



Genomic Evidence for the Complex Evolutionary History of Macaques (Genus *Macaca*)

Zhenxin Fan¹ · Rusong Zhang¹ · Anbo Zhou² · Jody Hey³ · Yang Song¹ · Naoki Osada⁴ · Yuzuru Hamada⁵ · Bisong Yue¹ · Jinchuan Xing² · Jing Li¹

Received: 16 November 2023 / Accepted: 20 March 2024 / Published online: 18 April 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

The genus *Macaca* is widely distributed, occupies a variety of habitats, shows diverse phenotypic characteristics, and is one of the best-studied genera of nonhuman primates. Here, we reported five re-sequencing *Macaca* genomes, including one *M. cyclopis*, one *M. fuscata*, one *M. thibetana*, one *M. silenus*, and one *M. sylvanus*. Together with published genomes of other macaque species, we combined 20 genome sequences of 10 macaque species to investigate the gene introgression and genetic differences among the species. The network analysis of the SNV-fragment trees indicates a reticular phylogeny of macaque species. Combining the results from various analytical methods, we identified extensive ancient introgression events among macaque species. The multiple introgression signals between different species groups were also observed, such as between *fascicularis* group species and *silenus* group species. However, gene flow signals between *fascicularis* and *sinica* group were not as strong as those between *fascicularis* group and *silenus* group. On the other hand, the unidirectional gene flow in *M. arctoides* probably occurred between the progenitor of *M. arctoides* and the common ancestor of *fascicularis* group. Our study also shows that the genetic backgrounds and genetic diversity of different macaques vary dramatically among species, even among populations of the same species. In conclusion, using whole genome sequences and multiple methods, we have studied the evolutionary history of the genus *Macaca* and provided evidence for extensive introgression among the species.

Keywords *Macaca* · Genome · Genetic diversity · Evolutionary history · Hybridization

Handling editor: **Liang Liu**.

Zhenxin Fan and Rusong Zhang have contributed equally to this work.

✉ Jing Li
ljtjf@126.com

- ¹ Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, Sichuan, People's Republic of China
- ² Department of Genetics, Rutgers, the State University of New Jersey, Piscataway, NJ 08854, USA
- ³ Department of Biology, Center for Computational Genetics and Genomics, Temple University, Philadelphia, PA, USA
- ⁴ Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido 060-0814, Japan
- ⁵ National Primate Research Center of Thailand, Chulalongkorn University, Bangkok, Thailand

Introduction

The genus *Macaca*, which diverged from other primates in northern Africa during the late Miocene from 7 to 8 million years ago (mya) (Delson 1980), is one of the most successful primate radiations. The invasion of *Macaca* in Eurasia occurred about 5.5 mya, followed by the splitting of several phyletic lineages in Asia. Currently, only *M. sylvanus* of North Africa serves as the genus' sole representative in Africa (Fooden 1979).

The extant macaque species were divided into four distinct species groups: *sylvanus*, *silenus*, *sinica*, and *fascicularis*, based on evidence from morphology (Delson 1980) and molecular studies (Hoelzer and Melnick 1996). However, other studies suggested that they should be classified into seven species groups, in which three groups were separated, respectively, from the *sinica*, *fascicularis*, and *silenus* to create the *arctoides* group, *mulatta* group, and *nigra* group (Zinner et al. 2013; Roos et al. 2014). In particular, *M. arctoides* showed the most uncertain phylogenetic relationship to other

macaques, either grouped to *sinica* group, or grouped to *fascicularis/mulatta* group, or formed its own *arctoides* group (Li et al. 2009; Jiang et al. 2016; Fan et al. 2017, 2018; Roos et al. 2019). The different groupings may reflect the complex evolutionary history of macaques. The speciation and radiation of the Asian lineage occurred about 3 mya, and then was further influenced by natural events, such as climatic and eustatic changes during the Late Pliocene and Pleistocene (Abbott et al. 2013). The significant changes of their habitat disturbance during that period could potentially promote the hybridization or secondary contact among different macaque lineages (Eudey 1979; Delson 1980; Fan et al. 2018). In fact, hybridizations have been reported in various macaques, such as in *M. mulatta*, *M. fascicularis* (Fooden 1995; Hamada et al. 2006; Yan et al. 2011), *M. nemestrina* (Ziegler et al. 2007; Vanderpool et al. 2020), and various *nigra* group species on the island of Sulawesi (Ciani et al. 1989; Evans et al. 2001, 2003). Some interspecies hybridization occurred during secondary contact after a period of isolation, such as the hybridization between *M. fascicularis* and *M. mulatta* (Ito et al. 2020). Additionally, ancient hybridizations that occurred between different species groups were also reported (e.g., *sinica* and *fascicularis* group) (Tosi et al. 2000, 2003a, b; Tosi et al. 2003a, b). Using complete genome sequence data of one *M. thibetana* and one *M. assamensis*, Fan et al. (2014) detected ancient hybridization between *M. thibetana* and *M. mulatta lasiota*. Recently, based on genome re-sequencing data from nine macaque species, strong gene flow signals were detected between the *fascicularis* group and *silenus* group (Song et al. 2021). Additionally, Zhang et al. (2014) revealed evidence of ancient hybridization between the *sinica* and *silenus* groups, and suggested that hybridization, rather than introgression, is the primary factor contributing to the complex evolutionary history of the *Macaca* genus. Different viewpoints have been proposed by Tan et al. (2023) and Rivas-González et al. (2023), who argue that incomplete lineage sorting (ILS) plays a significant role in the evolution of macaques. These studies introduce the notion that shared ancestral polymorphisms and the stochastic sorting of genetic variation during speciation events contribute to the observed complexity in macaque phylogenetics. These findings suggest the evolutionary history of macaques is very complex, especially involving hybridization and secondary contact.

In this study, we sequenced the whole genome of one *M. cyclopis*, one *M. fuscata*, one *M. thibetana*, one *M. silenus*, and one *M. sylvanus* at high coverage using the Illumina HiSeq X Ten platform. Combined with published genomes, we assembled a dataset with 20 macaque genomes to (1) investigate the genetic differences among macaque species; (2) confirm their recent demographic decline; and (3) assess introgression between different macaque species.

Materials and Methods

Samples and Sequencing

Genome re-sequencing was performed for five macaques. We sequenced the genome of one 20 year-old female *M. cyclopis* (TW), one 12 year-old female *M. fuscata* (JM), one female *M. thibetana* (TM, age unknown), as well as one *M. silenus* (LTM) and one *M. sylvanus* (BBM) for which the sex and age information unknown. Notably, all samples were collected from the wild. Genomic DNA was extracted from whole blood using the standard phenol–chloroform method (Sambrook and Russel 2001). Paired-end libraries with insert sizes of 300–500 bp were generated for each sample. Library preparation and all sequencing runs were performed according to the manufacturer's protocols. All samples were sequenced using Illumina HiSeq X Ten. The *M. cyclopis* and *M. fuscata* were sequenced at Novogene (Beijing, China), the *M. silenus* and *M. sylvanus* were sequenced in New York Genome Center (New York, USA), and the *M. thibetana* was sequenced in Biomarker Technologies Corporation (Beijing, China). The clean reads of the above macaques have been deposited in the NCBI Short Read Archive (SRR11921216-SRR11921219, SRR11927939-SRR11927943, SRR11927944-SRR11927948).

Combining the five genomes with published macaque genomes, we generated a data set with 20 macaque genome sequences (Table S1). These macaques covered all four species groups within the genus *Macaca*, including eleven individuals in *fascicularis* group (three *M. mulatta lasiota* (CR), one *M. cyclopis* (TW), two *M. fuscata* (JM), and five *M. fascicularis* (CE)), five individuals in *sinica* group (two *M. thibetana* (TM), two *M. arctoides* (SM), and one *M. assamensis* (XH)), three samples in *silenus* group (two *M. nemestrina* (PM) and one *M. silenus* (LTM)), and *M. sylvanus* (BBM) in *sylvanus* group, which is the only species within *sylvanus* group (Table S1). One Guinea baboon (*Papio papio*) (NCBI accession No. SRX652597 and SRX652598) was used as an outgroup.

Re-sequencing Reads Mapping, Genotyping, and Post-genotype Filters

The paired-end short reads were aligned to the *M. mulatta mulatta* reference genome using Bowtie2 (Langmead and Salzberg 2012) under the local alignment algorithm with the very sensitive model and proper insert sizes, while default options were used for other parameters. Next, Picard and GATK toolsets (DePristo et al. 2011) were used to process the alignments to single nucleotide

variation (SNV) calls in Variant Call Format (VCF). The pipeline is the same as used in our previous studies (Fan et al. 2014; Freedman et al. 2014). After the SNV calling, we performed several conservative data quality filters to control the data quality. Genome filters (GF) and sample filters (SF) described in Fan et al. (2014) were applied to eliminate unreliable SNVs as much as possible, such as SNVs in copy number variants regions, CpG sites, and proximity to Indel.

Phylogenetic, Network, PCA, and ADMIXTURE Analyses

SNVs of 20 macaque individuals were used for the analyses. To eliminate the effects of SNVs that are in linkage disequilibrium, SNVs that have a pairwise $r^2 > 0.2$ within 50 SNV windows were first filtered out using PLINK (Purcell et al. 2007). After pruning, 23,055,023 out of 95,801,537 SNVs remained.

For phylogenetic analysis, modeltest-ng (<https://github.com/ddarriba/modeltest>, version v0.1.2) was first used to test 24 different models, including six models of nucleotide substitution, in combination with four models of site rate heterogeneity. The nucleotide frequency was also allowed to be estimated from the dataset (parameter: “-h uigf -f ef -s 3”). RAxML (version 8.2.11) (Stamatakis 2014) was then used to perform the phylogenetic analysis with a full Maximum Likelihood (ML) search for the best tree and 1000 fast bootstraps to obtain a confidence level. The nucleotide substitution model was chosen based on the modeltest-ng result (parameter “-x -N 1000 -m GTRCATX -c 1 -V -f a”).

Considering the possible gene flow events between different species, we also built consensus networks with the following method. Sites from every non-overlapping 100 kb of the genome fragment were concatenated separately, which were named as “SNV-fragment.” In total, 26,346 SNV-fragments were obtained for consensus network analysis. For each SNV-fragment, a phylogenetic tree was computed with IQ-TREE (version 1.6.10) (Nguyen et al. 2015; Kalyaanamoorthy et al. 2017; Hoang et al. 2018) with parameters: -st DNA -bb 1000 -m MFP. Consensus networks of the 26,346 SNV-fragment trees were generated using phangorn (v2.5.5) (Schliep 2011) with different proportions.

Principal component analysis (PCA) was performed using the smartpca within the EIGENSOFT package (version 6.14) (Patterson et al. 2006). Genome-wide admixture estimates were obtained using a model-based algorithm implemented in ADMIXTURE (version 1.02) (Alexander et al. 2009).

Gene Flow Analysis

The D statistics were performed between closely related populations to test whether there was gene flow between

different species. The D statistic detected asymmetries in allele sharing between either of two receiving lineages (P1 and P2) and a source lineage (P3), given an outgroup (O) (Durand et al. 2011). For each comparison, the D statistic was calculated in 1 Mb windows along the genome. Only sites that passed GF and SF filters were considered. Following (Durand et al. 2011), the standard error of the statistic was calculated using a jackknife procedure, and a Z -score was obtained by dividing the value of the D statistic by its standard error. Z -scores with absolute value ≥ 3 were considered significant, indicating evidence for gene flow between the P3 and one of the receiving lineages (P1 for negative Z -scores, P2 for positive values). Different macaques were assigned as P1, P2, and P3, and the Guinea baboon was used as an outgroup (O) in all tests. The *M. mulatta lasiota* from Sichuan (CR2) was excluded from these analyses due to its low coverage. Various possible combinations were performed with our own Python script.

Demographic Analysis in Macaques

Phylogenetic and Demographic Inference with IMA3

We used IMA3 to estimate both the macaque phylogeny and the demographic history. IMA3 estimates a rooted ordered phylogenetic topology by integrating over IM models (Hey et al. 2018). Because of limitations on the number of populations that can be analyzed with IMA3, we firstly tested the model with fewer species consisting of five species (Table S2), to estimated phylogenetic relations and demographic history. The migration rate parameters and the prior on population splitting times were set to $U[0,0.2]$ and $U[0,2]$, respectively. Ghost populations were included in all runs.

Based on the results of the test sets, we focused on the history of the eight Asian species. From the aligned genomes of 17 individuals, we sampled 200 non-coding loci, each of them was not closer than 10,000 bp to a reported gene. We also filtered out regions with CpG elements and repeats. To generate sampled regions that do not show evidence of recombination, sequences were phased (Stephens et al. 2001) and subsampled using the 4-gamete criterion (Hudson and Kaplan 1985) as previously described (Hey and Wang 2019).

In order to estimate speciation times and effective population sizes, we required estimates of the mutation rate and the generation time (Hey and Nielsen 2004). For generation time, we used 10 years, consistent with estimates in the 9 to 11 years range from *M. fuscata* (Koyama et al. 1992; Takahata et al. 1998; Sugiyama et al. 2009). To estimate the mutation rate per unit of time, we obtained the estimated divergence times from a TimeTree database (Kumar et al. 2017) and regressed the observed pairwise

divergence for the 200 sampled loci upon these (Table S3 and Fig. S1). The regression showed a strong linear relationship ($R^2=0.8753$) at a substitution rate of 4.0×10^{-10} mutations/bp/year, which was corresponding to 4.0×10^{-9} mutations/bp/generation, assuming 10 years per generation.

The IMA3 program was run using hyperparameter upper bounds of 5.0 for the genetic drift parameters, and 0.2 for the migration rate parameters. Together these priors allow for population migration rates as high as 1.0. The upper bound on speciation times, scaled by mutation rate, was 2.0. To enhance the mixing of the Markov chain simulation, 420 heated chains were run simultaneously on 140 processors. Following an extensive burnin period, during which the run was monitored to assess convergence, a total of 123,121 phylogenetic topologies were recorded. Table S4 shows the estimated posterior probability distribution for all trees sampled at a frequency of 1% or higher. IMA3 estimates rooted phylogenies with the sequence of internal nodes is ordered in time. For example, a tree in which populations A and B join more recently than do C and D, is distinct from one in which populations C and D join more recently than the junction of A and B. We then ran IMA3 to estimate the demographic history conditional on the rooted ordered topology with the highest estimated posterior probability. The upper bounds on a uniform prior distributions were 5.0 for the genetic drift parameters, 0.2 for the migration rate parameters, and 1.5 for the speciation time terms. Following a burnin, using 420 heated chains on 140 processors, we sampled 29,760 genealogies for each locus. Finally, the IMfig program was used to generate figures of the combined phylogenetic and demographic history using the maximum posterior estimates of model parameters (Hey et al. 2018).

Inference of Population Size Changes Through Time with PSMC

The pairwise sequentially Markovian coalescent (PSMC) (Li & Durbin 2011) method was used to infer demographic history. Briefly, the method uses the distribution of heterozygote sites across the genome and a Hidden Markov Model to reconstruct the history of effective population sizes. The following parameters were used: numbers of iterations = 25, time interval = $1 \times 6 + 58 \times 1$, mutation rate per generation = 4.0×10^{-9} , and generation time = 10. The above settings were the same as in IMA3. To validate the confidence in PSMC findings, 100 bootstrap replicates were run for each genome. To sample a bootstrap replicate, the genome was divided into segments of 5 Mb in length, and the segments were then sampled with replacement to obtain a sequence with approximately the same length as the original genome defined by the “-b” option in the PSMC software.

Genetic Divergence

To estimate the genetic divergence between different macaques, genetic distance was calculated using the genetic distance metric described by (Gronau et al. 2011) at genome-wide with 50 kb non-overlapping windows, and then we made the pairwise comparison between and within different species across the genome. We also calculated the average pairwise differences (PD) based on 51,941 of 50 kb non-overlapping windows.

Results

Whole Genome Data Mapping

This study included the whole genome sequences of 20 macaque individuals and one baboon as an outgroup. Five samples were newly sequenced with the remaining sequences obtained from the previous studies (Fang et al. 2011; Yan et al. 2011; Higashino et al. 2012; Fan et al. 2014, 2018; Zhang et al. 2014; Osada et al. 2015) and NCBI submission. All the genomes had higher than $20 \times$ coverage, except CR2, which was $10.52 \times$ (Table S2). The average coverage of the 20 macaque genomes was $\sim 37 \times$. The number of total useable sites ranged from 1,637,370,536 (CR2) to 2,306,953,441 (CR3), and the only sample that contained less than two billion sites was CR2 (Table S5).

Phylogenetic Tree Across Macaques

All autosomal SNVs from re-sequencing data of 20 individuals were used to construct a genome-wide phylogenetic tree (Fig. 1A). The overall topology supported four major clades of all the macaques with 100% bootstrap support values. The four major clades were corresponding to the four previous described species groups in this genus. The *sinica* group (*M. thibetana*, *M. arctoides*, and *M. assamensis*) and *fascicularis* group (*M. mulatta*, *M. fascicularis*, *M. fuscata*, and *M. cyclopis*) were a sister group to each other, *silenus* group (*M. nemestrina* and *M. silenus*) was diverged from the *sinica* and *fascicularis* groups, whereas *M. sylvanus*, the only species of *sylvanus* group, was located at the most basal clade within macaques. Within *fascicularis* group, three *M. mulatta* shared a close relationship with *M. cyclopis*. A principal component analysis (PCA) without outgroup also divided sampled macaques into four clusters (Fig. 1B). The first principal component (PC1) accounted for 18% of the variance in the dataset, where the *fascicularis* group species were separated from the rest species. PC2, which accounted for 16% of the variance, further separated *sinica* group species from *sylvanus* group and *silenus* group species.

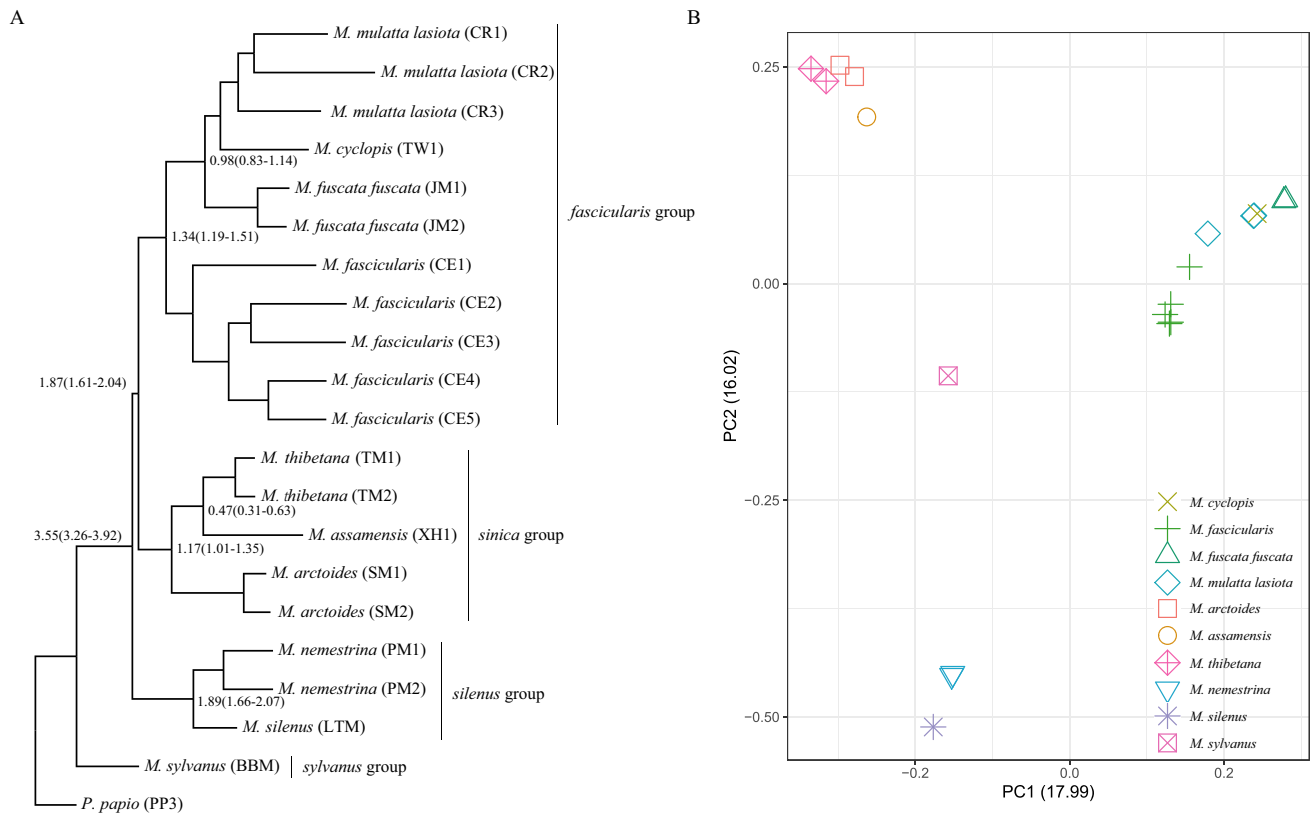


Fig. 1 The phylogeny of *Macaca*. **A** The phylogenetic tree across macaques. The genome-wide phylogenetic tree is based on autosome SNVs. All the branches are supported by 100% bootstrap runs. The

Because reconstructing the history of macaques is both a phylogenetic problem, as well as a population genetic problem because of the likely history of population size changes and gene flow or admixture, we also took a model-based approach to estimating that history. The IMA3 program can estimate population phylogeny, while allowing for population size changes and gene exchange, by integrating over the possible isolation-with-migration models that fall under a user-specified prior (Hey et al. 2018). The top six trees generated from IMA3, all shared the same phylogeny as Fig. 1A (Table S4). With eight species, IMA3 estimated a large and complex model that included seven speciation time estimates (Table S6). The estimated times of population separation were consistent with the Asian radiation of macaques beginning at about 3.5 mya. The most closely related species (*M. assamensis* and *M. thibetana*) were estimated to separate about only half a million years ago, whereas the divergence between *M. silenus* and *M. nemestrina* and the split between *sinica* group and *fuscularis* group were both

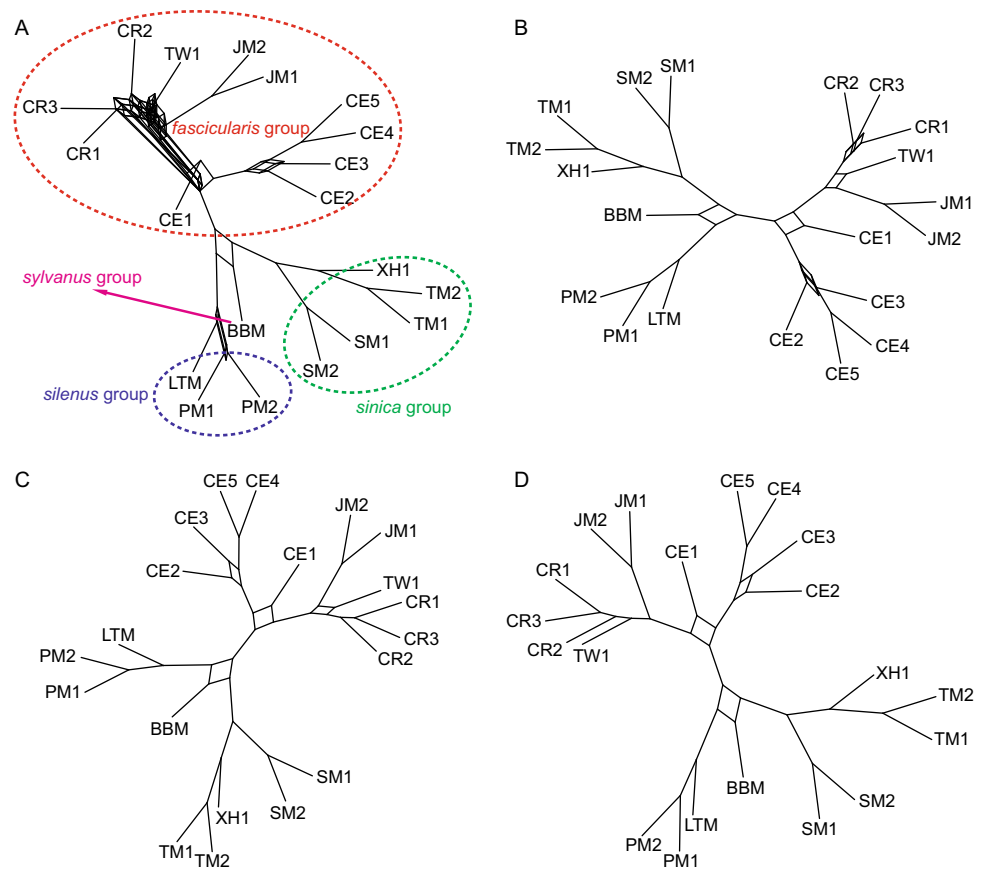
numbers on internal nodes are speciation time estimates by IMA3. **B** The genome-wide PCA results. The results from PC1 to PC2 and the variance explained by each PC are shown

about 1.9 mya. The mainland species *M. mulatta* and the island species *M. fuscata* separated about 0.98 mya.

Reticulate Evolutionary History of Macaques

Because of the complex evolutionary history of macaques, next we applied a network analysis approach to allow more complex phylogenetic relationship modeling among the species. A consensus network analysis of the SNV-fragment trees yielded reticulate structure of connecting alternative branches indicating reticulate phylogenetic of macaques (Fig. 2). There is obvious reticulate structure in the center of the network at different thresholds, indicating the evolutionary relationship of three species groups is unstable, either the *fuscularis* and the *sinica* groups share close relationship, or the *fuscularis* and the *silenus* group together. This suggests potential interspecific gene flow may exist among different species groups. Reticulate structure also occurs between species within *fuscularis* species groups, suggesting that there are phylogenetic conflicts within the species

Fig. 2 Consensus networks for macaques from SNV-fragment trees at different thresholds of trees to form an edge. **A** 11% threshold. **B** 16% threshold. **C** 21% threshold. **D** 26% threshold



groups. Especially for the crab-eating macaques, the CE1 had connections with the rest four *M. fascicularis* and (*M. mulatta*+*M. fuscata*+*M. cyclopis*), suggesting the complex genetic background of CE1.

Introgression in Macaques

The Ima3 analyses and *D* test were performed to investigate gene flow among macaque species. Based on the analysis of the test sets comprising five species, no migration between *sylvanus* group (*M. sylvanus*) and other groups was detected (Fig. S2A and B), suggesting that Asia macaques did not hybridize with African macaques after invading Eurasia. Therefore, the subsequent Ima3 analysis exclude *M. Sylvanus*. Furthermore, migration was also observed between species within the same groups (Fig. S2C and D).

The subsequent Ima3 analyses included 15 population size estimates and 98 separate migration rate estimates (Fig. 3A). Estimated effective population sizes varied widely, from about 14,000 for *M. silenus* to nearly half a million for *M. assamensis*. The *M. fuscata*, *M. arctoides*, and *M. thibetana* also had small effective population sizes (Table S7, Fig. 3A). With respect to gene flow, most of the parameters, including essentially all of those involving ancestral species, revealed estimated posterior densities that

were quite flat, consistent with there being little statistical power to resolve whether or not the estimated rate was different from zero. However, for pairs of sampled species, there is more power, and in several cases, a likelihood ratio test (Nielsen and Wakeley 2001) rejected a rate of zero migration. Figure 2B shows cases of non-zero migration, together with the corresponding estimate of the population migration rate (i.e., twice the product of the effective population size of the receiving species and the migration rate, or $2Nm$). In all cases, these values were less than 0.03, suggesting that gene flow had not had a large effect on the genetic structure of these sampled populations. We identified seven statistically significant gene flow events: (1) from *M. arctoides* to *M. fuscata*; (2) from *M. arctoides* to *M. fascicularis*; (3) from *M. fascicularis* to *M. nemestrina*; (4) from *M. nemestrina* to *M. fascicularis*; (5) from *M. mulatta lasiota* to *M. silenus*; (6) from *M. silenus* to *M. fascicularis*; (7) from *M. silenus* to *M. nemestrina* (Table S8, Fig. 3A).

In addition, *D* tests were performed to assess gene flow events between different macaques (Table S9). The results showed that *M. arctoides* had significant gene flow with three *fascicularis* group species (*M. mulatta lasiota*, *M. fuscata* and *M. cyclopis*). Since the three macaques had a common ancestor, thus the gene flow could have occurred between *M. arctoides* and the common ancestor of these

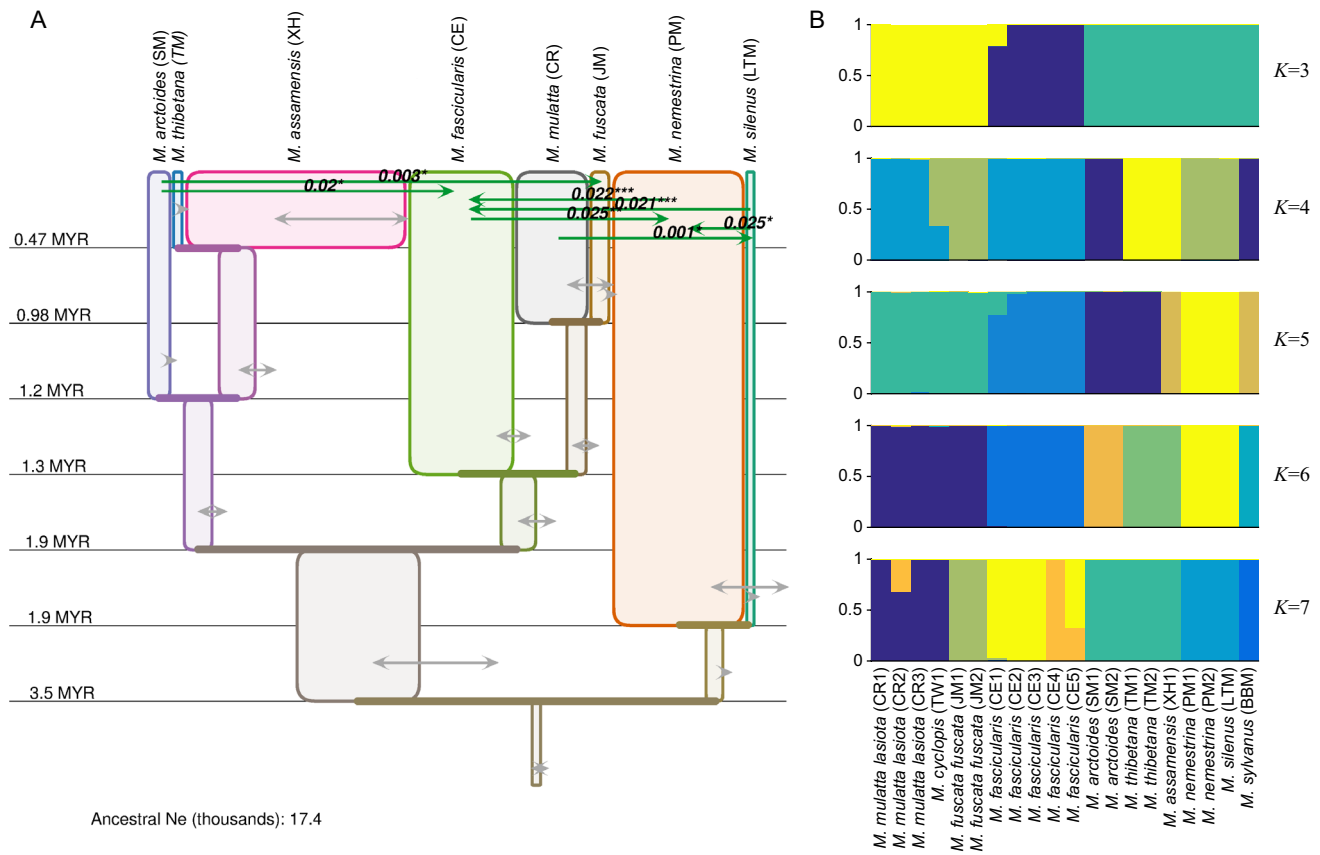


Fig. 3 Introgression within *Macaca*. **A** A representation of an estimated Isolation with Migration model generated by IMA3 and the IMfig program. The phylogeny is depicted as a series of boxes organized hierarchically, with ancestor boxes positioned in-between the corresponding descendants, and the width of boxes proportional to estimated effective population size (N_e). The 95% confidence intervals for N_e values are shown as gray arrows extending to the left and right of the right boundary of each population box. Splitting times are depicted as evenly spaced solid horizontal lines, with text values

three macaques. We also detected significant gene flow between *M. mulatta lasiota* and *M. nemestrina*, and the signal was consistent among different individuals in these species. However, the gene flow between *M. mulatta lasiota* and *M. thibetana* was only observed in one individual pairwise D test. The runs with different individuals of *M. mulatta lasiota* and *M. thibetana* did not detect significant gene flow. For *M. assamensis* and *M. thibetana*, both species had significant gene flow with multiple macaques within *fascicularis* group, indicating the gene flow probably happened between the ancestor of *M. assamensis* and *M. thibetana* and the ancestor of *fascicularis* group species.

Genome-wide admixture estimates were obtained using a model-based algorithm implemented in ADMIXTURE (version 1.02) (Alexander et al. 2009) to assess the

ancestry of each individual from 2 to 8 inferred ancestral populations (K) (Fig. 3B). The likelihood value reached the first peak when $K=3$, although the CV error was still high. The maximum likelihood was achieved at $K=6$. When $K=3$, *sylvanus* group, *silenus* group, and *sinica* group species formed one clade, whereas three *M. mulatta*, two *M. fuscata*, and *M. cyclopis* formed the second clade. Five *M. fascicularis* grouped with each other and formed the last clade. When $K=6$, the three *silenus* group species had their own component, and *M. sylvanus* also formed its own component. Within *sinica* group species, the two *M. thibetana* and *M. assamensis* formed another component (Fig. 3B), which indicates *M. arctoides* had a distinct genetic background.

Demographic Histories of Macaques

The pairwise sequentially Markovian coalescent model (PSMC) was conducted to test the ancestral demographic trajectories of sampled macaques (Fig. 4). Except for *M. sylvanus*, all the macaques exhibited similar demographic trajectories until about 700 thousand years ago (kya). Since then, some of the macaques, even the ones within the same species groups, showed very different trajectories. Within *sinica* group, *M. thibetana* and *M. assamensis* had very similar trajectories, and their effective population size (N_e) was also very similar across the entire history (Fig. 4C). However, two *M. arctoides* began experiencing population decline at ~ 2 mya and kept maintaining lower N_e . Within *fascicularis* group, the two island species, *M. cyclopis* and *M. fuscata*, exhibited very low N_e after 700 kya, whereas *M. mulatta* and *M. fascicularis* maintained relative high N_e after 700 kya (Fig. 4A). Moreover, *M. mulatta* and some of the *M. fascicularis* experienced population growth since ~ 100 kya. However, one *M. fascicularis* individual (CE1), which was from Vietnam, showed different trajectories when compared to the other four *M. fascicularis*. CE1 began the population decline at ~ 100 kya, whereas the other four *M. fascicularis* started population growth at ~ 60 kya (Fig. 4B). While the *sylvanus* group and the *silenus* group, both of which

diverged early in the genus, showed different trajectories from *sinica* group and *fascicularis* group species (Fig. 4). Both *M. silenus* and *M. nemestrina* started the population growth at ~ 700 kya, but then *M. silenus* began the decline at ~ 300 kya, whereas *M. nemestrina* kept the population growth. The most ancient living macaque, *M. sylvanus*, had very different trajectories. It had the highest N_e before 700 kya, and then experienced an extremely strong population decline and remained very low N_e (Fig. 4).

Genetic Divergence of Macaques

To quantify genome-wide heterozygosity, we calculated the number of heterozygous SNVs overall useable sites of all the genomes (Fig. S3). Within macaques, *M. sylvanus* had the lowest autosomal heterozygosity (0.000399), and *M. silenus* also had very low heterozygosity (0.000558). Within *fascicularis* group, two *M. fuscata* exhibited low heterozygosity (0.001148 and 0.001007). Two *M. thibetana* had the lowest values within *sinica* group (0.000898 and 0.000666), whereas their close relative *M. assamensis* had high heterozygosity (0.002723). *Macaca mulatta* and *M. fascicularis*, which both had large populations, had high heterozygosity (Fig. S3).

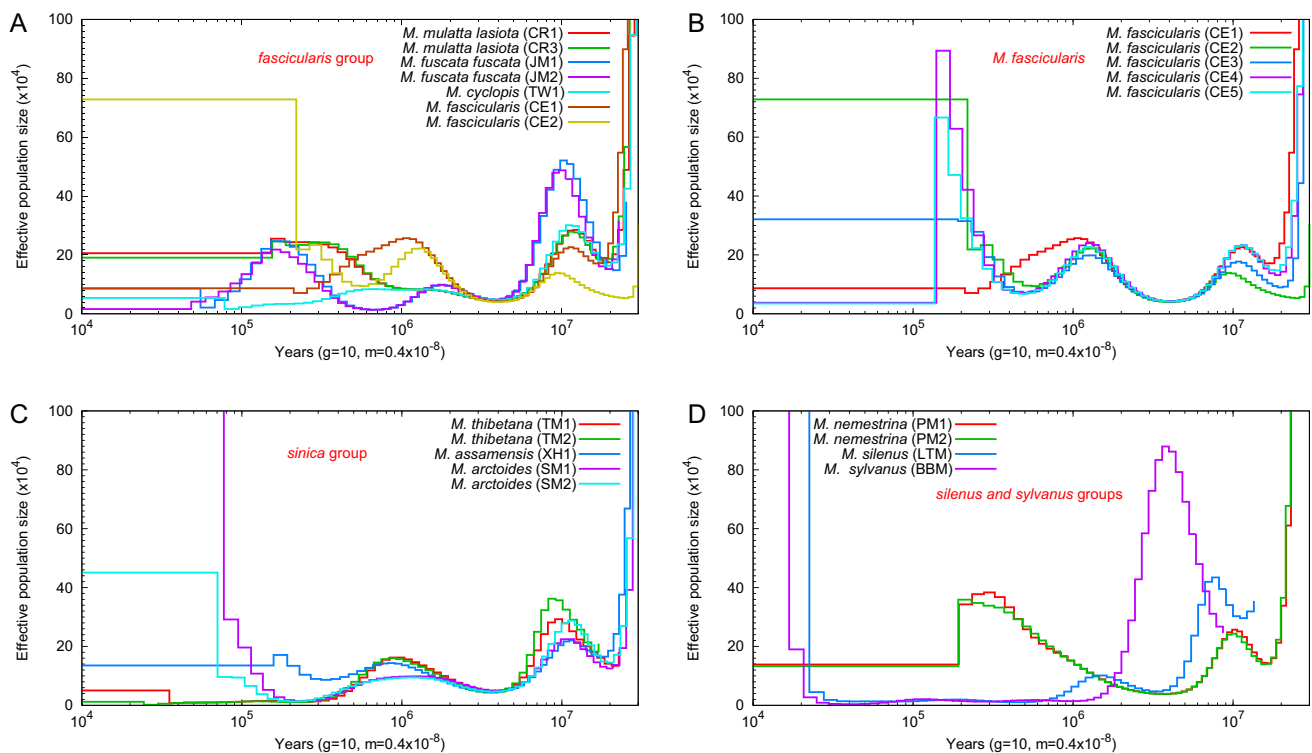


Fig. 4 Historical changes in effective population sizes of macaques. Reconstruction of historical patterns of effective population size for macaque genomes using the PSMC method. **A** *fascicularis* group

species. **B** five *M. fascicularis*. **C** *sinica* group species. **D** *sylvanus* group and *silenus* group species

To estimate the genetic divergence between samples, we calculated the genetic distance at whole genome level (Table S10). Overall, the genetic distances between different macaques within the same species groups were smaller than that between samples among different species groups. Within *fascicularis* group, genetic distances of *M. cyclopis* to *M. mulatta* (0.0565–0.0571) were lower than that of *M. cyclopis* to *M. fuscata* (0.0631–0.0635). Within *sinica* group, two *M. thibetana* had smaller genetic distances to *M. assamensis* (0.0625 and 0.0627) than that of *M. thibetana* to *M. arctoides* (0.818–0.834). *M. sylvanus* was the most ancient living macaque species, and we observed that it had the largest genetic distance between all the other macaques (0.1449–0.17). This study allowed to estimate the genetic distances within species given that some species contained more than one individual (Table S10). The genetic distances within two *M. fuscata* (0.0232), two *M. thibetana* (0.0194), two *M. arctoides* (0.0227), three *M. mulatta* (0.0484, 0.0499, and 0.05), and two *M. nemestrina* (0.0558) were smaller than distances within five *M. fascicularis* were very large (0.0433–0.0781).

Next, we calculated the pairwise genetic distance between different macaques in the 50 kb non-overlapping window (Figs. S4–S6). We also exhibited the average pairwise differences (PD) based on the above windows (Table S11). In general, the results were consistent with the overall genetic distance (Table S10), but provided details of the genetic distances among these macaques. The average PD showed that the genetic differences within *M. fascicularis* were larger than that within *M. mulatta*, *M. fuscata*, *M. thibetana*, and *M. nemestrina*. Moreover, *M. sylvanus* had very large genetic differences between all the rest macaques.

Discussion

Phylogeny of Genus *Macaca*

In this study, we sequenced the whole genome of 5 macaque species and analyzed genome sequences of 20 individuals from 10 species that covered all species groups across the genus. Compared to previous and recent study (Zhang et al. 2014; Song et al. 2021), this study included the most macaque individuals to date. The phylogenetic topology based on genome-wide SNVs showed that the ten species could be grouped into four well-supported clades, which was consistent with the four distinct species groups: *sylvanus*, *silenus*, *sinica*, and *fascicularis* (Zinner et al. 2013; Roos et al. 2014). Although species from *Sulawesi* macaques (*nigra* group) did not included in this study, our previous studies demonstrated that *Sulawesi* macaques (*M. nigra* and *M. tonkeana*) separated from the *silenus* group to form a sister group (Song et al. 2021). As the only living

macaque that is not distributed in Asia (Fooden 1979), *M. sylvanus* located at the most basal clade within macaques and the Asian species all grouped together. Nevertheless, some studies have advocated for the subdivision of the *Macaca* genus into seven species groups (Roos et al. 2019; Tan et al. 2023). This involves segregating *M. mulatta*, *M. fuscata*, and *M. cyclopis* from the *fascicularis* species group, thereby establishing the *mulatta* species group. Similarly, *M. arctoides* is isolated from the *sinica* species group, resulting in the formation of the *arctoides* species group. Furthermore, Sulawesi macaques are delineated from the *silenus* species group, constituting the *nigra* species group. This nuanced classification is refinement of above classification, providing a more detailed insight into the taxonomic relationships within the *Macaca* genus.

In addition to the phylogenetic relationship, our IMA3 analysis provides estimates on the divergence time among the branches. Current evidence suggests that macaques diverged from other primates in northern Africa during the later Miocene from 7 to 8 mya, and then they invaded Eurasia about 5.5 mya and split into phyletic lineages in Asia (Delson 1980; Roos et al. 2019). Our IMA3 analysis estimated that the times of separation within the Asian macaques beginning about 3.5 mya (95% HPD was 3.25–3.92 mya), which was consistent with the reports about the speciation and radiation of the Asian lineage occurred at about 3 mya (Fan et al. 2018; Roos et al. 2019). Then, the split between *sinica* group and *fascicularis* group was estimated at about 1.9 mya, and the divergence between *M. arctoides* and (*M. thibetana* + *M. assamensis*) was about 1.2 mya. These results indicated speciation of the genus *Macaca* was relative recent.

Complex Admixture History of Macaques

Natural hybridizations were reported in almost all the major evolutionary clades of primates (Cortes-Ortiz et al. 2007; Ackermann and Bishop 2010; Reich et al. 2010; Zinner et al. 2009, 2011; Tung and Barreiro 2017). In macaques, some characters, such as morphological characteristics, genital structure, and sexual behavior, are significantly different among species, but the chromosome karyotypes of macaques are almost the same, and hybridization between different macaques has been previously reported (Fooden 1995; Tosi et al. 2000; Tosi et al. 2003a, b; Yan et al. 2011; Fan et al. 2014, 2018; Hamada et al. 2006, 2016; Jiang et al. 2016; Evans et al. 2017; Matsudaira et al. 2018). Furthermore, the hybrid offspring of some species is fertile (Ciani et al. 1989; Yang and Shi 1994; Hamada et al. 2012; Bunlungsup et al. 2017; Evans et al. 2017).

Most previous genome-wide studies on ancient gene introgression among macaques used the popular method *D* test (also known as the “ABBA-BABA” test). However,

D test only considered biallelic sites and assumes that all ABBA or BABA sites arise due to either incomplete lineage sorting or introgression (Patterson et al. 2012). With additional divergence time, this assumption is likely to be violated due to convergent or multiple mutations at a single site (Edelman et al. 2019). To validate previous hybridization and detect unknown gene flows between macaques, we included more genomes and used multiple analytical methods such as consensus networks and IMA3. We have detected extensive admixtures between different macaques, some of which were new findings which were discussed below. We exclude the *M. sylvanus* in the analyses of the final IMA3 and *D* test because it is the only macaque species distributed out of Asia. Natural hybridization in *M. sylvanus* was not observed and reported before. Several test runs of IMA3 with the *M. sylvanus* also did not detect gene flow between the *M. sylvanus* and other species, which suggested that the *M. sylvanus* separated early and had little or no gene flow with the non-African macaque species.

IMA3 detected gene flow between *M. nemestrina* and *M. fascicularis* in both directions. The observed gene flow signals were probably caused by ancient hybridization events or even between their common ancestors, because the IMA3 also detected significant gene flows between more species in *silenus* group and *fascicularis* group, for instance, from CR (*M. mulatta lasiota*) to LTM (*M. silenus*) and from LTM (*M. silenus*) to CE (*M. fascicularis*). *D* test also found significant gene flow between CR and *M. nemestrina*. The ADMIXTURE analyses ($K=4$) showed *M. silenus*, *M. nemestrina*, *M. cyclopis*, and *M. fuscata* were grouped into one component. Therefore, the multiple signals between different species of the two groups are likely a result of ancestral introgression. Our results are consistent with Vanderpool et al. (2020) and Song et al. (2021) finding high-level gene flow signals between *fascicularis* group and *silenus* group. A recent study based on de novo assembled genomes of macaques concluded that the *fascicularis* group may have originated from an ancient hybridization between the progenitors of the *sinica* group and those of the *Silenus* group (Zhang et al. 2014). With more individuals from different species groups in the genus in our study, it allows to further investigate the hybridization hypothesis. If the hybridization origin hypothesis is true, gene flow signals between multiple species of *fascicularis* group and *sinica* group could be also detected. However, in our dataset, the detected gene flow signals between *fascicularis* and *sinica* group were not as strong as those between *fascicularis* group and *silenus* group (Fig. 3A), yet our results could not preclude the hypothesis hybridization origin of *fascicularis* group due to the occurrence of random genetic drift or incomplete lineage sorting during evolution.

Some previously reported gene flow events were detected or supported by new evidence. The phylogenetic

position of *M. arctoides* was discrepant based on different genetic markers (Tosi et al. 2000; Tosi et al. 2003a, b; Li et al. 2009). Our previous work showed that *M. arctoides* had a nuclear genome related to *sinica* species, and a mitochondrial genome closely related with *M. mulatta*. It suggested a secondary contact between *proto-arctoides* and *proto-mulatta* has resulted in the transfer of *mulatta*-type mitochondria into *proto-arctoides* (Fan et al. 2018). In this study, the results from IMA3 and *D* test showed that *M. arctoides* had significant gene flows with four *fascicularis* group species (*M. mulatta lasiota*, *M. fascicularis*, *M. fuscata*, and *M. cyclopis*), this finding suggested that gene flow occurred between the progenitor of *M. arctoides* and the common ancestor of *fascicularis* group.

Additionally, the ADMIXTURE analyses ($K=3, 5$, and 6) showed the Vietnamese *fascicularis* (CE1) had a mixture of genetic components with *M. mulatta lasiota*, which was congruous with previous results (Yan et al. 2011). The hybridization between *M. fascicularis* and *M. mulatta* has been extensively reported, and the Vietnamese *fascicularis* genome was shaped by introgression after hybridization with the *M. mulatta lasiota* (Stevison and Kohn 2008; Bonhomme et al. 2009; Yan et al. 2011; Hamada et al. 2016).

Divergence and Genetic Diversity

We assessed the genome-wide heterozygosity and genetic diversity among/within macaque species. In this study, we included the only African macaque species *M. sylvanus*. Compared to Asian macaques, *M. sylvanus* had the lowest autosomal heterozygosity. *M. silenus*, *M. thibetana*, and *M. fuscata* also exhibited low heterozygosity (Fig. S3). Species with large population sizes, such as *M. mulatta*, *M. fascicularis*, and *M. nemestrina*, had high heterozygosity. We also noticed that *M. mulatta* and *M. fascicularis* had relatively large genetic diversity within species (Table S10). *M. fascicularis* had the highest genome-wide heterozygosity and genetic diversity, which was probably due to their large distribution with complex evolutionary history in different populations. This study sampled five individuals from three or four populations (the origin of one sample was unclear). Previous studies have suggested that the *M. fascicularis* are divided into four major genetic groups (Smith et al. 2007; Blancher et al. 2008; Osada et al. 2010, 2015). The Vietnamese *M. fascicularis* (CE1) is from the Indochinese population that has significant gene flow with the *M. mulatta lasiota* (Stevison and Kohn 2008, 2009; Bonhomme et al. 2009; Yan et al. 2011), whereas the Malaysian *M. fascicularis* (CE2) is from the Indonesian-Malaysian population that maintains the highest genetic diversity and is thought to be the ancestral population (Delson 1980; Higashino et al. 2012; Osada et al. 2015; Fan et al. 2018). The Mauritian *M. fascicularis* (CE4 and CE5) was

introduced to the island of Mauritius around the sixteenth century, which experienced quick population expansion and extreme population bottleneck (Sussman and Tattersall 2008; Osada et al. 2015). The last major population is the Philippine population, which shows slightly reduced genetic diversity and is probably derived from the Indonesian-Malaysian population (Stevison and Kohn 2008; Bonhomme et al. 2009; Osada et al. 2015). Therefore, our genome-wide investigations confirm that *M. fascicularis* from different populations have different genetic backgrounds and the whole species maintains a high level of genetic diversity.

Our study showed macaque species had distinct population trajectories. IMA3 estimated the effective population sizes of macaque species. Some macaques, such as *M. assamensis*, *M. fascicularis*, *M. mulatta*, and *M. nemestrina*, maintained large effective population sizes, while *M. silenus*, *M. fuscata*, *M. arctoides*, and *M. thibetana* had small effective population sizes (Table S7, Fig. 3A). In addition, PSMC results showed a bottleneck for all macaques in our study occurred at about 0.7 mya. This bottleneck may be caused by the change of climate. During the mid-Pleistocene transition (MPT, 1.2–0.8 mya), the glacial-interglacial climate cycle changed from ~40,000 years to ~120,000 years, and this change led to subsequent longer, colder ice ages with larger continental ice sheets and lower global sea level (Chalk et al. 2017). There was a long ice age in the Quaternary period about 0.8–0.68 mya, which may have caused the bottleneck of all macaques at 0.7 mya. Since then, some macaque species showed very different trajectories. For example, the *M. fascicularis* had different trajectory compared with other *fascicularis* group species. Even within *fascicularis*, the trajectory of CE1 was also different from other *M. fascicularis* populations. In addition, the island species, such as *M. cyclopis* and *M. fuscata*, exhibited very low N_e after 700 kya. The *sylvanus* group and *silenus* group species also had different trajectories compared with *sinica* group and *fascicularis* group species (Fig. 4). The observed differences in the demographic histories of different macaque species reflected that they experienced complicated environment and climate changes, gene flow events, and habitat loss (Fan et al. 2014, 2018).

Conclusion

In conclusion, taking advantage of whole genome sequences of multiple species of the genus *Macaca*, the present study shows that genetic background varies among species, even among populations of the same species. Combining different genome analysis methods, we detected extensive gene flow among macaques and validated previously reported hybridizations. In particular, the gene flow between different species of *fascicularis* group and *silenus*

group is likely a result of ancestral introgression. Although we applied different methods to detect admixture and reconstruct the demography, the current analyses could not generate a unified result for the evolutionary history of different macaques, and the introgression patterns might be more complicated than we observed. Therefore, this study highlighted that the admixture has greatly shaped the evolutionary history of the genus *Macaca*.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00239-024-10166-z>.

Acknowledgements This work was supported by the National Natural Science Foundation of China (grant numbers 31770415 and 32171607) and US National Science Foundation (NSF1564659). Computing resources for IMA3 analyses were supported by the National Institutes of Health (grant number S10OD020095), National Science Foundation (grant number 1625061), and US Army Research Laboratory contract (W911NF-16-2-0189). We would like to thank Dr. Zhanlong He from the Chinese Academy of Medical Sciences (Kunming), Mrs Song Wang from Nanning Zoo, and professor Don J. Melnick from Columbia University for their kindly providing macaque samples.

Author Contributions ZF, BY, and JL contributed to the design of this research. NO, and YH collected the samples. ZF, RZ, YS, AZ, JH, JW, ML, and JX contributed to data analysis. ZF, RZ, YS, AZ, JH, JW, JX, and JL wrote the manuscript. All authors read and approved the final manuscript.

Funding Funding was provided by National Natural Science Foundation of China (Grant Nos. 32171607, 32371696), National Science Foundation (Grant Nos. NSF1564659, 1625061), Foundation for the National Institutes of Health (Grant No. S10OD020095), Army Research Laboratory (Grant No. W911NF-16-2-0189).

Data Availability The raw sequencing reads for de novo sequencing have been submitted to NCBI Short Read Archive under accession number PRJNA779036. The re-sequencing data have been deposited in the NCBI Short Read Archive (SRR11921216-SRR11921219, SRR11927939-SRR11927943, SRR11927944-SRR11927948).

Declarations

Competing interests The authors declare that they have no competing interests.

References

- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Most M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM, Väinölä R, Wolf JBW, Zinner D (2013) Hybridization and speciation. *J Evol Biol* 26(2):229–246
- Ackermann RR, Bishop JM (2010) Morphological and molecular evidence reveals recent hybridization between gorilla taxa. *Evolution* 64(10):271–290

- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664
- Blancher A, Bonhomme M, Crouau-Roy B, Terao K, Kitano T, Saitou N (2008) Mitochondrial DNA sequence phylogeny of four populations of the widely distributed cynomolgus macaque (*Macaca fascicularis fascicularis*). *J Hered* 99(3):254–264
- Bonhomme M, Cuartero S, Blancher A, Crouau-Roy B (2009) Assessing natural introgression in two biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*). *J Hered* 100(2):158–169
- Bunlungsup S, Kanthaswamy S, Oldt RF, Smith DG, Houghton P, Hamada Y, Malaivijitnond S (2017) Genetic analysis of samples from wild populations opens new perspectives on hybridization between long-tailed (*Macaca fascicularis*) and rhesus macaques (*Macaca mulatta*). *Am J Primatol* 79(3):e22621
- Chalk TB, Hain MP, Foster GL, Rohling EJ, Sexton PF, Badger M, Cherry SG, Hasenfratz AP, Haug GH, Jaccard SL, Martinez-Garcia A, Palike H, Pancost RD, Wilson PA (2017) Causes of ice age intensification across the mid-pleistocene transition. *Proc Natl Acad Sci USA* 114(50):13114–13119
- Ciani AC, Stanyon R, Scheffrahn W, Sampurno B (1989) Evidence of gene flow between Sulawesi macaques. *Am J Primatol* 17(3):257–270
- Cortes-Ortiz L, Duda TF, Canales-Espinosa D, Garcia-Orduna F, Rodriguez-Luna E (2007) Hybridization in large-bodied new world primates. *Genetics* 176(4):2421–2425
- Delson E (1980) Fossil macaques, phyletic relationships and a scenario of deployment. *Macaques Stud Ecol Behav Evol* 10:30
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498
- Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28(8):2239–2252
- Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, Garcia-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, Challis R, Kumar S, Moreira G, Salazar C, Chouteau M, Counterman BA, Papa R, Blaxter M, Reed RD, Dasmahapatra KK, Kronforst M, Joron M, Jiggins CD, McMillan WO, Di Palma F, Blumberg AJ, Wakeley J, Jaffe D, Mallet J (2019) Genomic architecture and introgression shape a butterfly radiation. *Science* 366(6465):594–599
- Eudey AA (1979) Differentiation and dispersal of macaques (*Macaca* spp.) in Asia. California University, California
- Evans BJ, Supriatna J, Melnick DJ (2001) Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. *Evolution* 55(9):1686–1702
- Evans BJ, Supriatna J, Andayani N, Setiadi MI, Cannatella DC, Melnick DJ (2003) Monkeys and toads define areas of endemism on Sulawesi. *Evolution* 57(7):1436–1443
- 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(9):170351
- Fan Z, Zhao G, Li P, Osada N, Xing J, Yi Y, Du L, Silva P, Wang H, Sakate R, Zhang X, Xu H, Yue B, Li J (2014) Whole-genome sequencing of Tibetan macaque (*Macaca thibetana*) provides new insight into the macaque evolutionary history. *Mol Biol Evol* 31(6):1475–1489
- 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
- Fan Z, Zhou A, Osada N, Yu J, Jiang J, Li P, Du L, Niu L, Deng J, Xu H, Xing J, Yue B, Li J (2018) Ancient hybridization and admixture in macaques (genus *Macaca*) inferred from whole genome sequences. *Mol Phylogenet Evol* 127:376–386
- Fang X, Zhang Y, Zhang R, Yang L, Li M, Ye K, Guo X, Wang J, Su B (2011) Genome sequence and global sequence variation map with 5.5 million SNPs in Chinese rhesus macaque. *Genome Biol* 12(7):R63
- Fooden J (1979) Taxonomy and evolution of the sinica group of macaques: I. Species and subspecies accounts of *Macaca sinica*. *Primates* 20(2):109–140
- Fooden J (1995) Systematic Review of Southeast Asian Longtail Macaques, *Macaca fascicularis* (Raffles, [1821]). *Fieldiana Zool* 81:1–206
- Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, Silva PM, Galaverni M, Fan Z, Marx P, Lorente-Galdos B, Beale H, Ramirez O, Hormozdiari F, Alkan C, Vila C, Squire K, Geffen E, Kusak J, Boyko AR, Parker HG, Lee C, Tadiogola V, Wilton A, Siepel A, Bustamante CD, Harkins TT, Nelson SF, Ostrander EA, Marques-Bonet T, Wayne RK, Novembre J (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet* 10(12):e1004016
- Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43(10):1031–1034
- 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(2):497–513
- 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 (*M. cyclopis*). *Primates* 53(4):397–411
- Hamada Y, San AM, Malaivijitnond S (2016) Assessment of the hybridization between rhesus (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*) based on morphological characters. *Am J Phys Anthropol* 159(2):189–198
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167(2):747–760
- Hey J, Wang K (2019) The effect of undetected recombination on genealogy sampling and inference under an isolation-with-migration model. *Mol Ecol Resour* 19(6):1593–1609
- Hey J, Chung Y, Sethuraman A, Lachance J, Tishkoff S, Sousa VC, Wang Y (2018) Phylogeny estimation by integration over isolation with migration models. *Mol Biol Evol* 35(11):2805–2818
- Higashino A, Sakate R, Kameoka Y, Takahashi I, Hirata M, Tanuma R, Masui T, Yasutomi Y, Osada N (2012) Whole-genome sequencing and analysis of the Malaysian cynomolgus macaque (*Macaca fascicularis*) genome. *Genome Biol* 13(7):R58
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35(2):518–522
- Hoelzer GA, Melnick DJ (1996) Evolutionary relationships of the macaques. *Evolution and ecology of macaque societies*. Cambridge University, Cambridge
- Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111(1):147–164
- 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(8):1164–1179

- Jiang J, Yu J, Li J, Li P, Fan Z, Niu L, Deng J, Yue B, Li J (2016) Mitochondrial genome and nuclear markers provide new insight into the evolutionary history of macaques. *PLoS ONE* 11(6):e0154665
- Kalyaanamoorthy S, Minh BQ, Wong T, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14(6):587–589
- Koyama N, Takahata Y, Huffman MA, Norikoshi K, Suzuki H (1992) Reproductive parameters of female Japanese macaques: thirty years data from the arashiyama troops, Japan. *Primates* 33(1):33–47
- Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol* 34(7):1812–1819
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359
- Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496
- 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(2):242–249
- 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(4):360–371
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1):268–274
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158(2):885–896
- Osada N, Uno Y, Mineta K, Kameoka Y, Takahashi I, Terao K (2010) Ancient genome-wide admixture extends beyond the current hybrid zone between *Macaca fascicularis* and *M. mulatta*. *Mol Ecol* 19(14):2884–2895
- Osada N, Hettiarachchi N, Adeyemi BI, Saitou N, Blancher A (2015) Whole-genome sequencing of six Mauritian Cynomolgus macaques (*Macaca fascicularis*) reveals a genome-wide pattern of polymorphisms under extreme population bottleneck. *Genome Biol Evol* 7(3):821–830
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2(12):e190
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D (2012) Ancient admixture in human history. *Genetics* 192(3):1065–1093
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575
- Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ, Kelso J, Slatkin M, Pääbo S (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468(7327):1053–1060
- Rivas-González I, Rousselle M, Li F, Zhou L, Dutheil JY, Munch K, Shao Y, Wu D, Schierup M, Zhang G (2023) Pervasive incomplete lineage sorting illuminates speciation and selection in primates. *Science* 380(6648):eabn4409
- Roos C, Boonratana R, Supriatna J, Fellowes J, Groves C, Nash S, Rylands A, Mittermeier R (2014) An updated taxonomy and conservation status review of Asian primates. *Asian Primates* 4:2–38
- 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
- Sambrook J, Russel DW (2001) Molecular cloning: a laboratory manual, 3rd edn. CSHL Press, New York
- Schliep KP (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* 27(4):592–593
- Smith DG, McDonough JW, George DA (2007) 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* 69(2):182–198
- 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(4):433–449
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68(4):978–989
- Stevison LS, Kohn MH (2008) Determining genetic background in captive stocks of cynomolgus macaques (*Macaca fascicularis*). *J Med Primatol* 37(5):311–317
- Stevison LS, Kohn MH (2009) Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. *Mol Ecol* 18(12):2457–2475
- Sugiyama Y, Kurita H, Matsui T, Kimoto S, Shimomura T (2009) Carrying of dead infants by Japanese macaque (*Macaca fuscata*) mothers. *Anthropol Sci* 117(2):113–119
- Sussman RW, Tattersall I (2008) Distribution, abundance, and putative ecological strategy of *Macaca fascicularis* on the Island of Mauritius, Southwestern Indian Ocean. *Folia Primatol* 46(1):28–43
- Takahata Y, Suzuki S, Agetsuma N, Okayasu N, Sprague DS (1998) Reproduction of wild Japanese macaque females of Yakushima and Kinkazan Islands: a preliminary report. *Primates* 39(3):339–349
- Tan X, Qi J, Liu Z, Fan P, Liu G, Zhang L, Shen Y, Li J, Roos C, Zhou X, Li M (2023) Phylogenomics reveals high levels of incomplete lineage sorting at the ancestral nodes of the macaque radiation. *Mol Biol Evol* 40(11):msad229
- 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
- Tosi AJ, Disotell TR, Morales JC, Melnick DJ (2003a) Cercopithecine Y-chromosome data provide a test of competing morphological evolutionary hypotheses. *Mol Phylogenet Evol* 27(3):510–521
- Tosi AJ, Morales JC, Melnick DJ (2003b) Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution* 57:1419–1435
- Tung J, Barreiro LB (2017) The contribution of admixture to primate evolution. *Curr Opin Genet Dev* 47:61–68
- Vanderpool D, Minh BQ, Lanfear R, Hughes D, Murali S, Harris RA, Raveendran M, Muzny DM, Hibbins MS, Williamson RJ, Gibbs RA, Worley KC, Rogers J, Hahn MW (2020) Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biol* 18:e3000954
- Yan G, Zhang G, Fang X, Zhang Y, Li C, Ling F, Cooper DN, Li Q, Li Y, van Gool AJ, Du H, Chen J, Chen R, Zhang P, Huang Z, Thompson JR, Meng Y, Bai Y, Wang J, Zhuo M, Wang T, Huang Y, Wei L, Li J, Wang Z, Hu H, Yang P, Le L, Stenson PD, Li B, Liu X, Ball EV, An N, Huang Q, Zhang Y, Fan W, Zhang X, Li Y, Wang W, Katze MG, Su B, Nielsen R, Yang H, Wang J, Wang X, Wang J (2011) Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nat Biotechnol* 29:1019–1023

- Yang F, Shi L (1994) Studies of the mitotic chromosomes, meiosis and spermatogenesis of a macaque hybrid. *J Genet Genomics* 21:24–29
- Zhang SJ, Liu CJ, Yu P, Zhong X, Chen JY, Yang X, Peng J, Yan S, Wang C, Zhu X, Xiong J, Zhang YE, Tan BC, Li CY (2014) Evolutionary interrogation of human biology in well-annotated genomic framework of rhesus macaque. *Mol Biol Evol* 31:1309–1324
- 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
- Zinner D, Groeneveld LF, Keller C, Roos C (2009) Mitochondrial phylogeography of baboons (*Papio* spp.): indication for introgressive hybridization? *BMC Evol Biol* 9:83
- Zinner D, Arnold ML, Roos C (2011) The strange blood: natural hybridization in primates. *Evol Anthropol* 20:96–103
- Zinner D, Kopp G, Roos C (2013) Family Cercopithecidae (old world monkeys). In: Setchell JM, Curtis DJ (eds) *Field and laboratory methods in primatology: a practical guide*. Cambridge University Press, Cambridge, pp 550–627

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.