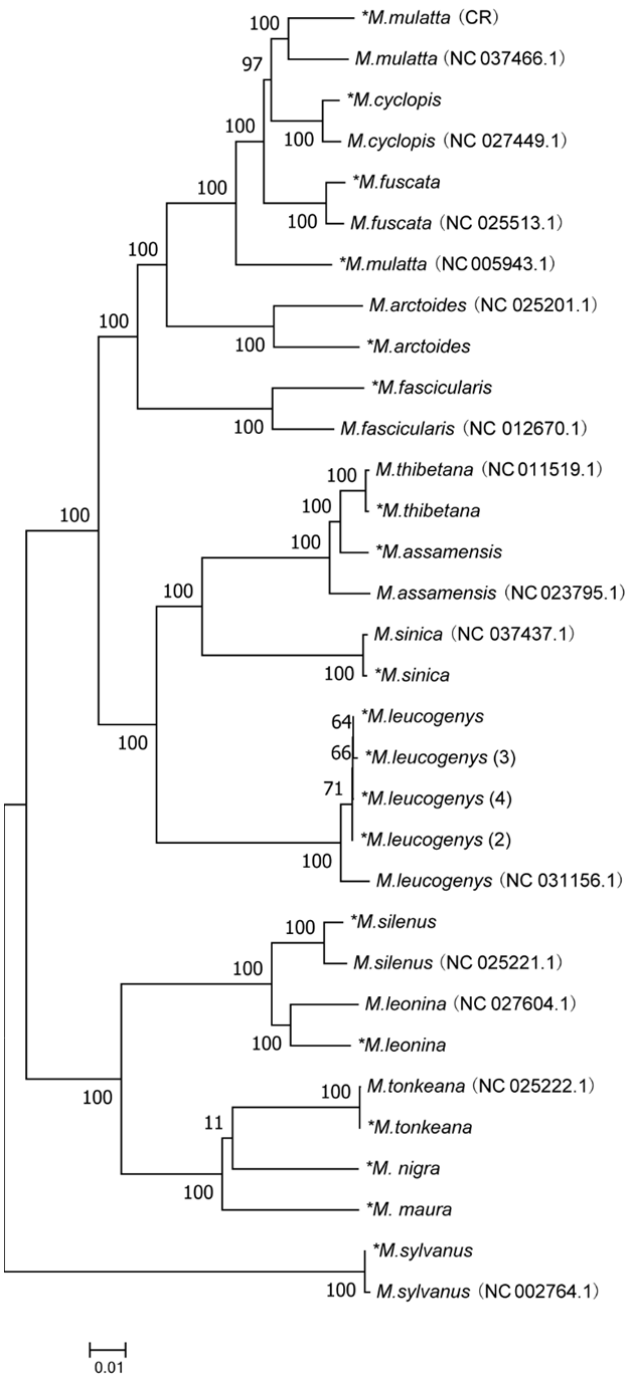


Figure S7. ML tree based on RAxML using mitogenomes. Newly generated sequences are marked with an asterisk. Accession numbers of published sequences are given in parentheses. Bootstrap values are given at nodes.

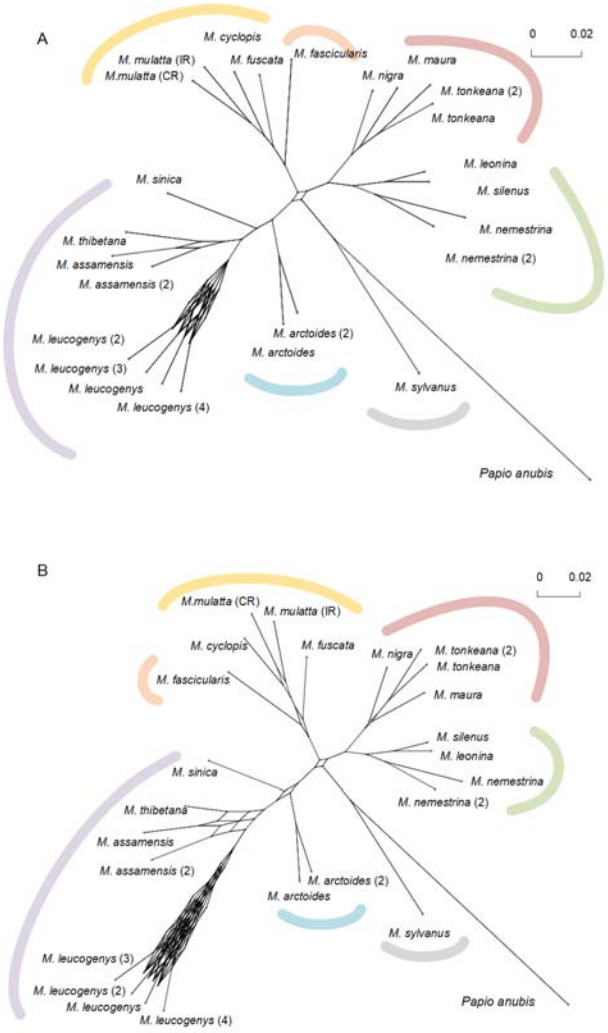


Figure S8. Consensus networks for macaques from 6,351 gene trees (200-kbp genome fragment) at different minimum thresholds (A: minimum threshold=20%; B: minimum threshold=10%) of gene trees to form an edge.

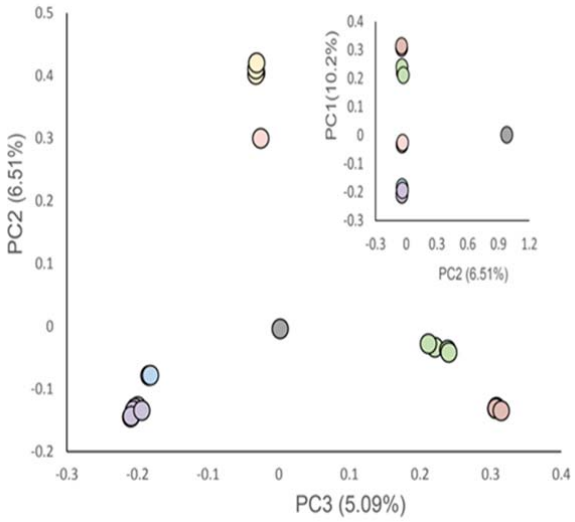


Figure S9. Plots of principal components 1 - 3 from PCA analysis of 24 macaques using autosomal SNVs.

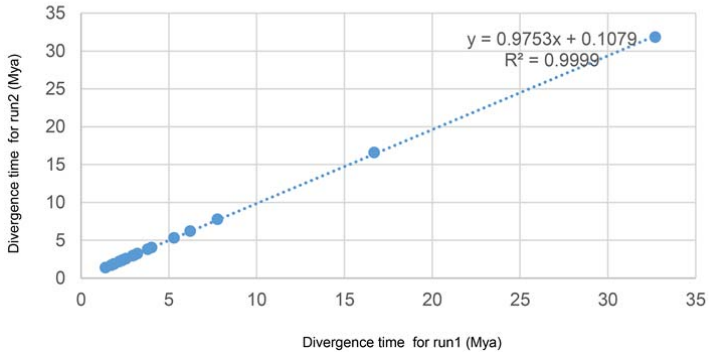


Figure S10. Convergence of the distribution for MCMCtree results.

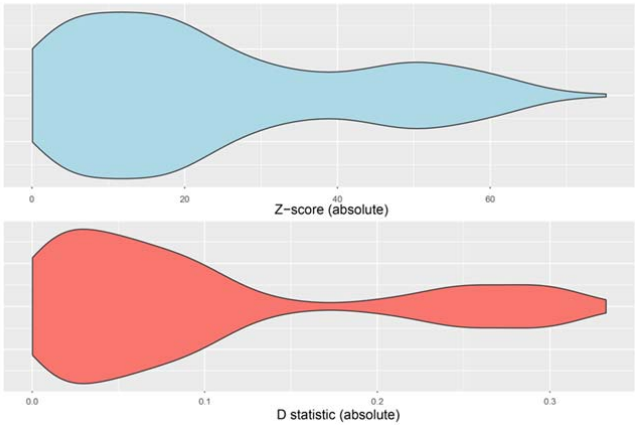


Figure S11. *D* statistic and Z values for *D* statistic for all triplets.

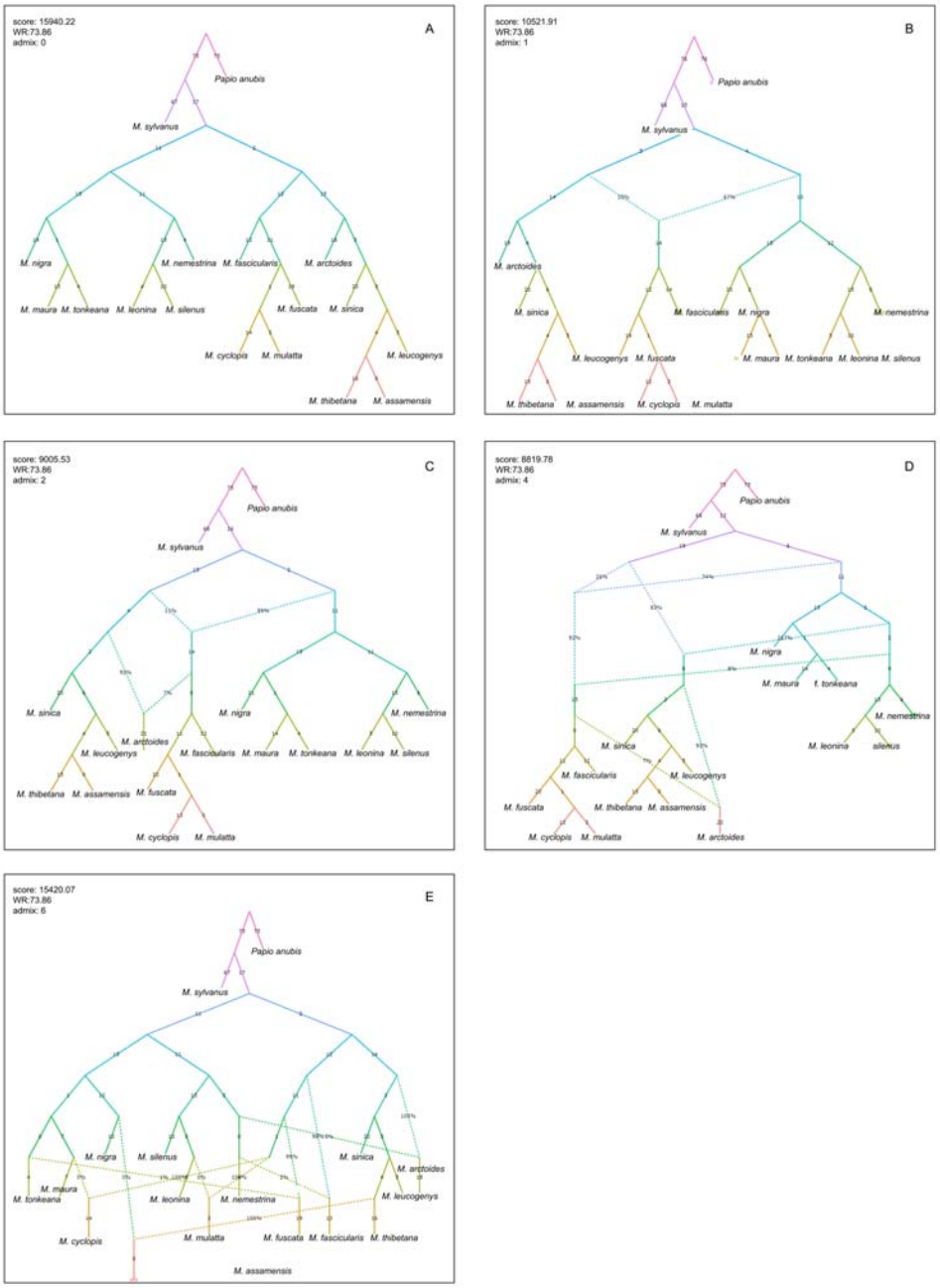


Figure S12. Qpgraph model with admixture fractions. Qpgraph is an admixture graph estimated using all pairwise *f* statistics between samples. Estimated genetic drift is shown along the solid lines, and estimated mixture proportions are given along the dotted lines.

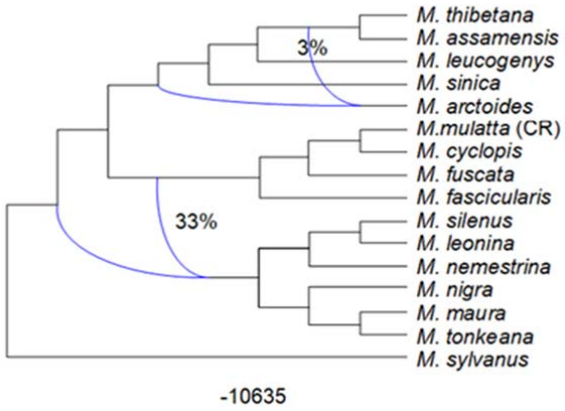


Figure S13. Rooted networks for the *Macaca* phylogeny with same reticulations with Fig. 3C inferred from PhyloNet. Reticulations are shown as blue arrows with inheritance probability denoted. Log-likelihood scores are shown below the networks. Notably, inheritance probability around 33% between *silenus*/*nigra* lineages and *fascicularis*/*mulatta* lineages resembles the distribution of quartet scores and the phylogenetic signals from genome fragments (Fig. 1).

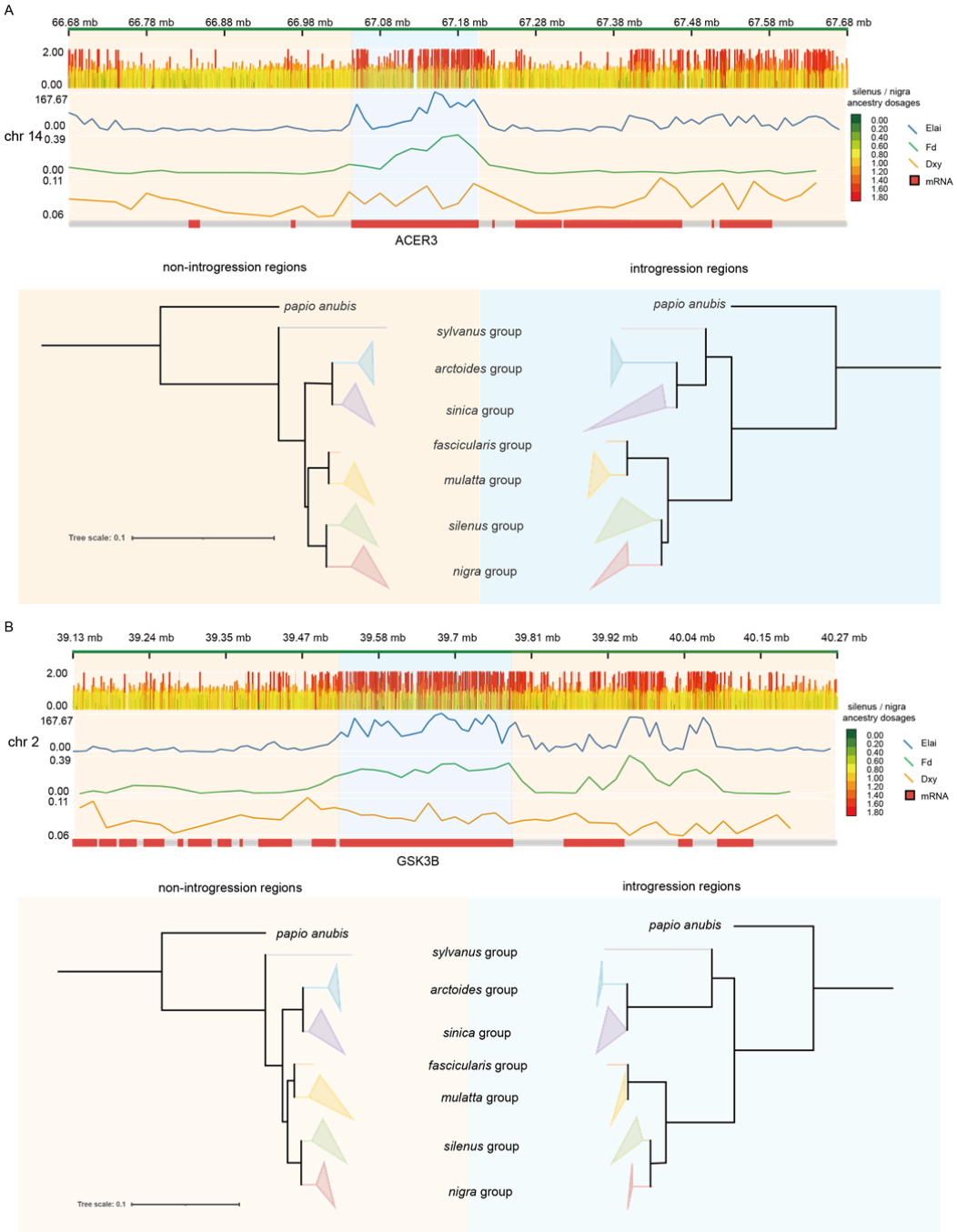


Figure S14. The schematic for inferred of local ancestry for *fascicularis/mulatta* lineage. First line showed 1 Mb regions which contains introgressed windows. The second part, which is a heatmap show the ancestry dosages for r *fascicularis/mulatta* source from *silenus/nigra* lineages. Sites with a proportion of source population (*silenus/nigra* lineage) ancestry greater than 1.5 were defined as introgression sites. The third line show the average value of a proportion of source population (*silenus/nigra* lineage) ancestry in 10 kb windows. The fourth and fifth lines show the *Fd* and *Dxy* and the last part showed the genes in this region. The phylogenetic trees at the bottom of the figure showed the topologies based on the introgression windows and the non-introgression regions

within this 1 Mb regions.

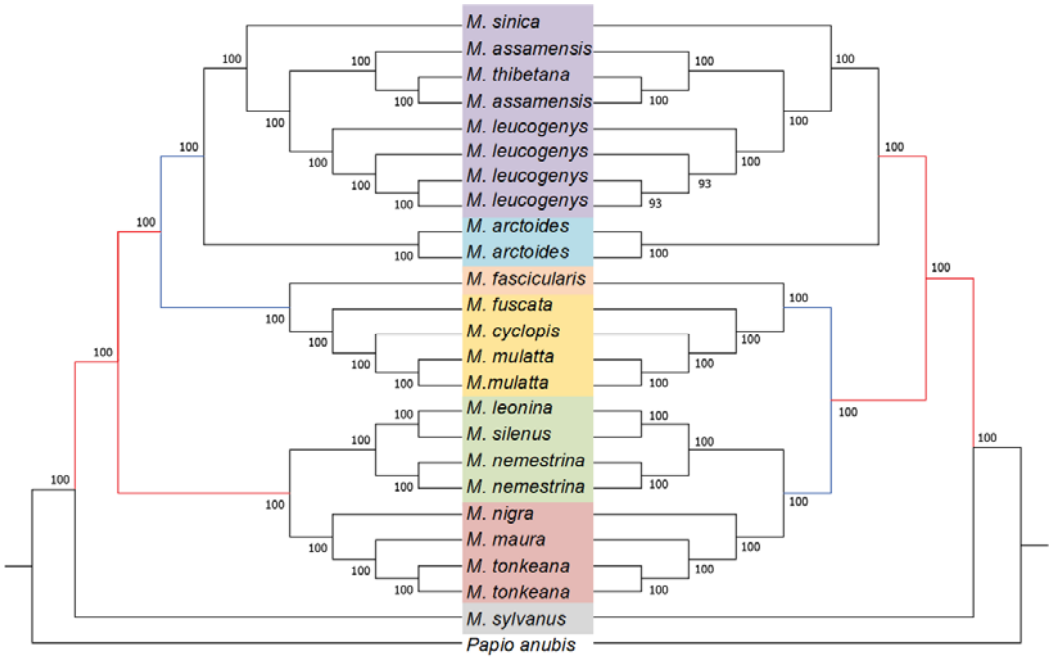


Figure S15. ML tree based autosomal SNVs located in introgressed windows and ML tree based on autosomal SNVs except those SNVs located in introgressed windows between the ancestors of *fascicularis/mulatta* and *nigra/silenus*. Bootstrap values are given at nodes.

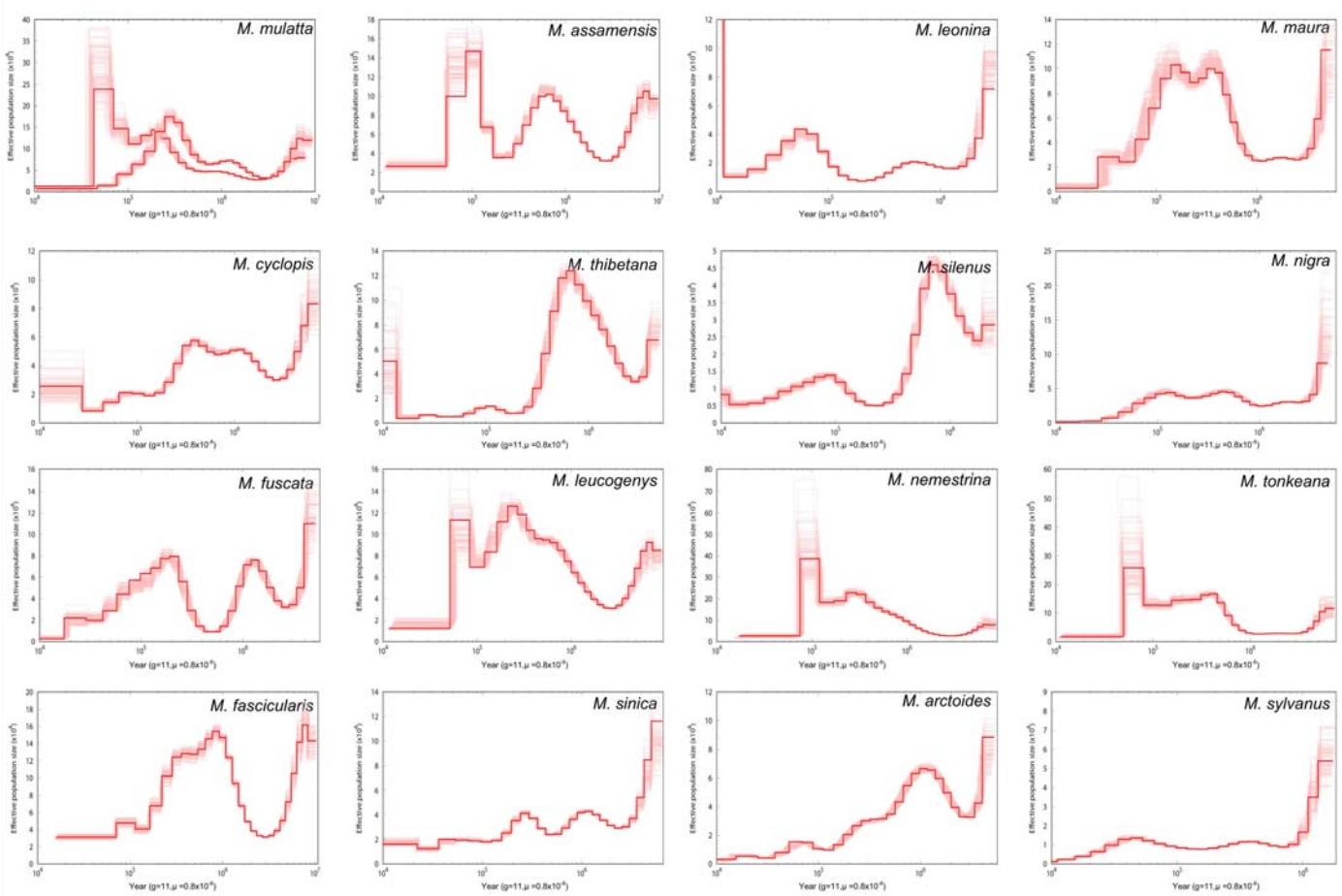


Figure S16. Demographic histories for each individual macaque genome with 100 bootstrap replicates. Each panel shows estimated ancestral effective population size calculated by PSMC. Bootstrap replicates are shown in light red. Sample names are given above each panel.