Mitonuclear interactions and the origin of macaque society

Mitonuclear interactions and the origin of macaque society

By Jianlong Zhu,

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University
Master of Science (2022)
Hamilton, Ontario (Department of Biology)

TITLE: Mitonuclear interactions and the origin of macaque society
AUTHOR: Jianlong Zhu (McMaster University)
SUPERVISOR: Dr. Ben Evans
NUMBER OF PAGES: xi, 81

Acknowledgements

In this project, many people gave me their hands. It's hard to find the exact word to express my appreciation toward them.

To my supervisor, Dr. Ben Evans, your suggestions are always insightful, and your responses are always timely. I can't imagine how far I could go without your support. Thank you for providing me with the opportunity in academia. If luck was limited in one's life, I'm sure I used half to become your student.

To my committee members, Dr. Ian Dworkin and Dr. Ben Bolker, thank you for your suggestions on the statistical details and how to present my results. Your recommendations improve the quality of my project a lot.

To my Evans lab members, Tharindu Premachandra, Lindsey Kukoly, and Emmanuela Anele, thank you for your emotional support. Experience with you is the treasure in my life. To Dr. Brian Golding, thank you for providing the computational resource.

Lastly, I would be remiss in not mentioning my family and friends, especially my parents. Your unconditional supports and encouragements are my power to move forward.

Abstract

In most eukaryotes, aerobic respiration requires interactions between autosomally-encoded genes (Ninteract genes) and mitochondrial DNA, RNA, and protein. In species where females are philopatric, contrasting distributions of genetic variation in mitochondrial and nuclear genomes creates variation in mitonuclear interactions that may be subject to natural selection. To test this expectation, we turned to a group with extreme female philopatry: the macaque monkeys. We examined four genomic datasets from (i) wild caught and (ii) captive populations of rhesus macaque, which is the most widely distributed non-human primate, and (iii) the stump-tailed macaque and (iv) a subspecies of longtail macaque, both of whose mitochondrial DNA is introgressed from a highly diverged ancestor. We identified atypically long runs of homozygosity, low polymorphism, high differentiation and/or rapid protein evolution associated with Ninteract genes compared to non-Ninteract genes. These metrics suggest a subset of Ninteract genes were independently subject to natural selection in multiple species. Selection on mitonuclear interactions is thus a factor in macaque genome evolution that could have influenced aspects of macaque societies including species diversity, ecological breadth, female-biased adult sex ratio and demography, sexual dimorphism, and mitonuclear phylogenomics.

Acronyms

ROH : Runs of Homozygosity

MRCA : Most recent common ancestor

SNP : Single nucleotide polymorphism

ANI : Average nucleotide identity

WGD : Whole genome duplication

PCA : Principal component analysis

M. f. aurea : *Macaque fascicularis aurea*

M. fascicularis : *Macaque fascicularis*

M. thibetana : *Macaque thibetana*

M. arctoides : *Macaque arctoides*

M. assamensis : *Macaque assamensis*

M. mulatta : *Macaque mulatta*

X. laevis : *Xenopus laevis*

X. tropicalis : *Xenopus tropicalis*

X. mellotropicalis : *Xenopus mellotropicalis*

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Jianlong Zhu, declare that this thesis titled, "Mitonuclear interactions and the origin of macaque society" and the work presented in it are my own.

This work represents original research conducted by myself in collaboration with Dr. Ben Evans undner the supervision of Dr. Ben Evans. Dr. Ian Dworkin and Dr. Ben Bolker provided suggestions on statistical models and results presentation. Ben Evans contibuted the design of the study, collected data, constructed mitochondrial phylogenetic tree, performed outlier analysis and introgression analysis.

This project has been submitted for peer-reviewed to the journal *Genome Biology and Evolution*

Chapter 1

Mitonuclear interactions and the origin of macaque society

1.1 Significance Statement

Endosymbiosis between the mitochondria and eukaryotic cell may have contributed to a myriad of important biological innovations (the nucleus, sex, the sequestered germline, speciation and more) yet it remains unclear whether ongoing natural selection on mitonuclear interactions is atypically strong, persistent, or prevalent. To address this question, we searched for and found signatures of natural selection on mitonuclear interactions in several species of macaque monkey that share an extreme version of a common behavioral paradigm – female philopatry and obligate male migration. These results argue that mitonuclear interactions could have effected key features of macaque societies such as species diversity, female-biased sex ratios, mitonuclear phylogenomic discordance, and sexual dimorphism. More generally, these results evidence links between behavior and genome evolution.

1.2 Data Accessibility

Data from this study were either downloaded from NCBI (accession numbers DRS139837, DRS139838, DRR219371, SRA023855, SRR1564766, SRR2981114, SRR1024051), provided by the authors of (Warren et al. 2020), or obtained from Supplementary Information (Liu et al. 2018; Bailey and Stevison 2021).

1.3 Introduction

In many organisms, genetic networks consist of components that are encoded by the mitochondria and nuclear genomes, even though these genomic compartments have different modes of inheritance, ploidy, and rates of recombination and mutation. For instance, the crucial processes of oxidative phosphorylation (OXPHOS genes), loading of amino acids on mitochondrial tRNAs (ARS2 genes), the formation of the mitochondrial

ribosome (MRP genes), and mitochondrial replication (REP genes) all require collaborative interactions between nuclear-encoded proteins and mitochondrial DNA, RNA, and proteins. Interacting components of these genetic systems therefore must co-evolve in the context distinctive evolutionary dynamics in each compartment (Rand et al. 2004; Burton and Barreto 2012; Hill 2020).

Some forms of sex-specific behavior lead to differences in the spatial distribution of genetic variation in mitochondrial DNA, which is maternally inherited, and autosomal DNA, which is biparentally inherited (Melnick and Hoelzer 1992; Caballero 1995; Hoelzer 1997; Evans and Charlesworth 2013). For example, female philopatry geographically anchors diverged mitochondrial lineages, and if this is coupled with obligate male migration, mitochondrial lineages from philopatric females will be constantly exposed to novel and not co-evolved genetic variation in the autosomes that is imported from afar by migrating males. Added to this, the rate of mutation in mitochondrial DNA is 20 times faster than in autosomal DNA in many mammals (Osada and Akashi 2012; Allio et al. 2017), and natural selection is expected to be less efficient at purging deleterious mutations from the mitochondrial genome compared to the autosomes because the former is non-recombining and has a smaller effective population size. Mitonuclear interactions thus are hypothesized to be an example of a Red Queen effect where nuclear genes must constantly evolve to maintain function with the mitochondrial genome; this effect has potentially ancient contributions to fundamental aspects of biology such as the origin of the nucleus, sexual reproduction, genome complexity, germline sequestration, and speciation (Martin and Koonin 2006; Lane 2014; Radzvilavicius et al. 2016; Hill 2017, 2019).

Macaque monkeys are compelling subjects for the study of mitonuclear interactions because females exhibit extreme philopatry and males generally migrate from their natal group (Dittus 1975; Pusey 1987; Swedell 2010; Clutton-Brock and Lukas 2012). These sex-differences in migratory behavior lead to sharp differences in mitochondrial diversity between populations while variation in nuclear DNA may be comparatively undifferentiated, even across vast geographic distances (Melnick and Hoelzer 1992; Melnick et al. 1993; Hoelzer 1997; Evans et al. 2001; Evans et al. 2003; Evans et al. 2020). For example, rhesus macaques have the largest distribution of any non-human primate (Wolfheim 1983; Southwick et al. 1996) and though as many as 13 subspecies have been previously proposed (Fooden 2000), recent taxonomic assessments consider this species to be morphologically homogeneous (Roos and Zinner 2015). Despite this, there is considerable divergence between intraspecific mitochondrial lineages (Smith and McDonough 2005; Zinner et al. 2013; and see below). Extant rhesus mitochondrial genomes evolved from a most recent common ancestor (MRCA) about two million years ago (Roos et al. 2019; Evans et al. 2020), which is roughly ten times older than the MRCA of human mitochondria (Fu et al. 2013). Highly diverged mitochondrial genomes are also found in different populations of other widespread macaque species, including pig-tailed (Evans et al. 2020) and long-tailed macaques (Yao et al. 2017), and in other papionin monkeys that share this behavioral system, such as anubis baboons (Roos et al. 2020) and mandrills

(Telfer et al. 2003). Additionally, there are striking examples of inter-species introgression of mitochondrial genomes in macaques. Mitochondria of stump-tailed macaques, *M. arctoides*, is most closely related to fascicularis group macaques (*M. mulatta*, *M. fascicularis*) whereas the nuclear genome is most closely related to sinica group macaques (*M. assamensis*, *M. thibetana*), suggesting ancient introgression (Tosi et al. 2000; Fan et al. 2018). Likewise, a subspecies of long-tailed macaques, *M. fascicularis aurea*, carries introgressed mitochondria from an ancestor of sinica group macaques (Matsudaira et al. 2018) – probably an ancestor of *M. sinica* (Evans et al. 2020). These examples illustrate the capacity of the nuclear genome to have compatible interactions with highly diverged mitochondrial genomes.

Recently, an analysis of *silenus* and Sulawesi group macaques recovered signs of natural selection on mitonuclear interactions (Evans et al. 2021). Here, we attempt to generalize these results across other components of macaque diversity by analyzing four genomic datasets: (i) 79 wild *M. mulatta* from China, (ii) 89 captive *M. mulatta* from India, (iii) *M. arctoides* and close relatives (ten individuals from five species), and (iv) *M. f. aurea* and close relatives (seven individuals from three species, including two subspecies of *M. fascicularis*). Our results provide evidence for pervasive and recurrent natural selection on autosomal genes that interact with mitochondrial DNA, RNA, and protein (Ninteract genes), arguing for the possibility that mitonuclear interactions may have influenced the evolution of several distinctive features of macaque societies.

1.4 Materials and Methods

1.4.1 Data

This study considered genomic data from four datasets: (i) 79 wild-sampled *M. mulatta* individuals from China (Liu et al. 2018), (ii) captive *M. mulatta* from India (subsetted from 850 to 89 individuals), (iii) *M. arctoide*s and close relatives (subsetted from 13 to ten individuals total to exclude three distantly related *silenus* group species), and (iv) *M. f. aurea* and close relatives (seven individuals). The first three datasets were previously published, and analysis was performed on genotype (vcf) files from those studies (Liu et al. 2018; Warren et al. 2020; Bailey and Stevison 2021). Additional information on these data is provided in the Supplement.

1.4.2 Mitochondrial genome assembly and phylogeny

As detailed above, we used a mitochondrial phylogeny to guide our analyses of autosomal genes. *De novo* assembly of mitochondrial genomes was performed as described in the Supplement. We combined de novo assemblies of 175 complete mitochondrial genomes with 222 complete genomes that were obtained from GenBank as detailed in (Evans et al. 2020) for a total of 397 complete genome sequences including three outgroup genomes (two baboons, one gelada). Evolutionary relationships among the mitochondrial genomes were estimated using Bayesian and maximum likelihood approaches as detailed in the Supplement.

1.4.3 Ninteract genes

Using annotation files of two reference genomes from rhesus macaques (versions 8 and 10), we searched for autosomal genes that are involved with any of the four mitonuclear interactions discussed above (OXPHOS, ARS2, MRP, REP; Gaspari et al. 2004; Sissler et al. 2017; Bogenhagen et al. 2018; Signes and Fernandez-Vizarra 2018) and that have direct or very close interactions with mitochondrial DNA, RNA or protein. We identified 199 autosomal Ninteract genes in each version in the following Ninteract categories: OXPHOS (101 and 100 in versions 8 and 10), ARS2 (77 and 78), MRP (17 in both assembly versions), and REP genes (4 in both versions). OXPHOS complex II was excluded from the Ninteract category because it comprises only autosomal encoded proteins. A total of 21,094 and 20,897 autosomal non-Ninteract genes were identified in versions 8 and 10, respectively. In both assembly versions, two and zero Ninteract genes were identified on the X and Y chromosomes, respectively. Because several factors influence genetic variation on the sex chromosomes in ways that differ from the autosomes (mutation rate, effective population size, hemizygosity, natural selection), the sex chromosomes were excluded from all analyses.

1.4.4 Analysis of polymorphism within and between lineages

We predicted that natural selection on Ninteract genes would decrease variation in these genes and their flanking regions within lineages and increase differentiation in pairwise comparisons between lineages. To test this, we analyzed runs of homozygosity (ROHs) and pairwise nucleotide diversity of polymorphic sites (π) in groups of individuals defined by mitochondrial variation and taxonomy, and population differentiation (F_{ST}) between these groups.

ROHs were identified using bcftools version 1.4 (Narasimhan et al. 2016). We defined Ninteract ROHs to be ROHs that contained the start of transcription of at least one Ninteract gene, and non-Ninteract ROHs to be ROHs that contained the start of transcription of at least one non-Ninteract gene but no Ninteract genes. We used linear models to test the hypothesis that the predictor variable Ninteract, which indicated whether (1) or not (0) an ROH was an Ninteract ROH, was positively correlated with the natural logarithm of ROH_length in 100 kilobase pairs. We log transformed the response variable because the distribution of the untransformed variable was skewed. An interaction term with the number of genes in each ROH (gene_number) was included in the linear models for each species or population in each dataset. While we did not have an a priori prediction for the interaction term in this model or in other linear models discussed below, we included this term as a way of avoiding making the strong assumption that our predictor variables were independent. ROHs lacking a start of transcription sites for at least one gene were excluded from this analysis.

F_{ST} and π were analyzed in genomic windows, and defined Ninteract windows to be any genomic window that contained the start of transcription of at least one Ninteract gene, and non-Ninteract windows to be any genomic window that contained start of transcription of non-Ninteract genes but no Ninteract genes. We considered two window

sizes (30kb, 100kb) and used linear models to test the hypothesis that F_{ST} and π were higher or lower, respectively, in Ninteract windows compared to non-Ninteract windows. We included an interaction term between gene_number, which is the number of genes in a genomic window, and the predictor variable Ninteract, which indicated whether (1) or not (0) a window was an Ninteract window, on the response variable (F_{ST} or π). Because their distributions were not especially skewed, these response variables were not transformed. F_{ST} was calculated using the method of Hudson et al. (1992) with a minimum allele frequency cutoff of 0.05; F_{ST} of genomic windows was calculate as the average of the F_{ST} values at each variable position.

Linked genomic regions tend to have similar genomic properties, which violates the assumption that the response variable of linear models is independently distributed. For this reason, we used block bootstrapping to estimate standard errors of the coefficients of our linear models of genomic windows using the block.glm function of the R package genomic.autocorr version 1.0-1 (Wallace and Burren 2017). Autocorrelation plots that were generated using the acf function of the R stats package version 3.6.2 (R_Core_Team 2018) suggested that autocorrelation decayed within a block size of 30; this value was used for the block bootstrapping for 30kb and 100kb wondows with 1000 replicates. P values were then calculated from standard errors following (Altman and Bland 2011).

Because ROHs were not divided into equally sized windows, we used a combination of linear models and permutation tests to evaluate whether the observed difference between Ninteract ROHs length and non-Ninteract ROHs length were significantly higher than expected by chance. In the permutations, the observed difference between the mean ROH length of Ninteract and non-Ninteract ROHs was compared to a distribution of differences calculated from by randomizing the assignment of Ninteract and non-Ninteract ROHs for 1000 randomizations. We had a one-sided expectation that observed differences should be atypically large compared to the randomized values. As a complement to the block boostrap analyses discussed above, we also performed these permutations on the genomic window metrics (F_{ST}, π).

Outlier analysis were performed for F_{ST} and π using Cook's distance, using linear models as in (Evans et al. 2021). This analysis evaluates whether a subset of the data (here Ninteract windows) has more extreme outliers (upper outliers for F_{ST}, lower outliers for π), than expected based on the rest of the data (the non-Ninteract windows). In all analyses, genomic windows lacking a start of transcription sites for at least one gene were excluded.

Tajima's D was calculated in 30kb and 100 kb genomic windows using vcftools version 0.1.16 (Danecek et al. 2011). Permutation tests with 1000 replications were used to test the hypothesis that the difference between the mean Tajima's D of Ninteract windows and non-Ninteract windows was more negative than expected by chance as described in Evans et al. (2021). This analysis was performed only on the two *M. mulatta* datasets because the intraspecific sample sizes were small for the *M. arctoides* and the *M. f. aurea* datasets.

1.4.5 *dNdS*

We expected protein evolution to be faster in Ninteract compared to non-Ninteract genes. To test this, for each dataset, a phylogeny was obtained among individuals by first converting the genotypes to nexus files using the vcf2philip script (https://github.com/edgardomortiz/vcf2phylip) and requiring at least 80% of the taxa to have called genotypes. For the *M. arctoides* and *M. f. aurea* datasets, variable positions within the complete genome were used; for the much larger datasets from *M. mulatta* variable positions only on chromosome 1 were used. A phenogram using the neighbor-joining algorithm using PAUP* (Swofford 2002) and this topology was used for analysis of *dN/dS*. For each gene, we used the codeml program within the PAML package version 4.9 (Yang 2007) to estimate the rate ratio of nonsynonymous to synonymous substitutions per site (*dN/dS*) over each phylogeny under a model with one *dN/dS* parameter. This analysis was performed for each Ninteract gene and 3000 randomly selected non-Ninteract genes. We discarded genes that lacked an "ATG" start site, contained premature stop codons, whose coding region was not a multiple of three nucleotides long, or if the estimated *dN/dS* value was greater than five. Permutation tests were used to test whether the observed difference between the mean *dN/dS* ratios of Ninteract and non-Ninteract genes was significantly higher than expected by chance as described above.

1.4.6 Comparisons across datasets

To assess whether the signatures of natural selection were shared across metrics (π, F_{ST}, ROHs, *dN/dS*) and to explore the possibility that the genes are repeatedly subject to natural selection in different species, we identified and quantified genomic windows that were lower π outliers and also upper F_{ST} outliers using the R package UpSetR (Conway et al. 2017), and we also quantified residence of these outliers in ROHs. In addition, we used permutation tests similar to those detailed above to test whether these outliers also tended to have higher *dN/dS* values compared to the non-Ninteract genes.

1.4.7 Analyses of introgression

We used two approaches to test for evidence of introgression and if detected, to permit us to evaluate hypotheses regarding introgression of Ninteract and non-Ninteract windows that are articulated in the results. The first approach uses a hidden Markov model to infer introgression in 100 kb windows as implemented by the software admixfrog (Peter 2020). Because the data produced errors under some SNP densities, for each chromosome we thinned the data to differing degrees, removing variant sites that were within 200-2000 positions from other variable sites. Because the extent of introgression that was detected in *M. arctoides* and *M. f. aurea* was low, no statistical analyses were performed on these introgression blocks.

The second approach we used to study introgression used Patterson's D statistic (Patterson et al. 2012). This analysis was performed using 30 kb and 100 kb windows and using taxon settings that matched established evolutionary relationships among

nuclear genes or genomes (Fan et al. 2018; Matsudaira et al. 2018), namely for the *M. f. aurea* dataset (((*M. fascicularis*, *M. f. aurea*), *M. assamensis*), *M. nemestrina*) and for the *M. arctoides* dataset (((*M. arctoides*, *M. assamensis*), *M. mulatta*), *M. nemestrina*). For both datasets we also analyzed Patterson's D after substituting the *M. thibetana* genome for the *M. assamensis* genomes. For the *M. f. aurea* dataset we used previously published genomic data from an *M. nemestrina* individual (PM1206) as an outgroup; genotypes for this individual were called along with the other individuals in this dataset for this analysis only using the pipeline described above. Following Evans et al. (2021), permutation tests were used to evaluate expectations based on evolutionary relationships among mitochondrial genomes of these samples that are described in the results.

1.5 Results

We used genomic data to perform de novo assemblies of mitochondrial genomes then estimated phylogenetic relationships among the mitochondrial genomes carried by each sample in each of four genomic datasets. Then, for each dataset, we performed tests for natural selection on autosomal Ninteract genes, including comparisons between Ninteract and non-Ninteract genes of residence in runs of homozygosity (ROHs), population differentiation (F_{ST}), pairwise nucleotide diversity of polymorphic positions (π), Tajima's D statistic, and the rate ratio of nonsynonymous to synonymous substitutions per site (dN/dS). These analyses were focused on groups of individuals defined by phylogenetic relationships among complete mitochondrial genomes and/or taxonomy. We additionally conducted comparisons between individuals in each dataset using partial Mantel tests.

1.5.1 Mitochondrial phylogenetics

Inferred evolutionary relationships among major lineages of 405 complete mitochondrial genomes, including 183 that were *de novo* assembled for this study, are depicted in Fig. 1. Within *M. mulatta*, three clades were carried by wild macaques from China (hereafter Red, Blue, Purple, with 19, 34, and 26 individuals, respectively, with the Blue clade also including a GenBank genome: KJ567051.1). Two other clades were carried by all but one captive *M. mulatta* from India (hereafter Orange and Brown, with 81 and 8 individuals, respectively, with the Orange clade also including two GenBank genomes: KJ567053.1, AY612638.1). One captive *M. mulatta* putatively from India (MMUL_IN_32754) carried mitochondria from the Red clade. The captive populations of *M. mulatta* from India and three of five subspecies of *M. mulatta* from China (all except *M. m. brevicaudus* and *M. m. tcheliensis*) have paraphyletic relationships among their mitochondrial genomes. Because we are interested in mitonuclear interactions, our analyses of *M. mulatta* are based on inferred mitochondrial clades (Blue, Red, Purple, Orange, Brown), as opposed to using subspecies taxonomy or captive research centers (See Supplement for more details). Mitochondrial relationships within the focal taxa *M. arctoides* and *M. f. aurea* are monophyletic based on four genomes from each

taxon. Other details about mitogenomic phylogenetic relationships are discussed in the Supplement.

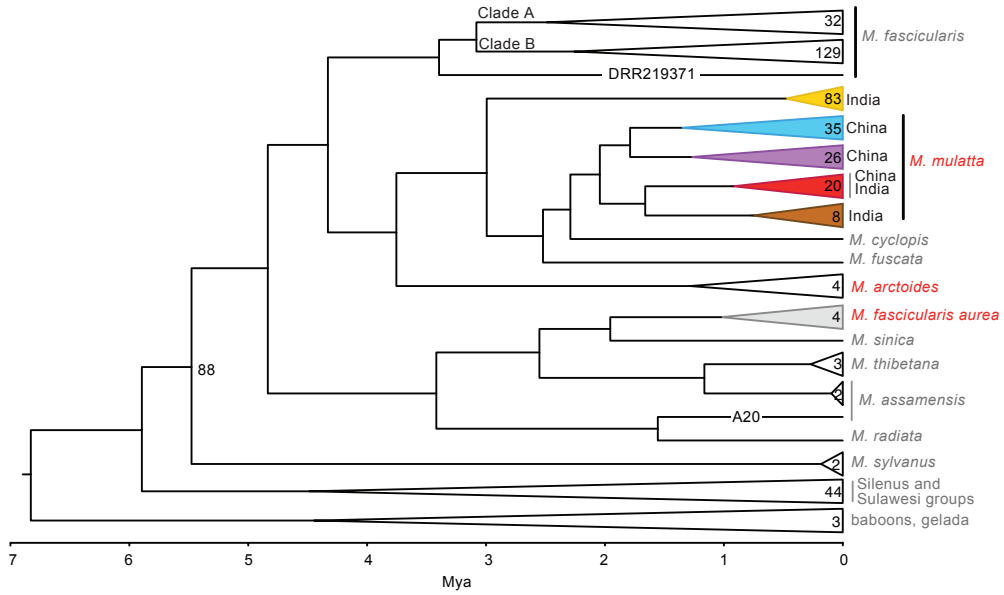

FIG. 1. Consensus tree from Bayesian analysis of 394 complete mitochondrial genome sequences from macaques and three outgroup genomes (two baboons, one gelada). Focal species are in red font and the scale bar indicates millions of years ago (Mya). All nodes have 100% posterior probability except where labeled with a posterior probability of 88%. Numbers inside collapsed clades refer to the number of genomes, which sometimes are greater than the number of individuals whose autosomal data were analyzed because other mitogenomes are included in this phylogeny (Methods). A diverged mitochondrial genome from a *M. fascicularis* individual from Vietnam is labeled on one branch with its accession number (DRR219371) and a diverged mitochondrial genome from a *M. assamensis* individual is labeled with its sample ID (A20). Colors correspond to the clade designations in the main text.

1.5.2 Runs of homozygosity (ROHs)

Runs of homozygosity (ROHs) are genomic regions are contiguous genomic tracks where homozygous genotypes are present. The lengths and abundances of ROHs are associated with several phenomena including inbreeding, demographic changes, recombination rates, and natural selection (Ceballos et al. 2018). In humans, for example, ROHs are more commonly observed surrounding alleles that have been subject to natural selection (Pemberton et al. 2012), such as alleles for lactase persistence (Tishkoff et al. 2007), than expected by chance. If Ninteract genes are frequently subject to natural selection, we expected them to more frequently reside in runs of homozygosity (ROHs), and for these ROHs to be longer than other ROHs that contain only non-Ninteract genes.

We recovered support for this expectation using linear models and permutation tests. For all linear models, there was a significant negative interaction between the effect of Ninteract genes and gene number; this indicates that the effect of Ninteract genes on

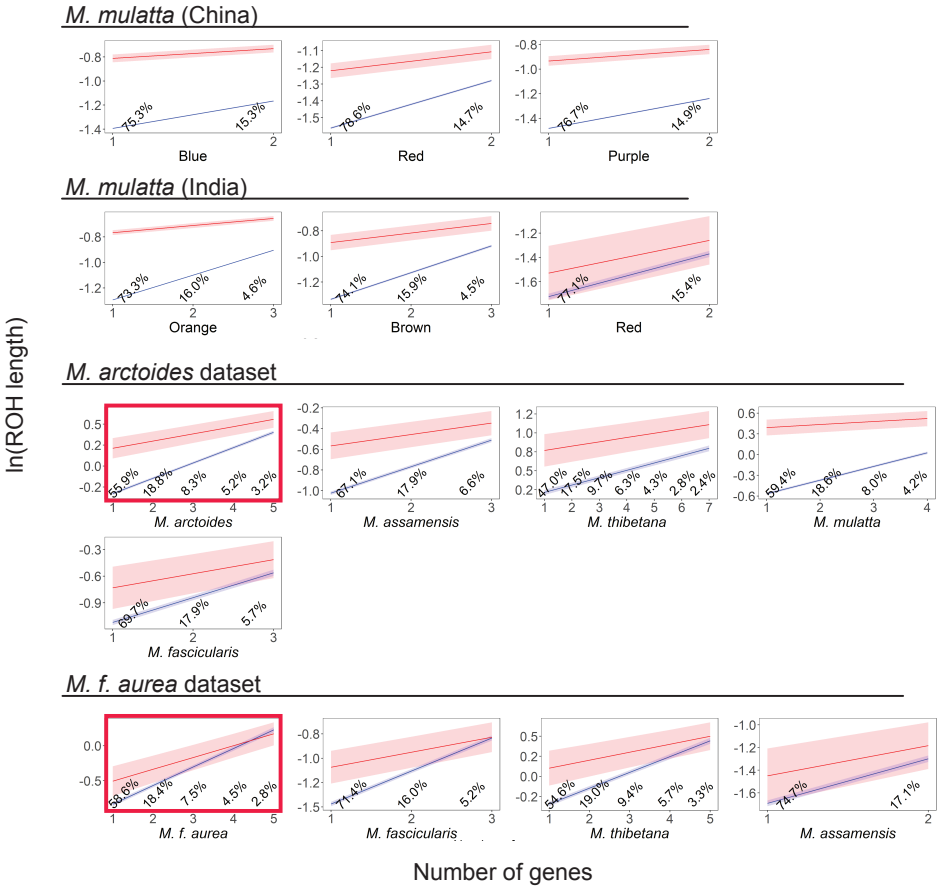

FIG. 2. In most species or population in all four datasets, the predicted values (marginal means) of the natural logarithm of the lengths of ROHs (in units of 100kb) are longer if they contain Ninteract genes (red) compared to those that contain only other genes (blue) across biologically relevant numbers of genes in each ROH (Number of genes). The limit of each x-axis is the 90^{th} percentile for number of genes in ROHs for each population or species; percentages indicate the proportion of all ROHs that contain each of the plotted numbers of genes. Shading indicates the 95% confidence interval of the predicted values. Red boxes highlight two species where we expected a particularly prominent effect of Ninteract genes on ROH length.

ROH length is smaller or even negative in ROHs that contained many genes (Table S1). One way to evaluate the biological relevance of this interaction is to consider ROHs in terms of the distribution of gene numbers that each ROH carries (from one to the maximum carried by any ROH that has at least one gene). For the ranges of ROH gene numbers from the 1st to the 90th percentiles, the estimated marginal means (predicted values) of all linear models indicated that ROHs that contained Ninteract genes are longer than those that contained only other (non-Ninteract) genes (Fig. 2, Table S1). These disparities are also evident in density plots which indicate that some ROHs with Ninteract genes are atypically long compared to ROHs with only other genes or ROHs with no genes (Figs. S1–S4). Put another way, for most ROHs that contain biologically

relevant gene numbers, ROHs with Ninteract genes tend to be longer than those that carry only other genes.

Permutation tests indicated that Ninteract ROHs were individually significantly longer than non-Ninteract ROHs in only one population (*M. mulatta* in the *M. arctoides* dataset; Table S1), but across all comparisons the permutation tests also indicated that Ninteract ROHs were significantly longer (P = 0.020, χ^2 test, degrees of freedom = 30, Fisher's method for combining probabilities). We note that these permutation tests do not control for the interaction between ROH length and gene number, but are less affected by autocorrelation than the linear models.

1.5.3 Genomic windows: F_{ST}, π, Tajima's D

F_{ST} is an index of population subdivision that varies from 0 (no subdivision) to 1 (fixation of different alleles in each population). We expected F_{ST} of Ninteract windows (genomic windows that contain the start of transcription of at least one Ninteract gene) to be higher than non-Ninteract windows (genomic windows that contain the start of transcription of at least one non-Ninteract gene but no Ninteract genes) due to species-specific natural selection on mitonuclear interactions. With the exceptions of *M. mulatta* from India and a handful of pairwise comparisons in other datasets, linear models generally found support for this expectation (Figs. 3, S5, Tables S2, S3). Permutation tests also found F_{ST} to be significantly higher in Ninteract windows compared to non-Ninteract windows in most individual comparisons (Tables S2, S3), and also across all comparisons (P « 0.0001 for both window sizes,χ^2 test, degrees of freedom = 44, Fisher's method for combining probabilities, Tables S2, S3).

Pairwise nucleotide diversity (π) is the average proportion of nucleotide differences between all pairs of sequences in a population (Nei 1987). We expected Ninteract windows to have lower π than to non-Ninteract windows due to selection on mitonuclear interactions. Consistent with this expectation, the predicted π was often lower in Ninteract windows compared to non-Ninteract windows over biologically relevant numbers of genes in each window (Figs. 4, S6, Tables S4, S5). This was particularly apparent in *M. f. aurea* and also in *M. arctoides* but with more variation among window sizes in *M. arctoides* (Fig. 4, S6). Permutation tests also found π to be significantly lower in Ninteract windows compared to non-Ninteract windows in several individual comparisons (Tables S4, S5), and across all comparisons (P « 0.0001 for both window sizes,χ^2 test, degrees of freedom = 30, Fisher's method for combining probabilities, Tables S3, S4).

Tajima's D (Tajima 1989) is a population genetic statistic that is influenced by demography and natural selection; negative values indicate an excess of low frequency polymorphisms, which is a hallmark of expansion in population size or of a recent selective sweep (Fay and Wu 1999). In a constant-sized Wright-Fisher population with no natural selection, the expected value of Tajima's D is zero, but demographic changes in natural populations may cause a different genome-wide mean value of Tajima's D, even in the absence of natural selection. For this reason, we compared the overall distributions of Tajima's D in genomic regions in Ninteract windows, non-Ninteract windows,

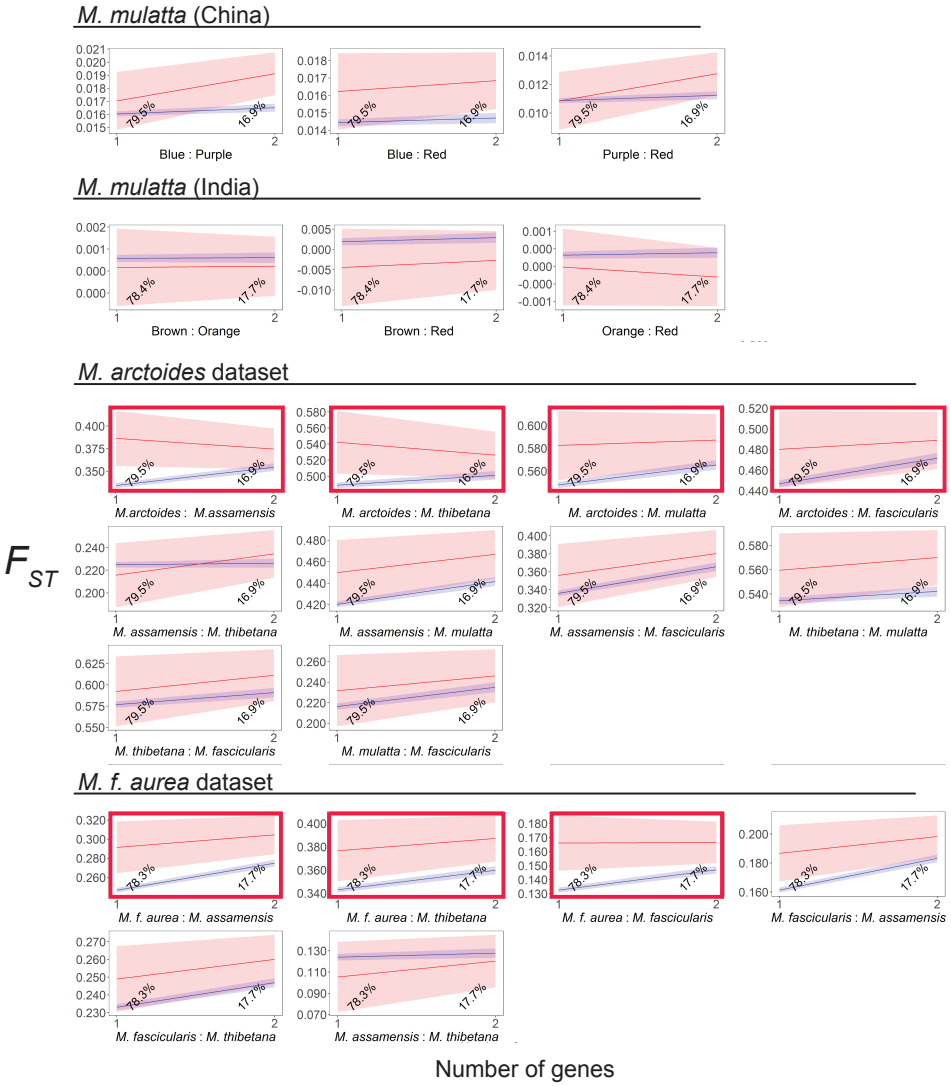

FIG. 3. In three of the four datasets (all but *M. mulatta* from India), the predicted values (marginal means) for most species or population of F_{ST} in 30 kb windows are generally higher in Ninteract windows (red) compared non-Ninteract windows (blue) across biologically relevant numbers of genes in each window (Number of genes). Plotting, shading, and highlights follow Fig. 2 except that here confidence intervals were obtained from block bootstrapping rather than standard errors from linear models.

and non-genic windows, with the expectation that the distribution of Tajima's D in Ninteract windows might be more negative or have a more negative tail than the other categories. Because the within species sample sizes were small for the *M. arctoides* and *M. f. aurea* datasets, we restricted our analysis of Tajima's D to the two *M. mulatta* datasets. Consistent with our expectation, our findings indicate that Tajima's D in Ninteract windows was more negative compared to non-Ninteract windows for nine of 10 of the *M. mulatta* populations we tested, and this difference was significant in four of

10 populations according to permutation tests (Supplementary Table S6) and across all comparisons for both window sizes ($P < 0.02$ for each window size,χ^2 test, degrees of freedom = 10, Fisher's method for combining probabilities).

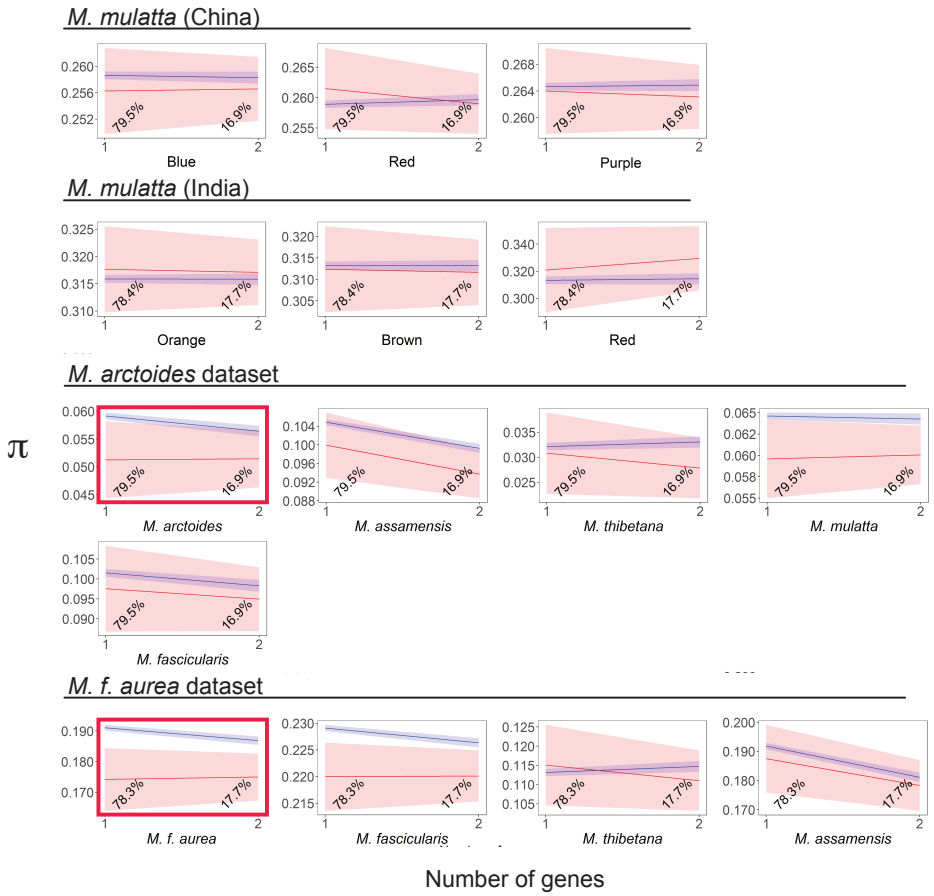

FIG. 4. In the M. arctoides and M. f. aurea datasets but in the M. mulatta from China and India datasets the predicted values (marginal means) for most species or population of π in 30 kb windows are generally lower in Ninteract windows (red) compared non-Ninteract windows (blue) across biologically relevant numbers of genes in each window (Number of genes). Plotting, shading, and highlights follow Fig. 3.

1.5.4 Outlier analysis of F_{ST} and π

The analyses of genomic windows detailed above were frequently consistent with our expectations associated with natural selection on mitonuclear interactions, but there were several exceptions where the expected difference between Ninteract and non-Ninteract windows was negligible. One possibility is that there could be some Ninteract genes that had atypically high F_{ST} and/or low π even though the trends for all Ninteract genes was sometimes similar to non-Ninteract genes. This is suggested by density plots that indicate that some Ninteract windows in most datasets have atypically high F_{ST} and atypically low π compared to non-Ninteract windows (Figs. S7-S14).

To explore this possibility, we performed an outlier analysis with the 30kb windows to test whether there was an excess of upper F_{ST} outliers or lower π outliers in Ninteract windows compared to non-Ninteract windows (Methods and Supplementary Information). This analysis identified an individually significant excess of upper F_{ST} outliers in almost all analyses ($P < 0.05$ for all comparisons except Brown:Orange and Brown:Red for the captive *M. mulatta* from India, binomial tests, Supplementary Results, Table S7, Figs. S15-S18). The excess of upper F_{ST} outliers were collectively significant across these datasets ($P \ll 0.0001$, χ^2 test, 44 degrees of freedom, Fisher's Method for combining probabilities). There also was an individually significant excess of lower π outliers for all taxa for all datasets ($P < 0.05$, binomial tests, Supplementary Results, Table S8, Figs. S19-S22) that were also collectively significant across all species or populations ($P \ll 0.0001$, χ^2 test, 30 degrees of freedom, Fisher's Method for combining probabilities).

1.5.5 The rate ratio of nonsynonymous to synonymous substitutions per site (*dN/dS*)

We expected protein evolution to be faster in Ninteract genes compared to non-Ninteract genes. Consistent with this expectation, for all four focal species, the difference between the mean *dN/dS* ratio of Ninteract genes and randomly selected non-Ninteract genes was positive, and permutation tests indicate that the observed difference was individually significant for two of four datasets (the *M. arctoides* and *M. f. aurea* datasets; Table S9). The rate of protein evolution was collectively significantly faster in Ninteract compared to non-Ninteract proteins across all datasets ($P = 0.0001$, χ^2 test, 8 degrees of freedom, Fisher's Method for combining probabilities).

1.5.6 Congruent signals of natural selection on the same Ninteract genes in different species

Comparison of statistics discussed above indicates that they provide both consistent and complementary insights into natural selection. For example, all (π) or almost all (F_{ST}) Ninteract outliers were also found in an Ninteract ROHs in at least one individual in the population or species being considered. Likewise, several genes that were upper outliers for F_{ST} were also lower outliers for π (Table S10). However, F_{ST} and π outliers were not observed to have higher dN/dS values compared to non-Ninteract genes ($P > 0.05$, permutation tests), which suggests that these metrics track distinctive signatures of natural selection.

These statistics also provide evidence that several Ninteract genes have been independently subjected to natural selection multiple times in different evolutionary lineages. Specifically, several Ninteract windows were upper F_{ST} outliers in multiple comparisons (Figs. S23-S26) or lower π outliers in multiple populations or species (Figs. S27-S30). Ten Ninteract windows were upper F_{ST} outliers and also were lower π outliers in at least four comparisons (F_{ST}) and populations/species (π). These ten windows contain the following Ninteract genes (and functional categories): CCDC56, CD14, NDUFA2, NDUFB8 (OXPHOS); HARS2 (ARS2); MRPL10, MRPL2, MRPL43, MRPL49,

MRPL55 (MRP). As discussed in the Supplement, many of these genes also were independently identified in a study of mitonuclear interactions in *silenus* group macaques that were not studied here (Evans et al. 2021).

1.5.7 Analyses of introgression

Another way that natural selection might influence mitonuclear interactions is by favoring or disfavoring gene flow of regions that contain Ninteract genes, depending on the phylogenetic affinities of the mitochondrial genomes of donor and recipient populations. To explore this possibility, we focused on the *M. arctoides* and *M. f. aurea* datasets because we were able to develop phylogenetically based expectations regarding introgression for each one. For *M. arctoides*, which carries mitochondrial DNA that is closely related to that of *M. mulatta* but whose nuclear genome is more closely related to sinica group macaques (*M. thibetana*, *M. assamensis*), we expected gene flow between *M. mulatta* and *M. arctoides* to be higher in Ninteract windows than in non-Ninteract windows. For *M. f. aurea*, which carries mitochondrial DNA that is closely related to that of sinica group macaques but whose nuclear genome is more closely related to other populations of *M. fascicularis*, we expected gene flow between *M. f. aurea* and *M. assamensis* (or *M. thibetana*) to be higher in Ninteract windows than in non-Ninteract windows.

Using an approach based on a hidden Markov model (Peter 2020), which is expected to be sensitive to recent introgression, we did not detect evidence for extensive recent introgression between *M. arctoides* and the other species in this dataset (Figs S31, S32). Using this same approach, we also did not detect evidence for extensive recent introgression between *M. f. aurea* and the other species in this dataset, including the other *M. fascicularis* genomes (Figs S33, S34). Thus, these analyses were uninformative with respect to our predictions. Interestingly and unrelated to our predictions concerning Ninteract genes, we detected evidence of gene flow from *M. arctoides* to one of the *M. assamensis* individuals, an observation that has not been previously reported (e.g., Song et al. 2021).

We also explored our expectations using Patterson's D statistic, which potentially could capture signatures of ancient introgression. Consistent with previous inferences, genome-wide introgression patters based on Patterson's D statistic indicate that gene flow between *M. mulatta* and *M. arctoides* is greater than between *M. mulatta* and *M. assamensis* or *M. thibetana*; this statistic also indicates that gene flow between *M. f. aurea* and *M. assamensis* or *M. thibetana* is greater than between *M. fascicularis* and *M. assamensis* or *M. thibetana*. However, no significant excess of gene flow was observed for Ninteract windows compared to non-Ninteract windows for any of these comparisons (Table S11). Taken together, these findings fail to provide evidence that recent introgression had a large effect on mitonuclear interactions in *M. arctoides* or *M. f. aurea*, though we note that this does not rule out the possibility of more subtle effects that we were unable to detect.

1.6 Discussion

At least three non-exclusive mechanisms are thought to contribute to selection on mitonuclear interactions: (i) adaptive divergence, where variation in mitonuclear interactions is tuned to unique local conditions, (ii) compensatory co-adaptation, where changes in the mitochondrial genome lead to selection on Ninteract genes for variants that restore or maintain function, and (iii) intergenomic conflict, where the maternal inheritance of mitochondrial DNA prevents efficient purging of deleterious mutations with male-specific effects or when the nuclear genome must evolve coping mechanisms to curtail selfish replication of mitochondrial DNA (Wolff and Gemmell 2013; Chou and Leu 2015; Havird et al. 2019). In general, the relative contributions of adaptive, neutral, and compensatory processes in sculpting mitonuclear interactions are unclear (Wolff et al. 2014), and we consider the combined effects of these factors here.

Across four genomic datasets from macaques, we found that Ninteract genes are embedded in atypically long ROHs, frequently reside in genomic windows with atypically high F_{ST} and low π and Tajima's D, and frequently have atypically rapid rates of protein evolution compared to non-Ninteract genes. After controlling for effects of gene number, every species or population had an individually significant excess of lower π outliers in Ninteract windows compared to non-Ninteract windows, and almost every comparison had an individually significant excess of upper F_{ST} outliers; across all datasets there was a significant excess of both of these types of outliers.

Polymorphism-based metrics of natural selection (ROHs, F_{ST}, π) were frequently concordant, which suggests they are influenced by similar or the same evolutionary processes, though these polymorphism metrics did not strongly overlap with rapid protein evolution (dN/dS). This disparity could reflect independent selection on regulatory and protein evolution and/or the relatively coarse perspective that our assay of protein evolution provided because the model assumed constraints on protein evolution to be constant across entire proteins and over time. Signatures of natural selection were frequently concordant in independent comparisons involving different sets of species or populations. This included independent comparisons in this study and also comparisons in the *silenus* and Sulawesi group species and populations (discussed further in the Supplement; Evans et al. 2021). Together, these findings argue that mitonuclear interactions of subsets of Ninteract genes from each of the four functional categories (OXPHOS, ARS2, MRP, REP) had a detectable, recurrent, and persistent influence on genome evolution across most or all macaques. This influence likely extends to other papionins that share this social system and perhaps also to other species with similar patterns of sex-biased dispersal.

How might mitonuclear interactions influence macaque societies?

1.6.1 Species diversity

Mitonuclear interactions are proposed to contribute to speciation via epistatic Dobzhansky-Muller interactions between interacting mitochondria-encoded and nuclear-encoded factors (Burton and Barreto 2012; Hill 2017). In macaques, female philopatry and obligate male dispersal causes genetic variation in the autosomes to be more evenly distributed over landscapes than the mitochondria. This could plausibly create conditions that increase the chances of epistatic Dobzhansky-Muller incompatibilities that promote speciation compared to species where both sexes disperse at similar rates. Indeed, macaques are one of the most diverse and widely distributed genera of Old World monkeys (Roos and Zinner 2015), and there is considerable evidence for natural selection on mitonuclear interactions in macaques (this study; Bailey and Stevison 2021; Evans et al. 2021). This raises the possibility that mitonuclear interactions associated with strong female philopatry increase the rate of speciation. However, some primate groups such as gibbons have less sex-biased migration but are also quite diverse. An interesting direction for future work would test in a phylogenetic context whether the rate of speciation is positively correlated with the extent of male-biased dispersal in mammals, or more specifically in primates.

1.6.2 Adaptation

Owing to their central role in energy metabolism, mitonuclear interactions influence essentially all phenotypes of eukaryotes (Rand and Mossman 2020). Macaques inhabit extraordinarily diverse habitats spanning rain forests of Southeast Asia, savanna and scrublands of India, mountainous regions of Morocco, Pakistan, Nepal, and China, areas with seasonal snow in the mountains of Japan, and even urban habitats. They are generally frugivorous, but their flexible diet also may include invertebrates, plants, and meat (Holzner et al. 2019). Might adaptive mitonuclear epistasis account for some component of the ecological breadth of macaque monkeys, wherein certain combinations of mitochondrial and nuclear variation are beneficial in certain contexts? One way to empirically test this is with xenomitochondrial hybrids (Kenyon and Moraes 1997; McKenzie et al. 2003), which could ethically be accomplished using cell culture in macaques. In fruit flies, this approach identified context-dependent tradeoffs that are associated with mitonuclear interactions (Camus et al. 2020), thereby demonstrating that natural selection can favor different mitochondrial variation in different environments, even while holding variation in the nuclear genome constant.

1.6.3 Discordance between mitochondrial and nuclear phylogenies

Many examples exist in macaques where relationships among mitochondrial DNA do not match those among nuclear DNA (mitonuclear phylogenomic discordance; please see Supplement for examples). As discussed above, differences in male and female dispersal contribute to mitonuclear phylogenomic discordance, and compensatory evolution in the nuclear genome may contribute as well. Another non-exclusive possible contributor to mitonuclear phylogenomic discordance is if the deleterious load on the mitochondrial

genome in one population creates opportunity for invasion of a less-loaded mitochondrial genome from a neighboring population or species (Sloan et al. 2017). Added to this, epistatic incompatibilities need not be symmetrical in the sense that a cross between a female from one group and a male from another may produce fit offspring, but not vice versa (Darwin's corollary to Haldane's Rule; Turelli and Moyle 2007). Together these mitogenomic factors likely contribute to dissimilar patterns of introgression of mitochondrial and nuclear variation in macaques, thereby increasing mitonuclear phylogenomic discordance.

1.6.4 Female-biased adult sex ratio

Macaques often have highly skewed female-biased sex ratios (Lindburg 1969; Van Schaik et al. 1989; Beisner et al. 2012) and this is also the case in other papionins (Swedell 2010). One possible contribution to this skew is that deleterious effects associated with mitonuclear incompatibilities disproportionately target males because natural selection is inefficient at removing deleterious mitochondrial mutations that have male-specific or male-biased effects ("mother's curse"; Gemmell et al. 2004; Wolff and Gemmell 2013; Gemmell 2015). This possibility is supported in human by associations between mitochondrial variation and male infertility (Frank and Hurst 1996; Gemmell et al. 2004; Frank 2012). The effects of mother's curse may be mitigated by positive assortative mating by creating circumstances that allow for selection against mitochondrial variation that is only detrimental in males (Hedrick 2012). However, migration of males from their natal groups brings them into contact with females carrying diverged mitochondria creates conditions that more closely match negative assortative mating, which has the opposite effect of increasing mother's curse (Hedrick 2012).

1.6.5 Sexual dimorphism

Macaque species vary extensively in secondary sexually dimorphic characters such as body size, canine length, and pelage (Nunn 1999; Plavcan 2001). Although not generally considered a component of sexual dimorphism, it is notable that there is considerable between species diversity in genital morphology and estrus swelling among macaques (Fooden 1976; Fooden 1989). The genitals of *M. arctoides* are particularly distinguished from other macaques, with males having the largest glans and baculum, and females having distinctions in internal and external morphology of the perineum, cervix, vagina, and position of the urethra (Fooden 1967). Is it a coincidence that the macaque species with the most apomorphic genital morphology also has introgressed mitochondria? An interesting possibility is that intraspecific and interspecific variation in sexual dimorphism could signal mitonuclear compatibility (Hill and Johnson 2013; Hill 2019), or that the expression of sexually dimorphic characteristics signals capacity of the nuclear genome to control selfish mitochondrial replication (Havird et al. 2019). Macaques have intraspecific variation in sexually dimorphic characteristics (Higham et al. 2012; Petersdorf et al. 2017; Rosenfield et al. 2019), but whether and to what degree these features vary with condition (or specifically the efficacy of cellular respiration) is unclear. In fruit flies, variation in mitochondrial genomes has a sex-specific trans-acting effect on

expression of nuclear genes (Mossman et al. 2016), demonstrating that mitochondrial variation can have sex-specific effects on the nuclear genome.

1.6.6 Sex biased dispersal

Dispersal can be a risky undertaking because it exposes individuals to unfamiliar habitats that may increase food insecurity, predation, intraspecific competition, and poses challenges to finding a mate (Bonte et al. 2012), but carries benefits associated with outbreeding, decreasing conspecific density and competition among kin, and location of new resources (Hamilton and May 1977; Matthysen 2005). Our efforts focused on macaques because they have an extreme female philopatry that we expected to drive natural selection on mitonuclear interactions. But it is also possible that sex-biased dispersal is an evolutionary response to mitonuclear interactions or (perhaps even more likely) there could be evolutionary feedback between these phenomena. Scrutiny of mitonuclear interactions in *Papio papio* and *P. hamadryas*, which secondarily lost female philopatry (Kopp et al. 2015; Staedele et al. 2015), and comparison to other papionins may provide further insights into these evolutionary dynamics.

1.7 Conclusions

Our analyses of the most widely distributed non-human primate – *M. mulatta* – and two macaques with introgressed mitochondrial genomes – *M. arctoides* and *M. f. aurea* – identifies signatures of natural selection in polymorphism and protein evolution of Ninteract genes. These findings, coupled with theory and findings from other species, argue that mitonuclear interactions potentially have implications for speciation, adaptation, phylogenomics, demography, and sexual selection in macaques and other species with similar patterns of sex-biased dispersal.

References

Allio, R., Donega, S., Galtier, N., and Nabholz, B. (July 16, 2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution* 34(11), 2762–2772.

Altman, D. and Bland, J. (Aug. 8, 2011). How to obtain the P value from a confidence interval. *BMJ* 343(aug08 1), d2304–d2304.

Bailey, N. and Stevison, L. (Aug. 18, 2021). Mitonuclear conflict in a macaque species exhibiting phylogenomic discordance. *Journal of Evolutionary Biology* 34(10), 1568–1579.

Beisner, B., Jackson, M., Cameron, A., and Mccowan, B. (Mar. 2012). Sex ratio, conflict dynamics, and wounding in rhesus macaques (*Macaca mulatta*). *Applied Animal Behaviour Science* 137(3-4), 137–147.

Bogenhagen, D., Ostermeyer-Fay, A., Haley, J., and Garcia-Diaz, M. (Feb. 2018). Kinetics and mechanism of mammalian mitochondrial ribosome assembly. *Cell Reports* 22(7), 1935–1944.

Bonte, D., Van Dyck, H., Bullock, J., Coulon, A., Delgado, M., Gibbs, M., Lehouck, V., Matthysen, E., Mustin, K., Saastamoinen, M., Schtickzelle, N., Stevens, V., Vandewoestijne, S., Baguette, M., Barton, K., Benton, T., Chaput-Bardy, A., Clobert, J., Dytham, C., Hovestadt, T., Meier, C., Palmer, S., Turlure, C., and Travis, J. (2012). Costs of dispersal. *Biological Reviews* 87(2), 290–312.

Burton, R. and Barreto, F. (Sept. 21, 2012). A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Molecular Ecology* 21(20), 4942–4957.

Caballero, A. (Feb. 1, 1995). On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* 139(2), 1007–1011.

Camus, M., O'leary, M., Reuter, M., and Lane, N. (2020). Impact of mitonuclear interactions on life-history responses to diet. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375(1790), 20190416.

Ceballos, F., Joshi, P., Clark, D., Ramsay, M., and Wilson, J. (Jan. 15, 2018). Runs of homozygosity: Windows into population history and trait architecture. *Nature Reviews Genetics* 19(4), 220–234.

Chou, J.-Y. and Leu, J.-Y. (May 19, 2015). The Red Queen in mitochondria: Cytonuclear co-evolution, hybrid breakdown and human disease. *Frontiers in Genetics* 6, 187.

Clutton-Brock, T. and Lukas, D. (2012). The evolution of social philopatry and dispersal in female mammals. *Molecular Ecology* 21(3), 472–492.

Conway, J., Lex, A., and Gehlenborg, N. (June 22, 2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33(18), 2938–2940.

Danecek, P., Auton, A., Abecasis, G., Albers, C., Banks, E., Depristo, M., Handsaker, R., Lunter, G., Marth, G., Sherry, S., Mcvean, G., and Durbin, R. (June 7, 2011). The variant call format and VCFtools. *Bioinformatics* 27(15), 2156–2158.

Dittus, W. (Dec. 31, 1975). Population Dynamics of the Toque Monkey, *Macaca sinica*. In: *Socioecology and Psychology of Primates*. DE GRUYTER MOUTON, 125–152.

Evans, B., Supriatna, J., and Melnick, D. (Aug. 2001). Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. *Evolution* 55(8), 1686–1702.

Evans, B. and Charlesworth, B. (Feb. 1, 2013). The effect of nonindependent mate pairing on the effective population size. *Genetics* 193(2), 545–556.

Evans, B., Gansauge, M.-T., Tocheri, M., Schillaci, M., Sutikna, T., Jatmiko, Saptomo, E., Klegarth, A., Tosi, A., Melnick, D., and Meyer, M. (Sept. 2020). Mitogenomics of macaques (*Macaca*) across Wallace's Line in the context of modern human dispersals. *Journal of Human Evolution* 146, 102852.

Evans, B., Peter, B., Melnick, D., Andayani, N., Supriatna, J., Zhu, J., and Tosi, A. (Oct. 6, 2021). Mitonuclear interactions and introgression genomics of macaque monkeys (*Macaca*) highlight the influence of behaviour on genome evolution. *Proceedings of the Royal Society B: Biological Sciences* 288(1960), 20211756.

Evans, B., Supriatna, J., Andayani, N., and Melnick, D. (Aug. 2003). Diversification of Sulawesi macaque monkeys: Decoupled evolution of mitochondrial and autosomal DNA. *Evolution* 57(8), 1931–1946.

Fan, Z., Zhou, A., Osada, N., Yu, J., Jiang, J., Li, P., Du, L., Niu, L., Deng, J., Xu, H., Xing, J., Yue, B., and Li, J. (Oct. 2018). Ancient hybridization and admixture in macaques (genus *Macaca*) inferred from whole genome sequences. *Molecular Phylogenetics and Evolution* 127, 376–386.

Fay, J. and Wu, C.-I. (July 1, 1999). A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Molecular Biology and Evolution* 16(7), 1003–1005.

Fooden, J. (May 1967). Complementary specialization of male and female reproductive structures in the bear macaque, *Macaca arctoides*. *Nature* 214(5091), 939–941.

Fooden, J. (1976). rovisional classification and key to living species of macaques (primates: *Macaca*). *Folia Primatologica* 25(2-3), 225–236.

Fooden, J. (1989). Malaria in macaques. *International Journal of Primatology* 15(4). Ed. by S. S. P. K. Seth, 573–596.

Fooden, J. (2000). Systematic review of the rhesus macaque, *Macaca mulatta* (Zimmermann, 1780) / Jack Fooden. *Fieldiana Zoology* 96, 1–180.

Frank, S. and Hurst, L. (Sept. 1996). Mitochondria and male disease. *Nature* 383(6597), 224–224.

Frank, S. (Sept. 2012). Evolution: Mitochondrial burden on male health. *Current Biology* 22(18), R797–R799.

Fu, Q., Mittnik, A., Johnson, P., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A., Martini, F., Cremonesi, R., Svoboda,

J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Pääbo, S., and Krause, J. (Apr. 2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology* 23(7), 553–559.

Gaspari, M., Larsson, N.-G., and Gustafsson, C. (Dec. 2004). The transcription machinery in mammalian mitochondria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1659(2-3), 148–152.

Gemmell, N. (2015). Why are males dying younger? Evolutionary and functional aspects of maternal mitochondrial inheritance and mitochondria in ageing. *Human Reproduction (Oxford)* 30, 38.

Gemmell, N., Metcalf, V., and Allendorf, F. (May 2004). Mother's curse: The effect of mtDNA on individual fitness and population viability. *Trends in Ecology & Evolution* 19(5), 238–244.

Hamilton, W. and May, R. (May 1977). Dispersal in stable habitats. *Nature* 269(5629), 578–581.

Havird, J., Forsythe, E., Williams, A., Werren, J., Dowling, D., and Sloan, D. (2019). Selfish Mitonuclear Conflict. *Current Biology* 29, R496–R511.

Hedrick, P. (2012). Reversing mother's curse revisited. *Evolution* 66(2), 612–616.

Higham, J., Heistermann, M., Saggau, C., Agil, M., Perwitasari-Farajallah, D., and Engelhardt, A. (2012). Sexual signalling in female crested macaques and the evolution of primate fertility signals. *BMC Evolutionary Biology* 12(1), 89.

Hill, G. (Apr. 2017). The mitonuclear compatibility species concept. *The Auk* 134(2), 393–409.

Hill, G. (2019). Mitonuclear ecology. *Molecular Biology and Evolution* 32(8), 1917–1927.

Hill, G. (June 2020). Mitonuclear compensatory coevolution. *Trends in Genetics* 36(6), 403–414.

Hill, G. and Johnson, J. (Oct. 7, 2013). The mitonuclear compatibility hypothesis of sexual selection. *Proceedings of the Royal Society B: Biological Sciences* 280(1768), 20131314.

Hoelzer, G. (Apr. 1997). Inferring phylogenies from mtdna variation: Mitochondrial-gene trees versus nuclear-gene trees revisited. *Evolution* 51(2), 622–626.

Holzner, A., Ruppert, N., Swat, F., Schmidt, M., Weiß, B., Villa, G., Mansor, A., Mohd Sah, S., Engelhardt, A., Kühl, H., and Widdig, A. (Oct. 2019). Macaques can contribute to greener practices in oil palm plantations when used as biological pest control. *Current Biology* 29(20), R1066–R1067.

Hudson, R., Slatkin, M., and Maddison, W. (Oct. 1, 1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* 132(2), 583–589.

Kenyon, L. and Moraes, C. (Aug. 19, 1997). Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids. *Proceedings of the National Academy of Sciences* 94(17), 9131–9135.

Kopp, G., Fischer, J., Patzelt, A., Roos, C., and Zinner, D. (Apr. 10, 2015). Population genetic insights into the social organization of Guinea baboons (*Papio papio*): Evidence for female-biased dispersal. *American Journal of Primatology* 77(8), 878–889.

Lane, N. (May 1, 2014). Bioenergetic Constraints on the Evolution of Complex Life. *Cold Spring Harbor Perspectives in Biology* 6(5), a015982–a015982.

Lindburg, D. (Nov. 28, 1969). Rhesus monkeys: Mating season mobility of adult males. *Science* 166(3909), 1176–1178.

Liu, Z., Tan, X., Orozco-Terwengel, P., Zhou, X., Zhang, L., Tian, S., Yan, Z., Xu, H., Ren, B., Zhang, P., Xiang, Z., Sun, B., Roos, C., Bruford, M., and Li, M. (Aug. 27, 2018). Population genomics of wild Chinese rhesus macaques reveals a dynamic demographic history and local adaptation, with implications for biomedical research. *GigaScience* 7(9), 1–14.

Martin, W. and Koonin, E. (Mar. 2006). Introns and the origin of nucleus–cytosol compartmentalization. *Nature* 440(7080), 41–45.

Matsudaira, K., Hamada, Y., Bunlungsup, S., Ishida, T., San, A., and 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. *Journal of Heredity* 109(4), 360–371.

Matthysen, E. (June 2005). Density-dependent dispersal in birds and mammals. *Ecography* 28(3), 403–416.

Mckenzie, M., Chiotis, M., Pinkert, C., and Trounce, I. (Apr. 25, 2003). Functional respiratory chain analyses in murid xenomitochondrial cybrids expose coevolutionary constraints of cytochrome b and nuclear subunits of complex III. *Molecular Biology and Evolution* 20(7), 1117–1124.

Melnick, D., Hoelzer, G., Absher, R., and Ashley, M. (1993). MtDNA diversity in rhesus monkeys reveals overestimates of divergence time and paraphyly with neighboring species. *Molecular Biology and Evolution* 10, 282–295.

Melnick, D. and Hoelzer, G. (Aug. 1992). Differences in male and female macaque dispersal lead to contrasting distributions of nuclear and mitochondrial DNA variation. *International Journal of Primatology* 13(4), 379–393.

Mossman, J., Tross, J., Li, N., Wu, Z., and Rand, D. (Oct. 1, 2016). Mitochondrial-Nuclear interactions mediate sex-specific transcriptional profiles in *Drosophila*. *Genetics* 204(2), 613–630.

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (Jan. 30, 2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32(11), 1749–1751.

Nei, M. (Dec. 31, 1987). *Molecular evolutionary genetics*. New York.

Nunn, C. (Aug. 1999). The evolution of exaggerated sexual swellings in primates and the graded-signal hypothesis. *Animal Behaviour* 58(2), 229–246.

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'hara, R., Simpson, G., Solymos, P., Henry, M., Stevens, H., and Wagner, H. (2013). *Vegan: Community Ecology Package. R package version 2.0-7*.

Ortiz, E. (2019). *0: Convert a VCF matrix into several matrix formats for phylogenetic analysis*.

Osada, N. and Akashi, H. (2012). Mitochondrial–nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome c oxidase complex. *Molecular Biology and Evolution* 29(1), 337–346.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (Nov. 1, 2012). Ancient admixture in human history. *Genetics* 192(3), 1065–1093.

Pemberton, T., Absher, D., Feldman, M., Myers, R., Rosenberg, N., and Li, J. (Aug. 2012). Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics* 91(2), 275–292.

Peter, B. (Mar. 15, 2020). 100,000 years of gene flow between Neandertals and Denisovans in the Altai mountains. *bioRxiv* 100.

Petersdorf, M., Dubuc, C., Georgiev, A., Winters, S., and Higham, J. (Sept. 11, 2017). Is male rhesus macaque facial coloration under intrasexual selection? *Behavioral Ecology* 28(6), 1472–1481.

Plavcan, J. (2001). Sexual dimorphism in primate evolution. *American Journal of Physical Anthropology* 116(S33), 25–53.

Pusey, A. (1987). The primate perspective on dispersal. In: *Animal Dispersal*. Springer Netherlands, 243–259.

Radzvilavicius, A., Hadjivasiliou, Z., Lane, N., and Pomiankowski, A. (2016). Selection for mitochondrial quality drives evolution of the germline. *PLoS Biology* 14, e2000410.

Rand, D., Haney, R., and Fry, A. (Dec. 2004). Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology & Evolution* 19(12), 645–653.

Rand, D. and Mossman, J. (2020). Mitonuclear conflict and cooperation govern the integration of genotypes, phenotypes and environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375(1790), 20190188.

Roos, C., Knauf, S., Chuma, I., Maille, A., Callou, C., Sabin, R., Portela Miguez, R., and Zinner, D. (Nov. 27, 2020). New mitogenomic lineages in *Papio* baboons and their phylogeographic implications. *American Journal of Physical Anthropology* 174(3), 407–417.

Roos, C., Kothe, M., Alba, D., Delson, E., and Zinner, D. (Aug. 2019). The radiation of macaques out of Africa: Evidence from mitogenome divergence times and the fossil record. *Journal of Human Evolution* 133, 114–132.

Roos, C. and Zinner, D. (2015). Diversity and evolutionary history of macaques with special focus on *Macaca mulatta* and *Macaca fascicularis*. In: *The Nonhuman Primate in Nonclinical Drug Development and Safety Assessment*. Ed. by J. Blmel, S. Korte, E. Schenck, and G. F. Weinbauer. Amsterdam: Elsevier, 3–16.

Rosenfield, K., Semple, S., Georgiev, A., Maestripieri, D., Higham, J., and Dubuc, C. (2019). Experimental evidence that female rhesus macaques (*Macaca mulatta*) perceive variation in male facial masculinity. *Royal Society Open Science* 6, 181415.

Signes, A. and Fernandez-Vizarra, E. (July 20, 2018). Assembly of mammalian oxidative phosphorylation complexes I–V and supercomplexes. *Essays in Biochemistry* 62(3), 255–270.

Sissler, M., González-Serrano, L., and Westhof, E. (Aug. 2017). Recent advances in mitochondrial aminoacyl-tRNA synthetases and disease. *Trends in Molecular Medicine* 23(8), 693–708.

Sloan, D., Havird, J., and Sharbrough, J. (Jan. 27, 2017). The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology* 26(8), 2212–2236.

Smith, D. and Mcdonough, J. (Jan. 2005). Mitochondrial DNA variation in Chinese and Indian rhesus macaques (*Macaca mulatta*). *American Journal of Primatology* 65(1), 1–25.

Song, Y., Jiang, C., Li, K.-H., Li, J., Qiu, H., Price, M., Fan, Z.-X., and Li, J. (2021). Genome-wide analysis reveals signatures of complex introgressive gene flow in macaques (genus *Macaca*). *Zoological Research* 42(4), 433–449.

Southwick, C. and Siddiqi, M. (1996). Population status of nonhuman primates in Asia, with emphasis on rhesus macaques in India. *American Journal of Primatology* 34(1). Ed. by J. E. Fa and D. G. Lindburg, 51–59.

Städele, V., Van Doren, V., Pines, M., Swedell, L., and Vigilant, L. (Jan. 2015). Fine-scale genetic assessment of sex-specific dispersal patterns in a multilevel primate society. *Journal of Human Evolution* 78, 103–113.

Swedell, L. (2010). African papionins: Diversity of social organization and ecological flexibility. *Primates in Perspective*.

Swofford, D. and Sullivan, J. (2002). Phylogeny inference based on parsimony and other methods using PAUP. In: *The Phylogenetic Handbook*. Sunderland: Cambridge University Press, 267–312.

Tajima, F. (Nov. 1, 1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3), 585–595.

Team, R. C. (2018). *Package 'Stats'. the R stats package*.

Telfer, P., Souquière, S., Clifford, S., Abernethy, K., Bruford, M., Disotell, T., Sterner, K., Roques, P., Marx, P., and Wickings, E. (June 2, 2003). Molecular evidence for deep phylogenetic divergence in *Mandrillus sphinx*. *Molecular Ecology* 12(7), 2019–2024.

Tishkoff, S., Reed, F., Ranciaro, A., Voight, B., Babbitt, C., Silverman, J., Powell, K., Mortensen, H., Hirbo, J., Osman, M., Ibrahim, M., Omar, S., Lema, G., Nyambo, T., Ghori, J., Bumpstead, S., Pritchard, J., Wray, G., and Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* 39(1), 31–40.

Tosi, A., Morales, J., and Melnick, D. (Nov. 2000). Comparison of Y chromosome and mtDNA phylogenies leads to unique inferences of macaque evolutionary history. *Molecular Phylogenetics and Evolution* 17(2), 133–144.

Turelli, M. and Moyle, L. (June 1, 2007). Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* 176(2), 1059–1088.

Van Schaik, C., Netto, W., Van Amerongen, A., and Westland, H. (1989). Social rank and sex ratio of captive long-tailed macaque females (*Macaca fascicularis*). *American Journal of Primatology* 19(3), 147–161.

Wallace, C. and Burren, O. (2017). *Package 'genomic.autocorr': Models dealing with spatial dependency in genomic data*.

Warren, W., Harris, R., Haukness, M., Fiddes, I., Murali, S., Fernandes, J., Dishuck, P., Storer, J., Raveendran, M., Hillier, L., Porubsky, D., Mao, Y., Gordon, D., Vollger,

M., Lewis, A., Munson, K., Devogelaere, E., Armstrong, J., Diekhans, M., Walker, J., Tomlinson, C., Graves-Lindsay, T., Kremitzki, M., Salama, S., Audano, P., Escalona, M., Maurer, N., Antonacci, F., Mercuri, L., Maggiolini, F., Catacchio, C., Underwood, J., O'connor, D., Sanders, A., Korbel, J., Ferguson, B., Kubisch, H., Picker, L., Kalin, N., Rosene, D., Levine, J., Abbott, D., Gray, S., Sanchez, M., Kovacs-Balint, Z., Kemnitz, J., Thomasy, S., Roberts, J., Kinnally, E., Capitanio, J., Skene, J., Platt, M., Cole, S., Green, R., Ventura, M., Wiseman, R., Paten, B., Batzer, M., Rogers, J., and Eichler, E. (Dec. 18, 2020). Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. *Science* 370(6523).

Wolff, J. and Gemmell, N. (2013). Mitochondria, maternal inheritance, and asymmetric fitness: Why males die younger. *BioEssays* 35(2), 93–99.

Wolff, J., Ladoukakis, E., Enríquez, J., and Dowling, D. (July 5, 2014). Mitonuclear interactions: Evolutionary consequences over multiple biological scales. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1646), 20130443.

Wolfheim, J. (1983). Primates of the world: Distribution, abundance, and conservation. *International Journal of Primatology* 6(3), 323–325.

Yang, Z. (Apr. 18, 2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24(8), 1586–1591.

Yao, L., Li, H., Martin, R., Moreau, C., and Malhi, R. (Nov. 2017). Tracing the phylogeographic history of Southeast Asian long-tailed macaques through mitogenomes of museum specimens. *Molecular Phylogenetics and Evolution* 116, 227–238.

Zinner, D., Fickenscher, G., and Roos, C. (2013). Family cercopithecidae (old world monkeys). In: *The handbook of the mammals of the world*. Ed. by A. B. R. R. A. Mittermeier and D. E. Wilson. Barcelona, 550–627.

Appendix A

Chapter 1 Supplement

A1 Supplementary Introduction

Macaque dispersal. In mammals, dispersal is often sex-biased, with the drivers of sex-specific dispersal thought to be influenced by factors such as social system and spatiotemporal variation in sex-specific resources (Hamilton and May 1977; Li and Kokko 2019). Many mammals have male-biased dispersal (Greenwood 1980; Dobson 1982) but with less male-bias than most papionin monkeys (Thierry 2007; Clutton-Brock and Lukas 2012). The extreme male-biased dispersal in most macaque monkeys is due to unusually strong female philopatry (as opposed to unusually long-distance male dispersal) because females generally stay in their natal group for their entire life (Dittus 1975; Pusey 1987; Swedell 2010; Clutton-Brock and Lukas 2012; Fischer et al. 2019).

A2 Supplementary Methods

A2.1 Data

For the captive rhesus from India, we selected a subset of nine individuals randomly from each of eight captive colonies; for a ninth colony (WNPRC) we analyzed 18 individuals including nine randomly selected individuals from each of two batches of samples from this center, as detailed in (Warren et al. 2020). For one captive rhesus from India, a mitochondrial genome was assembled from raw reads, but genotypic data were lacking, leaving a total of 89 individual captive rhesus from India for which autosomal data was analyzed. For the *M. arctoides* dataset, we excluded data from three individuals from the *silenus* species group to focus our findings on the more closely related species and so that our results would be evolutionarily independent from (Evans et al. 2021). This left ten individuals in the *M. arctoides* dataset, including three *M. arctoides* (SRS6488501, SRS1196878 (SRR2981139), SRS1196879 (SRR2981140) – provenances not reported), two *M. assamensis* (SRS6491954, SRR2981114 – provenance not reported), one *M. thibetana* (SRR1024051 – Sichuan, China), one *M. fascicularis* (SRS117874 - Vietnam), and three *M. mulatta* (SRS115022 – SW China, SRS212016 – SW China, SRS114988 – provenance not reported) individuals. For the *M. f. aurea* dataset, genomic data from two *M. f. aurea* individuals (DRS139837, DRS139838) were compared to genomic data

from three *M. f. fascicularis* individuals (DRR219371, SRA023855, SRR1564766, from Thailand, Vietnam, and Mauritius respectively), one *M. assamensis* (SRR2981114) and one *M. thibetana* (SRR1024051) individual. For this dataset, data were trimmed, aligned to version 10 of the rhesus genome, de-duplicated, genotyped and filtered as described elsewhere (Evans et al. 2021).

A2.2 Ninteract genes

More than 500 nuclear-encoded proteins are imported into the mitochondrial organelle (Mootha et al. 2003; Sickmann et al. 2003) that participate in processes discussed in the main text (OXPHOS, ARS2, MRP, REP) but also a spectrum of other functions, e.g. mitochondrial membrane biogenesis and maintenance, nucleotide synthesis and transport, cytosolic chaperones, mitochondrial mRNA degradation and apoptosis (Smits et al. 2010). We focused here on a subset of these proteins that directly interact with mitochondrial DNA, RNA, or protein because we expected the strongest signature of natural selection on mitonuclear interactions involving these proteins. OXPHOS complex II genes were not considered to be Ninteract genes because this complex does not contain mitochondria-encoded proteins. Only two Ninteract genes occur on the sex chromosomes (both are OXPHOS genes on the X chromosome) and these were excluded from analysis (see main text).

A2.3 Mitochondrial phylogenomics

De novo assembly of mitochondrial genomes was performed using NOVOplasty version 4.3.1 (Dierckxsens et al. 2017). Many genomes were successfully assembled using a kmer size of 29, but we also attempted higher (33, 38) and lower (24) values if the initial assembly was unsuccessful. We included one sample from the captive rhesus from India (MMUL.IN-32510) where we assembled a mitochondrial genome even though there was no genotype data available; this individual was included in the phylogenetic analysis of the mitochondrial genomes but not in the population genetic and molecular evolutionary analyses of autosomal data. For four captive rhesus individuals from India (MMUL.IN-39345, MMUL.IN-38591, MMUL.IN-32510, MMUL.IN-35091), assembly of complete mitochondrial genomes was only achieved after we first mapped raw data to a mitochondrial reference genome (KJ567053.1, a *M. mulatta* from India) and the performed de novo assembly using NOVOplasty with only the mapped reads. We were unable to assemble the mitochondrial genome from one *M. fascicularis* individual (SRR1564766, from Mauritius) and we instead assumed that the phylogenetic placement of this individual's mitochondrial genome was the same as a GenBank accession of another longtail macaque from Mauritius (KM851000.1). Similarly, because raw data from the *M. arctoides* dataset were not available at the time of analysis, we included publicly available mitochondrial genomes from each species and geographic region represented in this dataset instead of de novo assembled genomes.

Evolutionary relationships among the mitochondrial genomes were estimated using Bayesian and maximum likelihood approaches using the software BEAST version 2.6.3

(Bouckaert et al. 2019) and Iqtree version 1.6.12 (Nguyen et al. 2015) respectively. For the BEAST analysis, the model of evolution, calibration times, and priors were the same as a previous study of mitogenomics (Evans et al. 2020). Ten independent runs were performed, each with a chain length of 43.3–73.2 million generations and sampling parameters every 1000 generations; a burn-in of 50% of each chain was discarded. Tracer was used to evaluate the effective sample sizes of each parameter after discarding the burn-in; all values were greater than 100 and most were greater than 200. The Logcombiner program from the BEAST package was used to combine the post-burn-in posterior distribution of trees and further thin the trees to sample every 5000 generations. Then the Treeannotator program from the BEAST package was used to generate a maximum clade credibility tree. For the maximum likelihood analysis, the TIM2+F+I+G4 model was used based on the Bayesian Information Criterion as implemented by Iqtree; confidence of nodes was evaluated using the ultrafast bootstrap approach (Minh et al. 2013). For both analyses, we excluded a portion of the alignment where the assembly of five wild rhesus samples (C_rhe_28, 29, 30, 32, and 34) had an insertion with homology to the COX2 gene because we were unsure of the validity of this portion of the alignment. Because major topographical relationships were identical in both analyses, we present only the Bayesian consensus tree.

A3 Supplementary Results

A3.1 Mitochondrial phylogeny

The estimated mitochondrial phylogeny is consistent with previous inferences (Roos et al. 2019) that the mitochondrial genome of *M. arctoides* is most closely related to a clade containing mitochondrial lineages carried by *M. mulatta*, *M. cyclopis*, and *M. fuscata*. The phylogeny also supports previous inferences (Matsudaira et al. 2018) that there was an ancient introgression of mitochondrial genomes to an ancestor of *M. f. aurea* from an ancestor of *M. assamensis* and *M. thibetana*. As previously (Evans et al. 2020), the inclusion of data from *M. sinica* indicates that this introgression event occurred between an ancestor of *M. sinica* after divergence from the ancestor of *M. assamensis* and *M. thibetana*.

For the captive rhesus macaques, animals that carry the orange clade are derived from ten institutes; sample sizes follow institutional abbreviation from (Warren et al. 2020): (JC_UoC: 6, JH_CPRC: 9, EV_NEPRC: 9, SKDS_CNPRC: 9, ZJ_YNPRC: 6, LPVGTI_OHCU: 9, RWDO_WNPRC: 9, MK_TNPRC: 9, BF_ONPRC: 8, RWNK _WNPRC: 6); animals that carry the brown clade are derived from four institutes (JC_UoC: 3, ZJ_YNPRC: 3, BF_ONPRC: 1, RWNK_WNPRC: 1), and the one individual that carried the red clade came from RWNK_WNPRC.

An unexpected inference came from the mitochondrial genome of a *M. fascicularis* individual from Vietnam (DRR219371) from the *M. f. aurea* dataset, which was highly diverged from the other two widespread mitochondrial lineages in this species (Fig. 1). To our knowledge this diverged mitochondrial lineage has not been previously reported.

Also unexpected was the nested position of the *M. sylvanus* lineage relative to the silenus group. However, this relationship had lower support in the Bayesian (Fig. 1) and maximum likelihood analysis (bootstrap support of 90%) compared to nodes subtending other major lineages (all had posterior probability and bootstrap support of 100%). Most other analyses suggest that mitochondrial DNA from *M. sylvanus* is sister to that of all other extant macaque species (Roos et al. 2019; Evans et al. 2020).

A3.2 ROH analyses

For each population or species in each dataset, portions of density plots of ROHs with Ninteract genes are right shifted compared to ROHs with only other genes and compared to ROHs with no genes (Figs. S1–4).

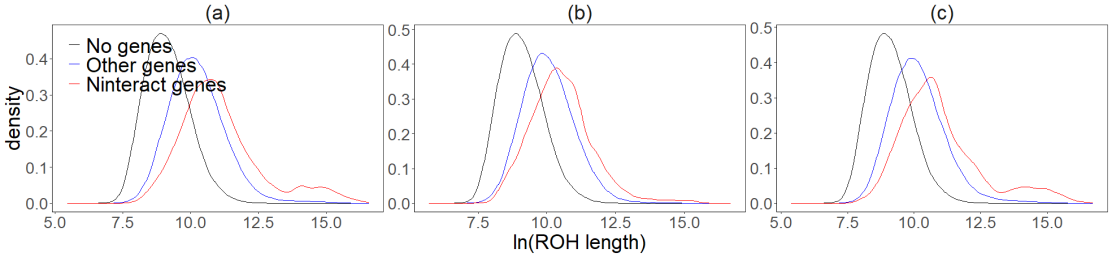

FIG. S1. Density plot of ROHs of wild *M. mulatta* from China illustrates longer length of Ninteract ROHs (red) compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S1.

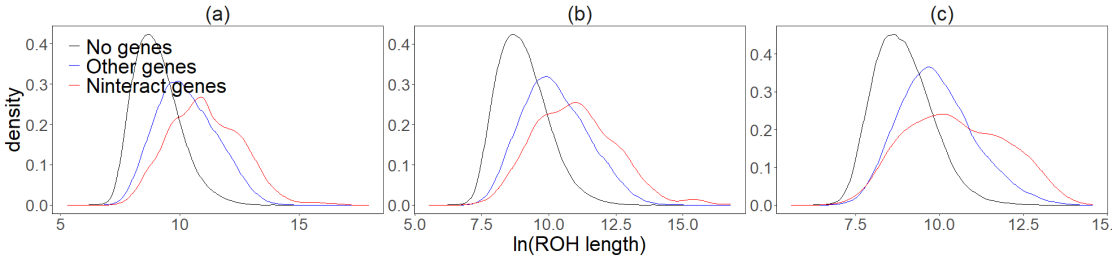

FIG. S2. Density plot of ROHs of captive *M. mulatta* from India illustrates longer length of Ninteract ROHs (red) compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S2.

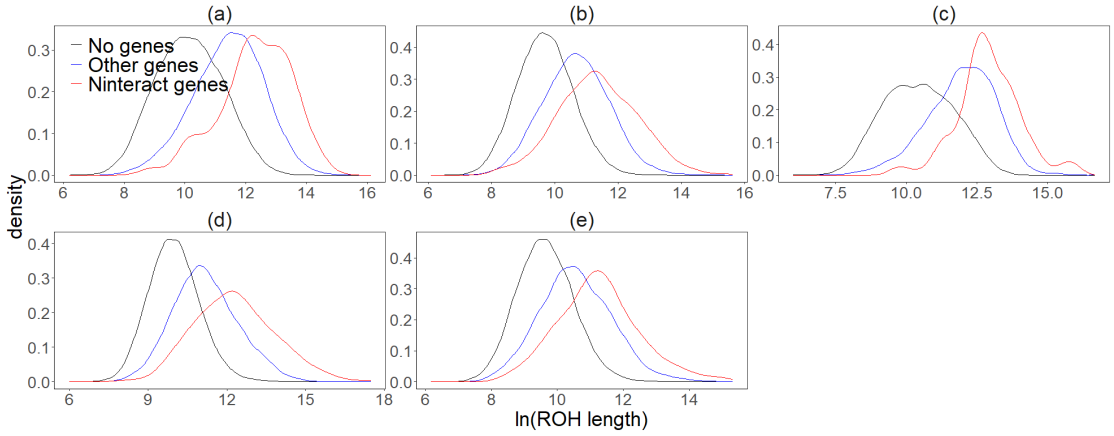

FIG. S3. Density plot of ROHs of *M. arctoides* dataset illustrates longer length of Ninteract ROHs (red) compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S3.

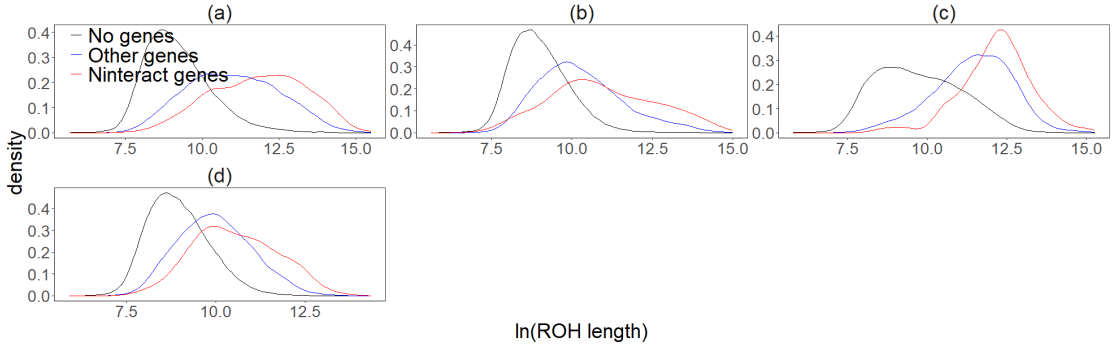

FIG. S4. Density plot of ROHs of *M. f. aurea* dataset illustrates longer length of Ninteract ROHs (red) compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S4.

A3.3 F_{ST} and π in 100kb windows

With exceptions discussed in the main text, linear models fitted to data from 100kb windows (Figs. S5, S6) recovered similar trends as those fitted to data from 30kb windows (Figs. 3, 4) in that F_{ST} predicted by the linear fit was generally higher and π predicted by the linear fit was generally lower in Ninteract compared to non-Ninteract windows. Consistent with this, portions of density plots of F_{ST} with Ninteract windows are right shifted compared to non-Ninteract windows and compared to windows with no genes (Figs. S7–14).

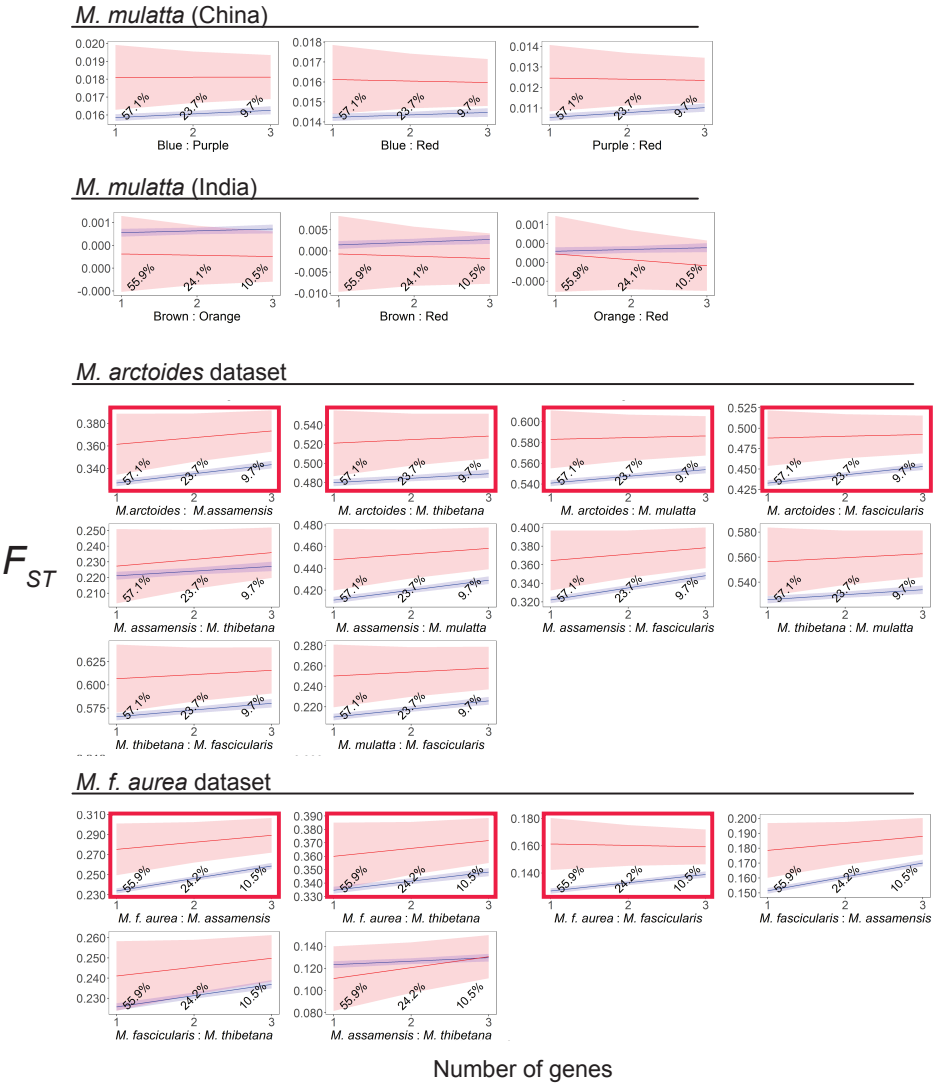

FIG. S5. In three of the four datasets (all but *M. mulatta* from India), the predicted values (marginal means) for most species or population of F_{ST} in 100 kb windows are generally higher in Ninteract windows (red) compared non-Ninteract windows (blue) across biologically relevant numbers of genes in each window (Number of genes). Plotting, shading, and highlights follow Fig. 3.

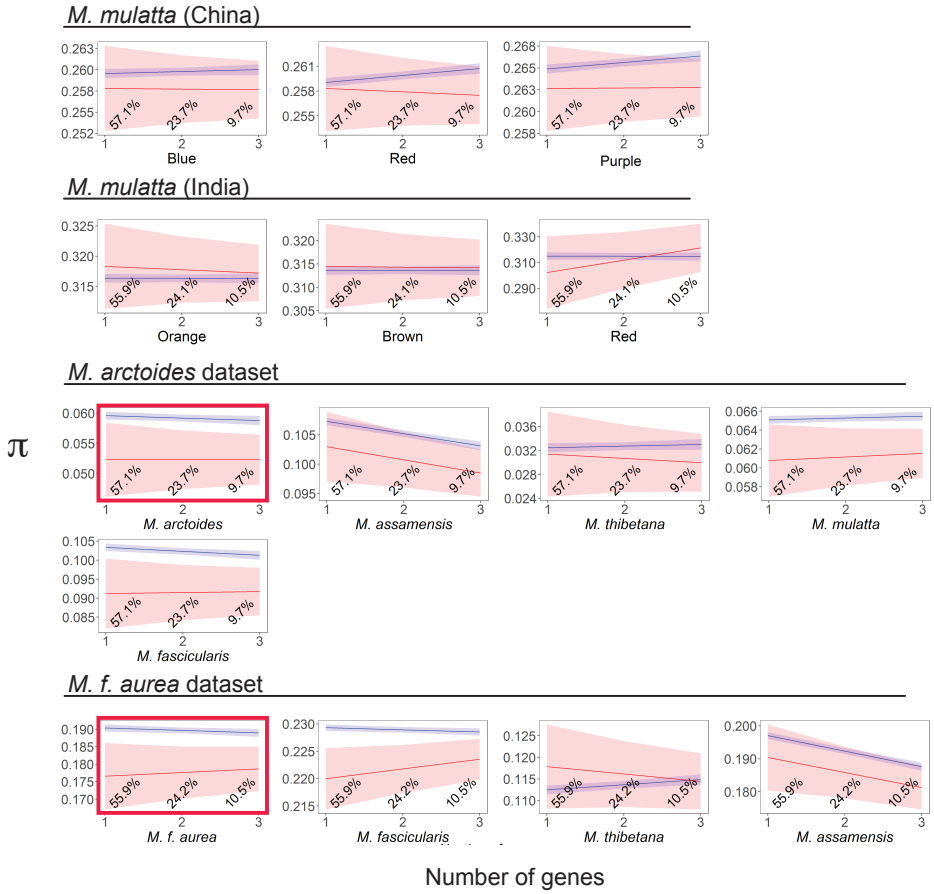

FIG. S6. In the *M. arctoides* and *M. f. aurea* datasets but in the *M. mulatta* from China and India datasets the predicted values (marginal means) for most species or population of π in 100 kb windows are generally lower in Ninteract windows (red) compared non-Ninteract windows (blue) across biologically relevant numbers of genes in each window (Number of genes). Plotting, shading, and highlights follow Fig. 4.

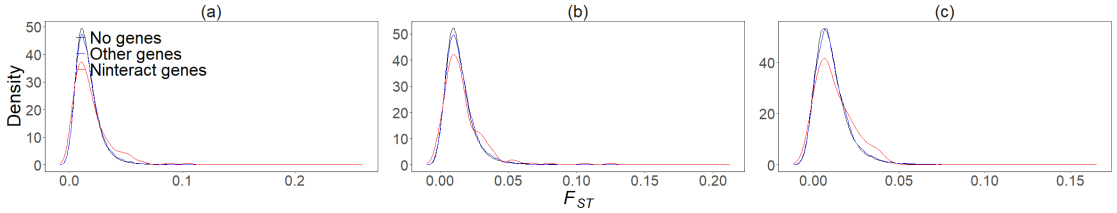

FIG. S7. Density plot of F_{ST} of wild *M. mulatta* from China for 30kb windows suggests some Ninteract windows (red) have atypically high F_{ST} compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S1.

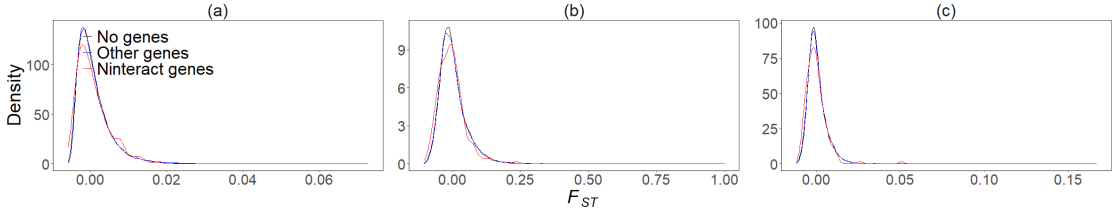

FIG. S8. Density plot of F_{ST} of captive *M. mulatta* from India for 30kb windows does not suggest that some Ninteract windows (red) have atypically high F_{ST} compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S2.

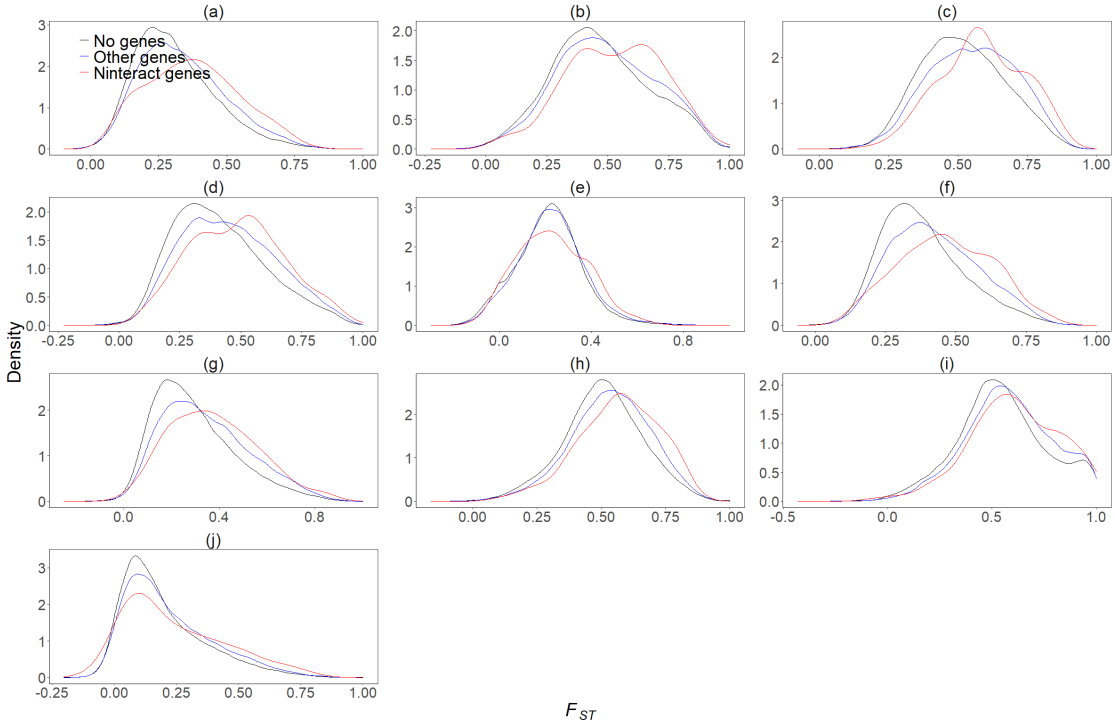

FIG. S9. Density plot of F_{ST} of *M. arctoides* dataset for 30kb windows suggests some Ninteract windows (red) have atypically high F_{ST} compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S3.

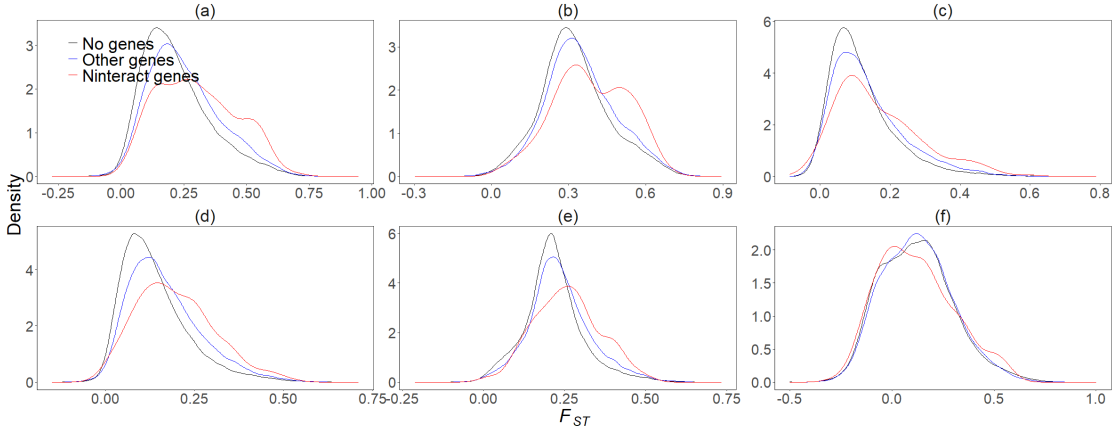

FIG. S10. Density plot of F_{ST} of *M. f. aurea* dataset for 30kb windows suggests some Ninteract windows (red) have atypically high F_{ST} compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S4.

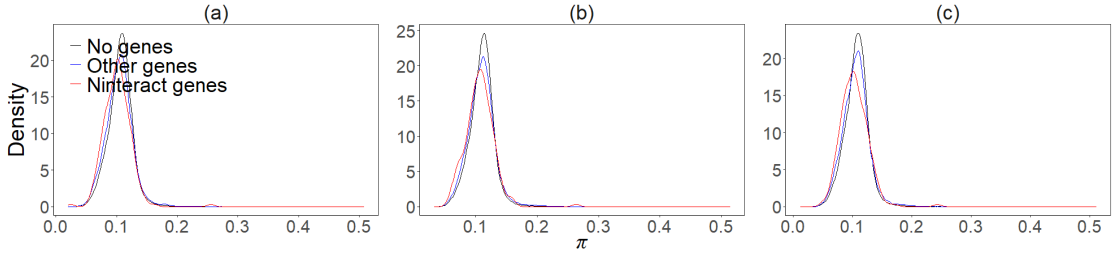

FIG. S11. Density plot of π for the wild *M. mulatta* from China for 30kb windows suggests some Ninteract windows (red) have atypically low π compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S1.

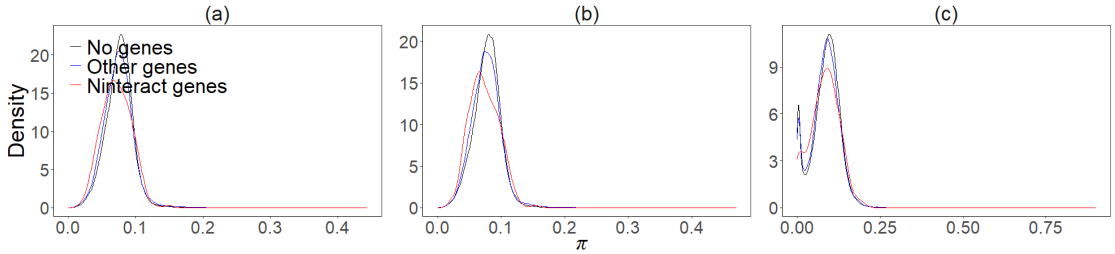

FIG. S12. Density plot of π for the captive *M. mulatta* from India for 30kb windows suggests some Ninteract windows (red) have atypically low π compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S2.

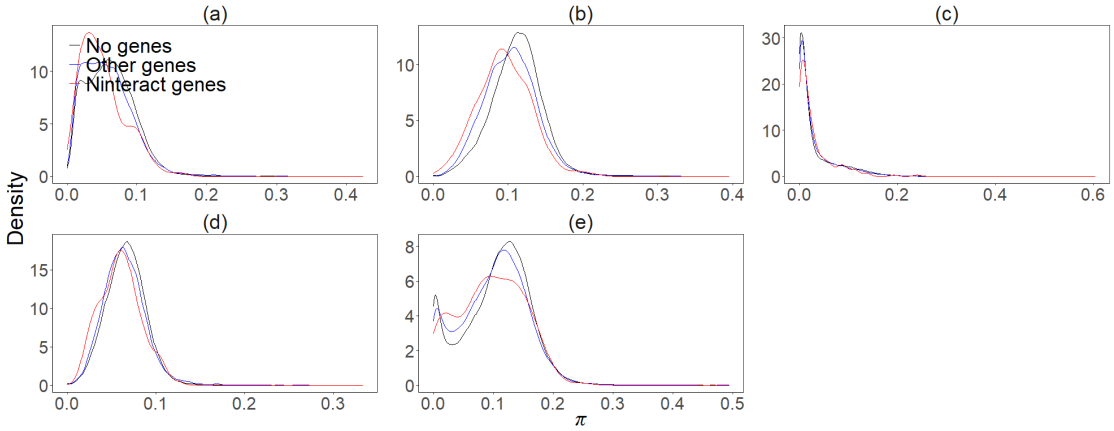

FIG. S13. Density plot of π for the *M. arctoides* dataset for 30kb windows suggests some Ninteract windows (red) have atypically low π compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S3.

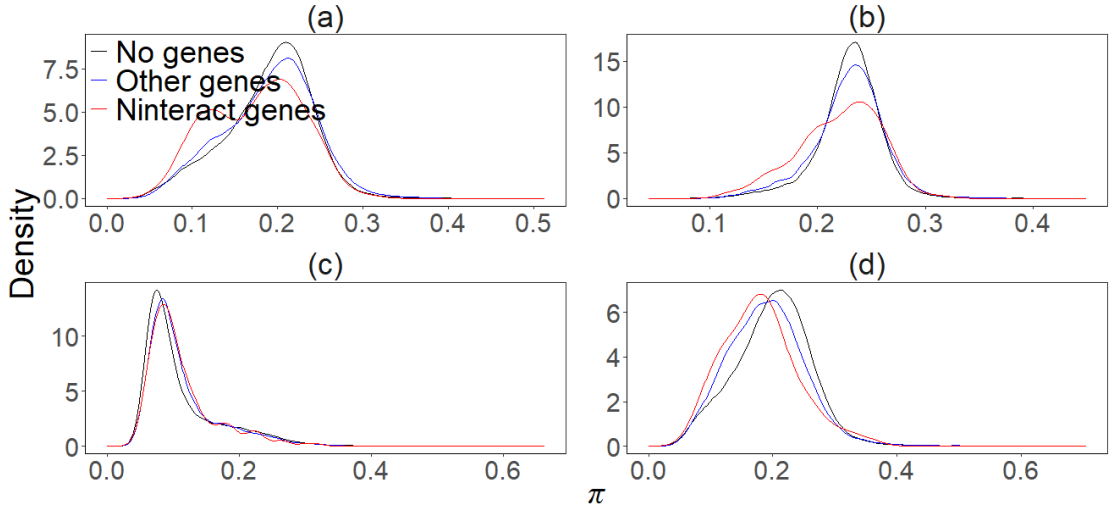

FIG. S14. Density plot of π for the *M. f. aurea* dataset for 30kb windows suggests some Ninteract windows (red) have atypically low π compared to ROHs that contain only other genes (blue) or no genes (black). The order of panels matches Fig. S4.

A3.4 F_{ST} outlier analyses: wild macaques from China

Outlier analysis (Fig. S16) also identified an excess of Ninteract windows that were upper F_{ST} outliers (comparisons between Brown and Orange: 4.8%, between Brown and Red: 4.2%, and between Orange and Red: 4.8%) compared to non-Ninteract windows (comparisons between Brown and Orange: 3.5%, between Brown and Red: 3.3%, and between Orange and Red: 2.8%). However, this excess was not individually significant for these comparisons ($P > 0.05$; binomial tests).

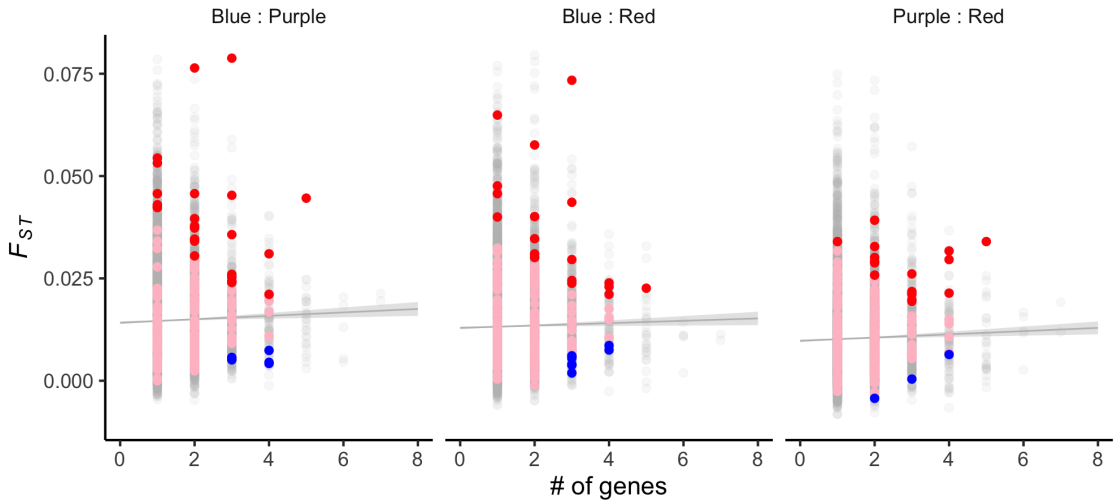

FIG. S15. F_{ST} outlier analysis for wild *M. mulatta* from China identifies a significant excess of upper F_{ST} outliers in Ninteract windows (red dots above fitted line) for all three pairwise comparisons. Blue dots indicate Ninteract windows with atypically low differentiation in each pairwise comparison.

A3.5 F_{ST} outlier analysis: captive *M. mulatta* from India

Outlier analysis (Fig. S16) also identified an excess of Ninteract windows that were upper FST outliers (comparisons between Brown and Orange: 4.8%, between Brown and Red: 4.2%, and between Orange and Red: 4.8%) compared to non-Ninteract windows (comparisons between Brown and Orange: 3.5%, between Brown and Red: 3.3%, and between Orange and Red: 2.8%). However, this excess was not individually significant for these comparisons ($P > 0.05$; binomial tests).

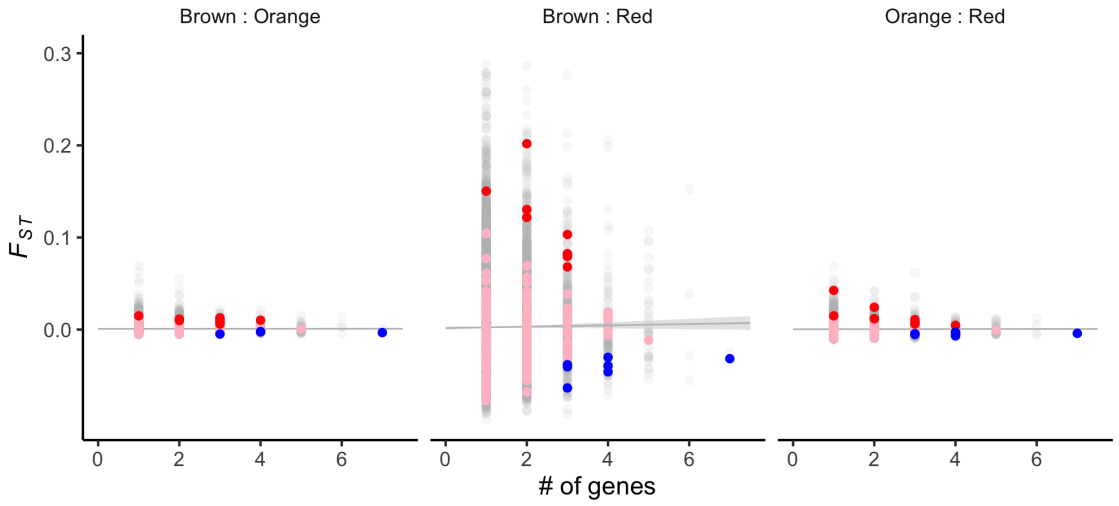

FIG. S16. F_{ST} outlier analysis of captive *M. mulatta* from India data identifies a significant excess of upper F_{ST} outliers in Ninteract windows (red dots above fitted line) in one of three comparisons (Orange : Red). Blue dots indicate Ninteract windows with atypically low differentiation in each pairwise comparison.

A3.6 F_{ST} outlier analysis: M. arctoides dataset

For the *M. arctoides* dataset, ten pairwise comparisons were considered between five populations/species including two (*M. thibetana* and *M. fascicularis*) that had only one individual. Outlier analysis (Fig. S17) identified a significant excess of Ninteract windows that were upper F_{ST} outliers (comparisons between *M. arctoides* and *M. assamensis*: 7.7%, between *M. arctoides* and *M. fascicularis*: 7.7%, between *M. arctoides* and *M. mulatta*: 5.7%, between *M. arctoides* and *M. thibetana*: 4.6%, between *M. assamensis* and *M. fascicularis*: 6.7%, between *M. assamensis* and *M. mulatta*: 8.2%, between *M. assamensis* and *M. thibetana*: 9.3%, between *M. mulatta* and *M. fascicularis*: 10.8%, between *M. thibetana* and *M. fascicularis*: 7.7%, and between *M. thibetana* and *M. mulatta*: 7.2%) compared to non-Ninteract windows (2.5%, 2.1%, 1.2%, 1.5%, 2.5%, 2.1%, 2.4%, 3.4%, 1.4%, and 1.5%, respectively). This excess was individually significant for all comparisons ($P < 0.05$; binomial tests).

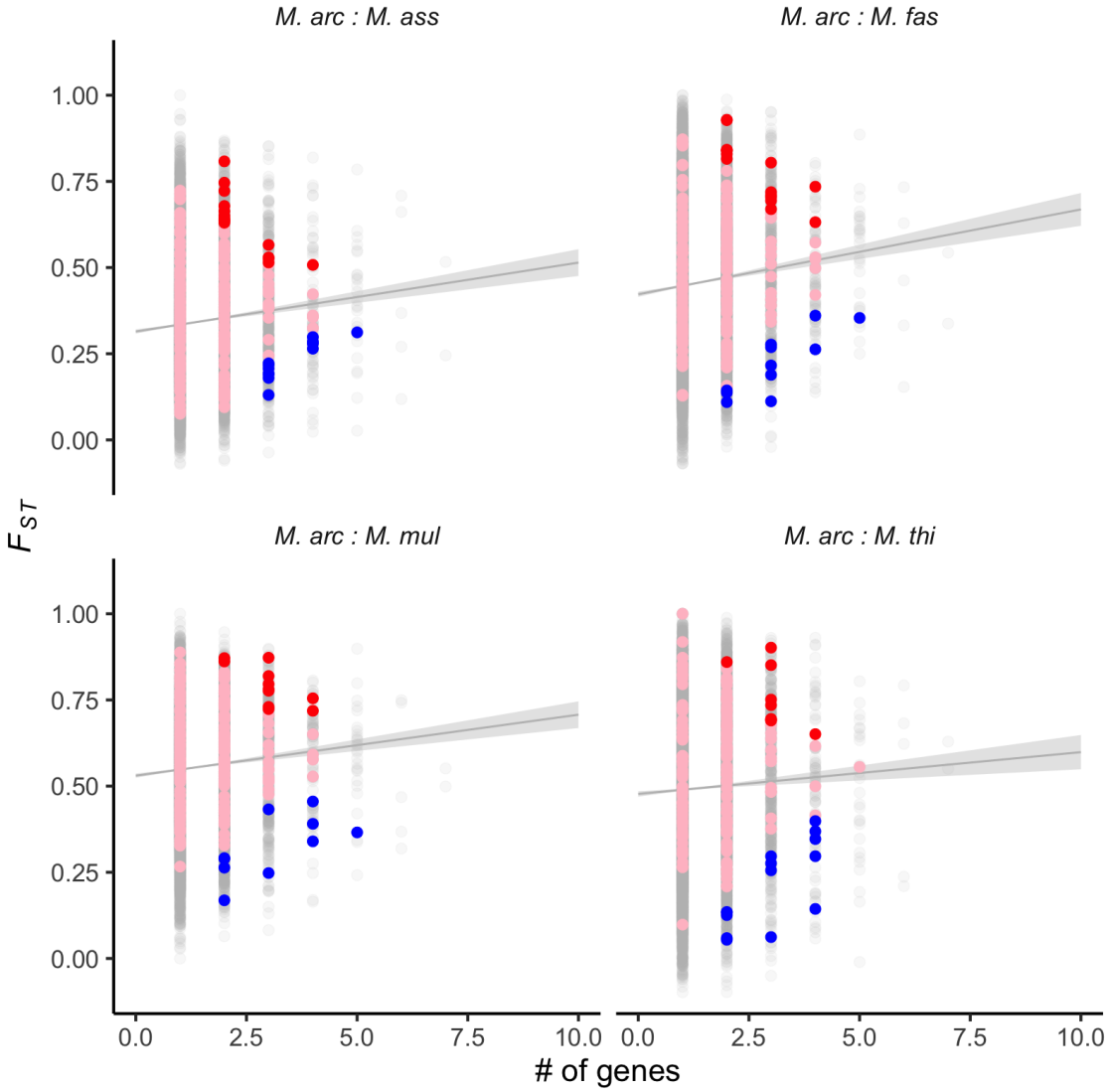

FIG. S17. F_{ST} outlier analysis for *M. arctoides* dataset identifies a significant excess of upper F_{ST} outliers in Ninteract windows (red dots above fitted line). Species names are abbreviated with first three letters; only pairwise comparisons involving the focal species *M. arctoides* are shown. Blue dots indicate Ninteract windows with atypically low differentiation in each pairwise comparison.

A3.7 F_{ST} analysis; *M. f. aurea*

For the *M. f. aurea* dataset, three pairwise comparisons were considered between three groups including one (*M. thibetana* and *M. assamensis*) that is a group comprised of two closely related species. Outlier analysis (Fig. S18) identified a significant excess of Ninteract windows that were upper F_{ST} outliers (comparisons between *M. f. aurea* and *M. fascicularis*: 9.8%, between *M. f. aurea* and *M. thibetana* 5.2%, between *M. f. aurea* and *M. assamensis*: 7.2%, between *M. fascicularis* and *M. thibetana*: 6.2%, between *M. fascicularis* and *M. assamensis*: 9.2%, and between *M. thibetana* and *M. assamensis*: 7.7%) compared to non-Ninteract windows (3.5 %, 2.1 %, 2.5 %, 2.4 %, 3.0 %, and 2.3%, respectively). This excess was individually significant for all comparisons ($P < 0.05$; binomial tests).

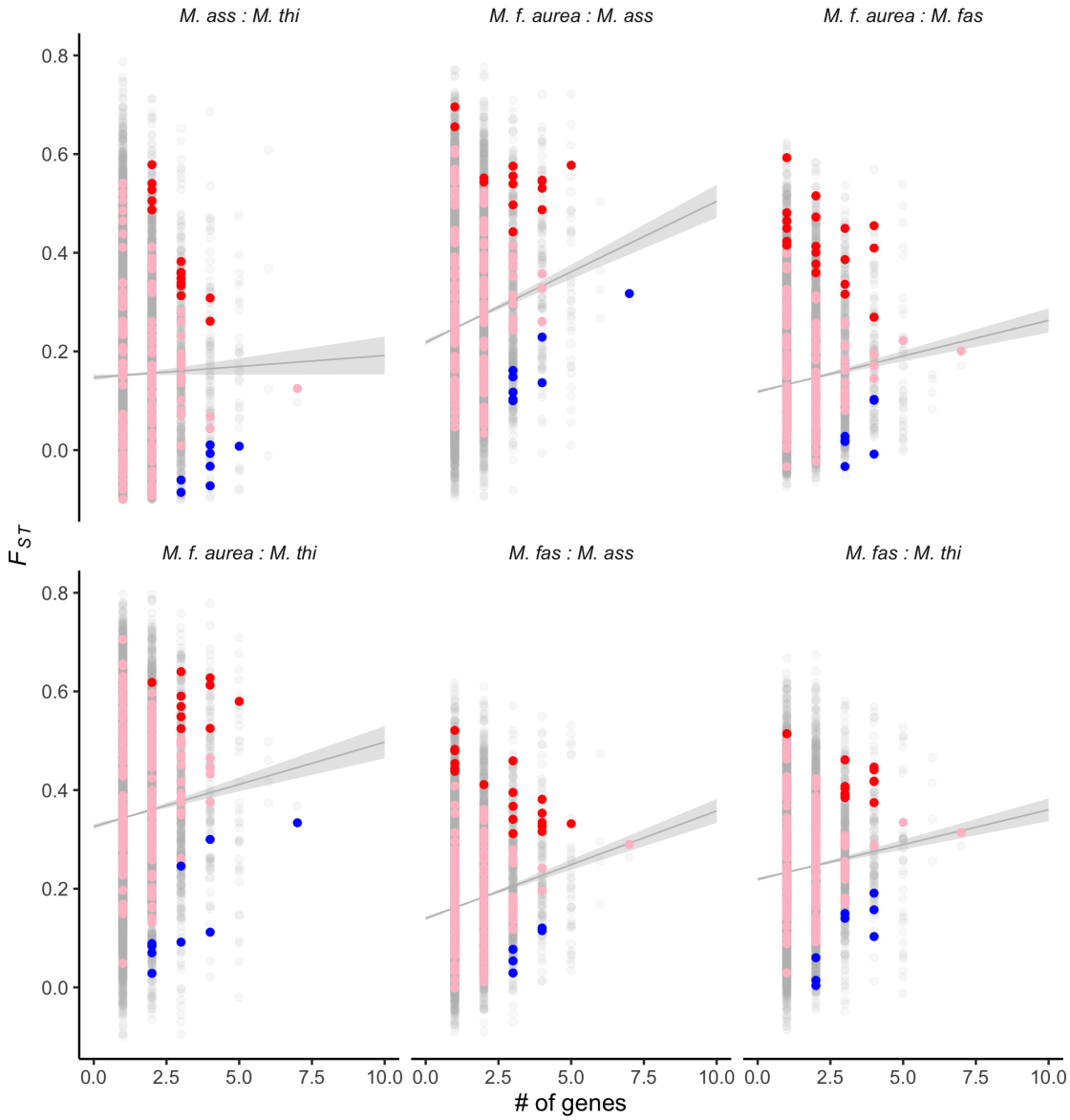

FIG. S18. F_{ST} outlier analysis for *M. f. aurea* data identifies a significant excess of upper F_{ST} outliers in Ninteract windows (red dots above fitted line). Blue dots indicate Ninteract windows with atypically low differentiation in each pairwise comparison.

A3.8 π analysis; wild *M. mulatta* from China

For the three populations of wild *M. mulatta* from China defined based on mitochondrial clades, outlier analysis identified a significant excess of Ninteract windows that were lower π outliers (Blue: 4.6%, Red: 6.2%, and Purple: 5.7%) compared to non-Ninteract windows (Blue: 1.6%, Red: 1.5%, and Purple: 1.4%, binomial tests, $P < 0.01$, Fig. S19).

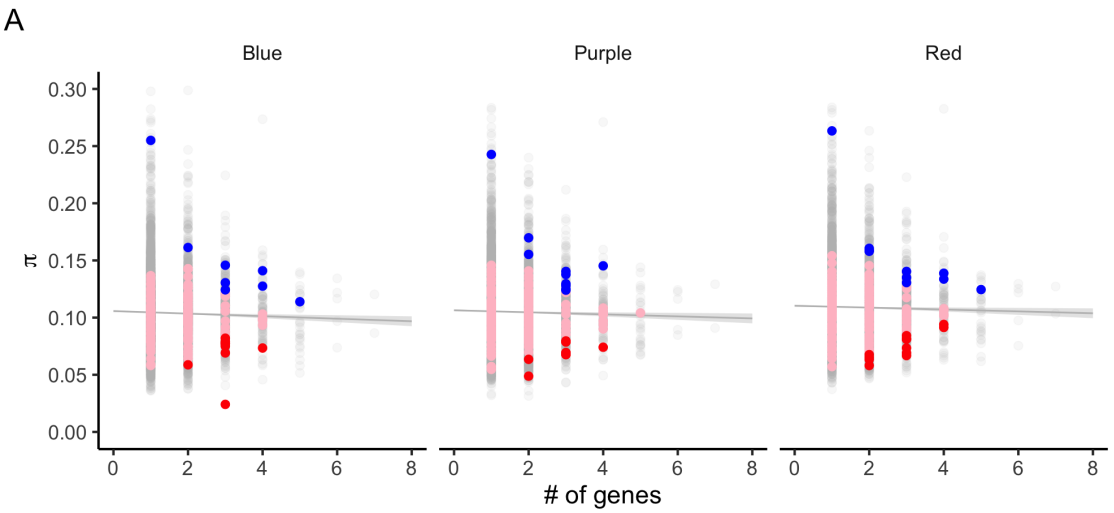

FIG. S19. π outlier analysis for wild *M. mulatta* from China identifies a significant excess of lower π outliers in Ninteract windows (red dots below fitted line). The very low value for π in the Blue group is a window that contains the Ninteract gene NDUFB8. Blue dots indicate Ninteract windows with atypically high diversity.

A3.9 π analysis; captive *M. mulatta* from India

For the three populations of captive *M. mulatta* from India defined based on mitochondrial clades, outlier analysis identified a significant excess of Ninteract windows that were lower π outliers (Brown: 5.2%, Orange: 4.1%, and Red: 6.2%) compared to non-Ninteract windows (Brown: 1.5%, Orange: 1.6%, and Red: 2.1%, binomial tests, $P < 0.05$, Fig. S20).

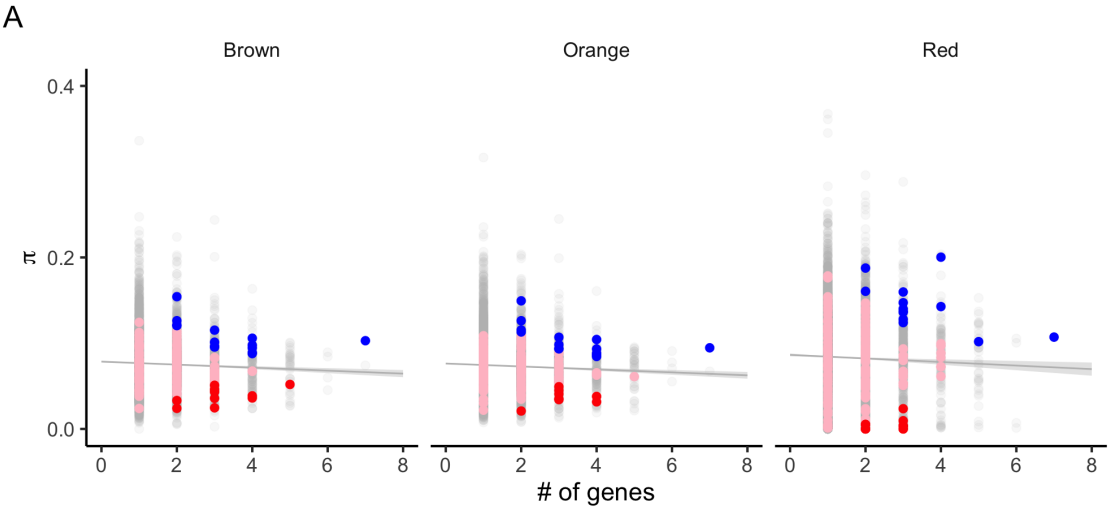

FIG. S20. π outlier analysis for captive *M. mulatta* from India identifies a significant excess of lower π outliers in Ninteract windows (red dots below fitted line). For the Red group, there are 11 lower π outliers, but many of them overlap in the plot. Blue dots indicate Ninteract windows with atypically high diversity.

A3.10 π analysis; *M. arctoides* dataset

For *M. arctoides* dataset, outlier analysis identified a significant excess of Ninteract windows that were lower π outliers (*M. arctoides*: 2.6%, *M. fascicularis*: 6.2%, *M. mulatta*: 8.2%, *M. thibetana*: 2.6%, *M. assamensis*: 8.2%) compared to non-Ninteract windows (0.1%, 1.6%, 1.2%, 0.2%, and 1.4%, respectively, binomial tests, $P < 0.05$, Fig. S21).

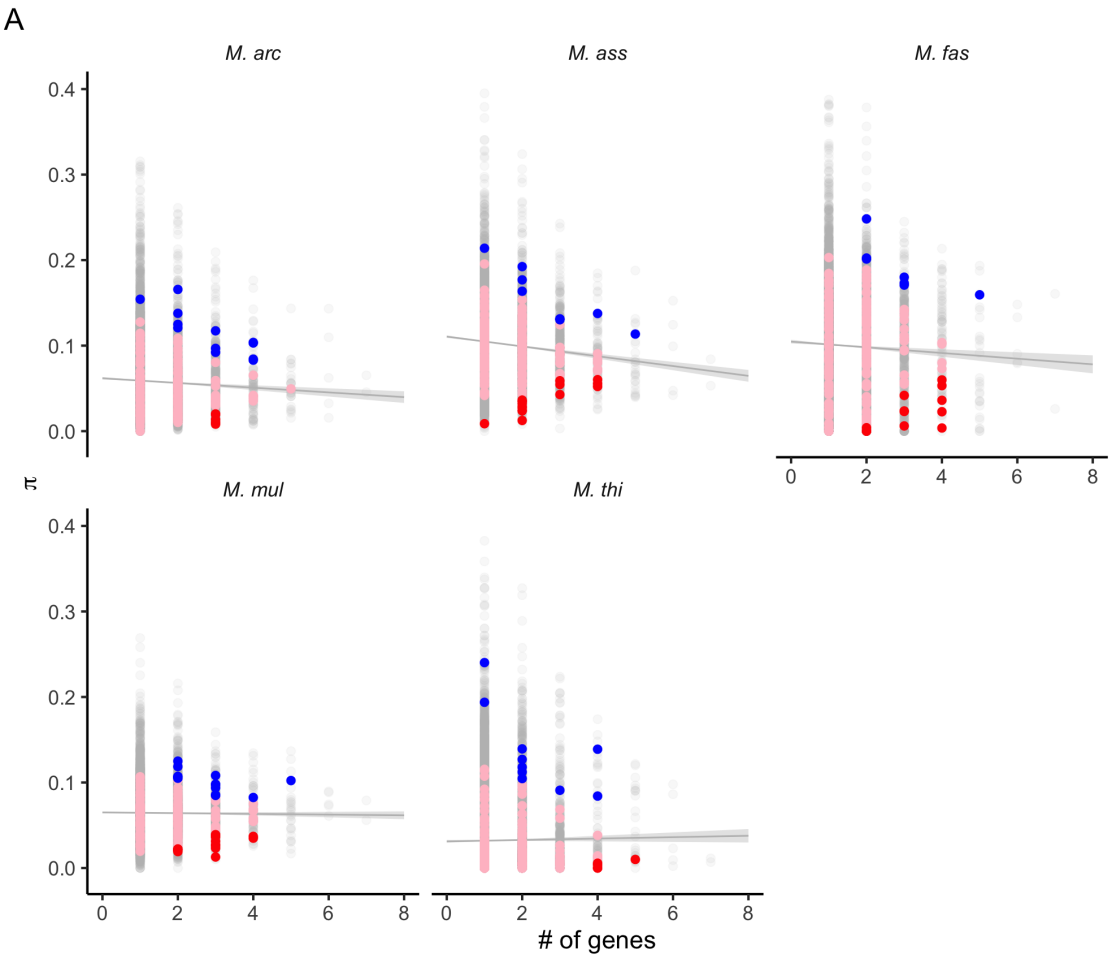

FIG. S21. π outlier analysis for the *M. arctoides* dataset identifies a significant excess of lower π outliers in Ninteract windows (red dots below fitted line). For *M. arctoides*, there are 5 lower π outliers, but many of them overlap in the plot. Blue dots indicate Ninteract windows with atypically high diversity.

A3.11 π analysis; *M. f. aurea* dataset

For *M. f. aurea* dataset, outlier analysis identified a significant excess of Ninteract windows that were lower π outliers (*M. f. aurea*: 5.7%, *M. fascicularis*: 10.3%, *M. thibetana*: 4.1%, and *M. assamensis*: 4.6%) compared to non-Ninteract windows (1.9%, 2.7%, 0.7%, and 1.1%, respectively, binomial tests, $P < 0.05$, Fig. S22).

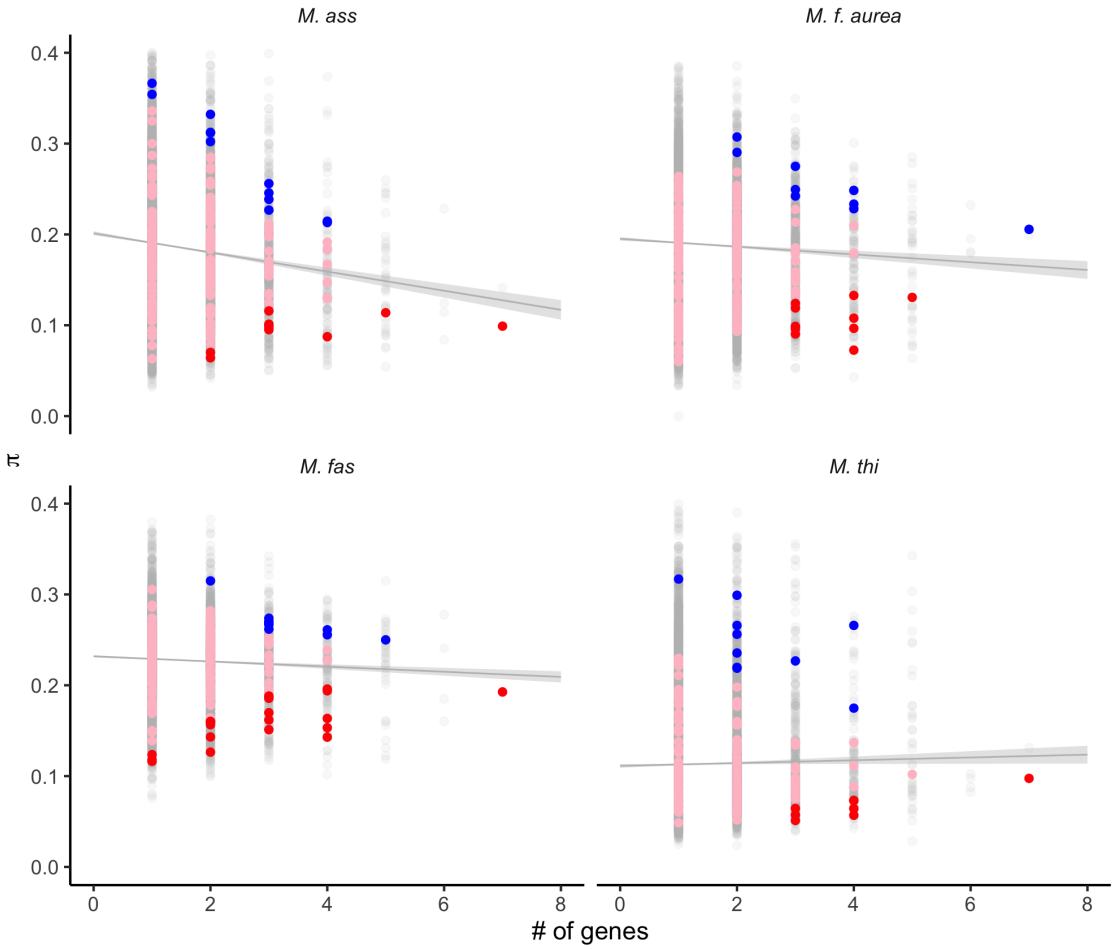

FIG. S22. π outlier analysis for *M. f. aurea* dataset identifies a significant excess of lower π outliers in Ninteract windows (red dots below fitted line). Blue dots indicate Ninteract windows with atypically high diversity.

A3.12 F_{ST} outlier intersection: wild *M. mulatta* from China

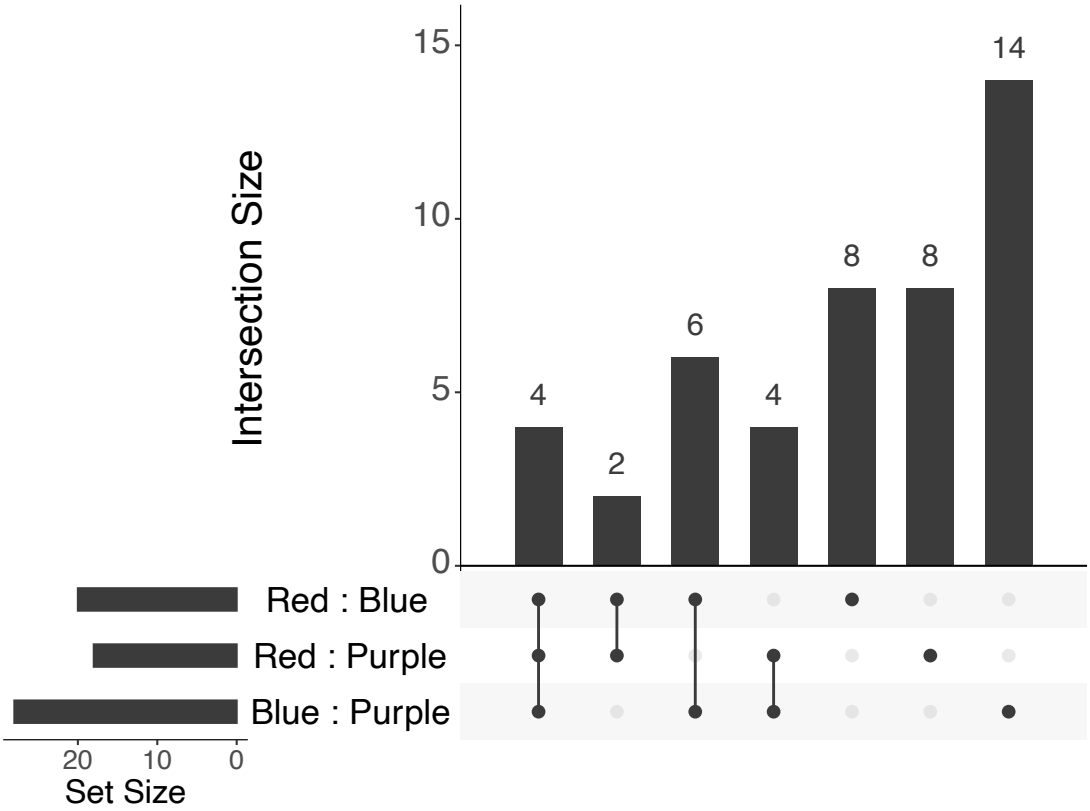

FIG. S23. Upset plot illustrating common outliers from the 30kb window F_{ST} analysis of the *M. mulatta* from China dataset. On the bottom left, the set size histogram illustrates the total number of outliers in each comparison; on the top, the intersection size illustrates the number of outliers that are shared across comparisons. For example, there are 19 F_{ST} outliers in the Red: Blue comparison (listed in Table S7); four of these are found in all three pairwise comparisons, seven are found only in this pairwise comparison.

A3.13 F_{ST} outlier intersection: captive *M. mulatta* from India

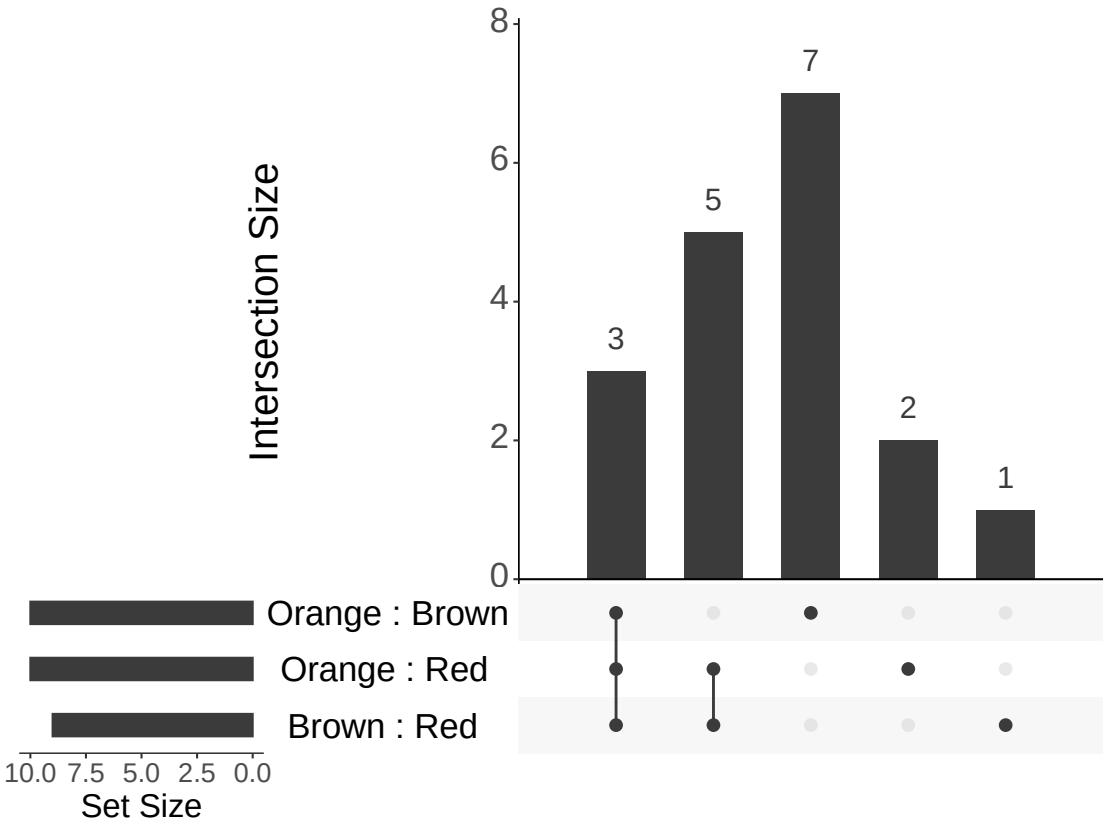

FIG. S24. Upset plot illustrating common outliers from the 30kb window F_{ST} analysis of the *M. mulatta* from India dataset. Labeling follows Fig. S23.

A3.14 F_{ST} outlier intersection: *M. arctoides* dataset

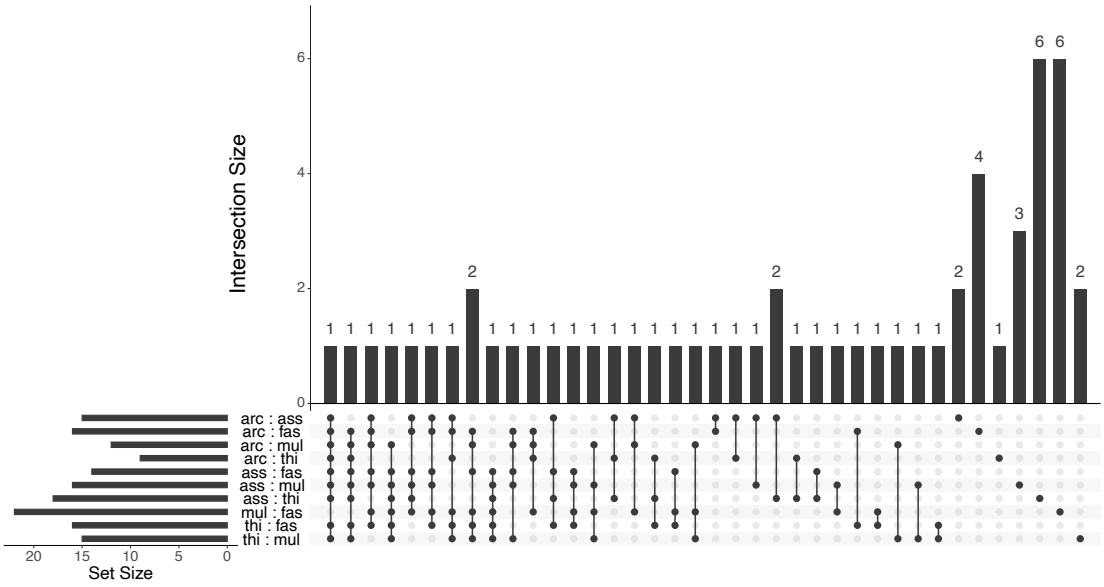

FIG. S25. Upset plot illustrating common outliers from the 30kb window F_{ST} analysis of the *M. arctoides* dataset. Labeling follows Fig. S23.

A3.15 F_{ST} outlier intersection: *M. f. aurea* dataset

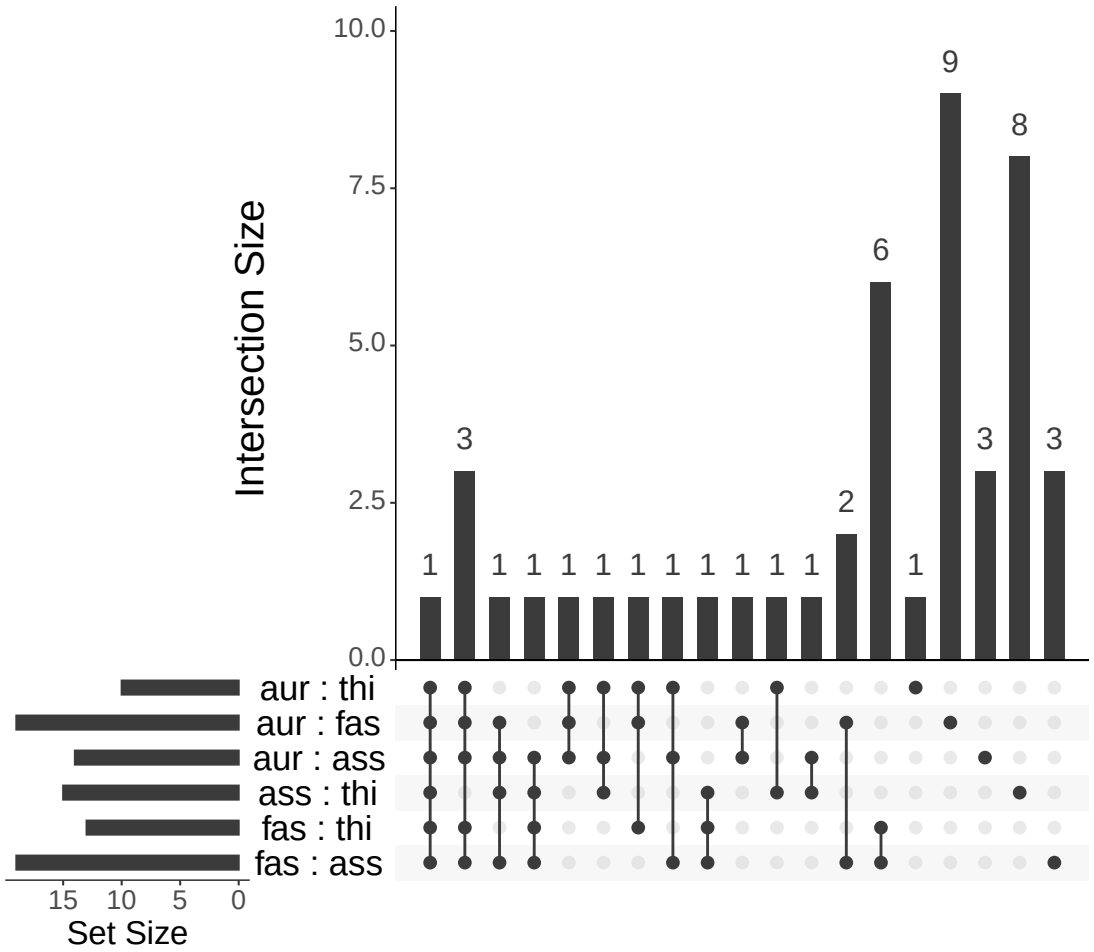

FIG. S26. Upset plot illustrating common outliers from the 30kb window F_{ST} analysis of the *M. f. aurea* dataset. Labeling follows Fig. S23.

A3.16 π outlier intersection: wild *M. mulatta* from China

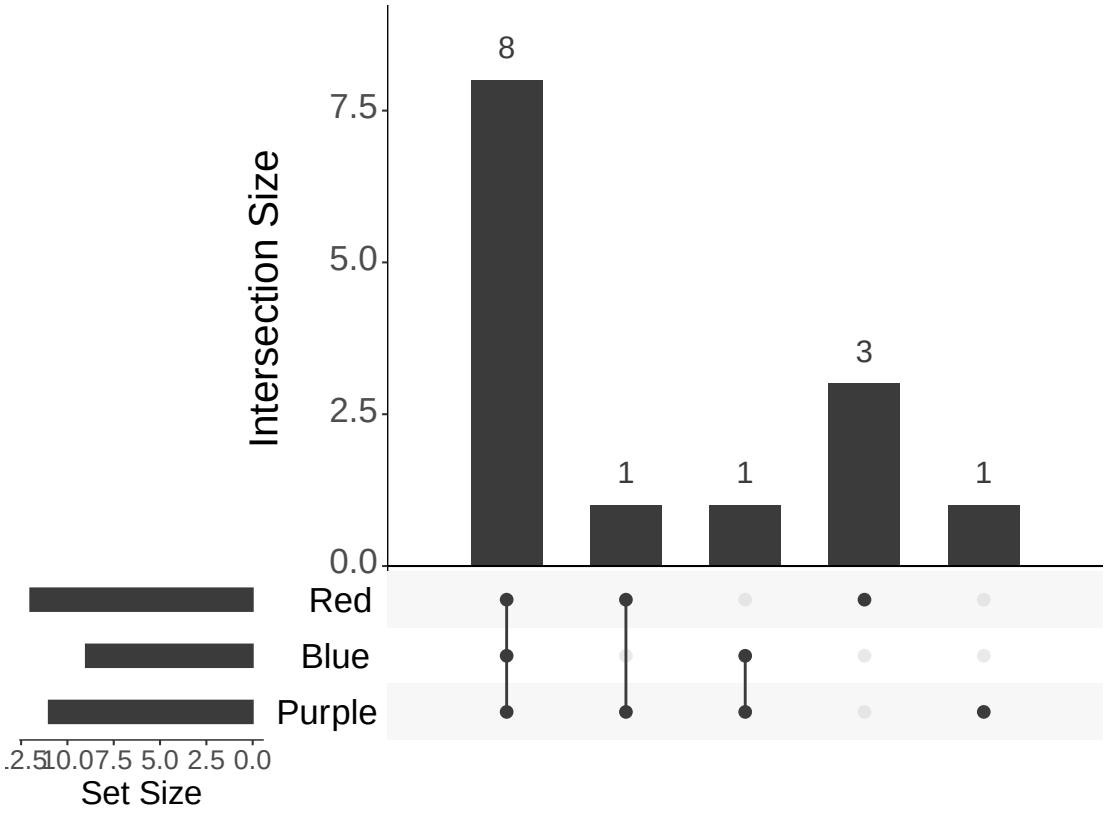

FIG. S27. Upset plot illustrating common outliers from the 30kb window π analysis of the *M. mulatta* from China dataset. Labeling follows Fig. S23.

A3.17 π outlier intersection: wild *M. mulatta* from India

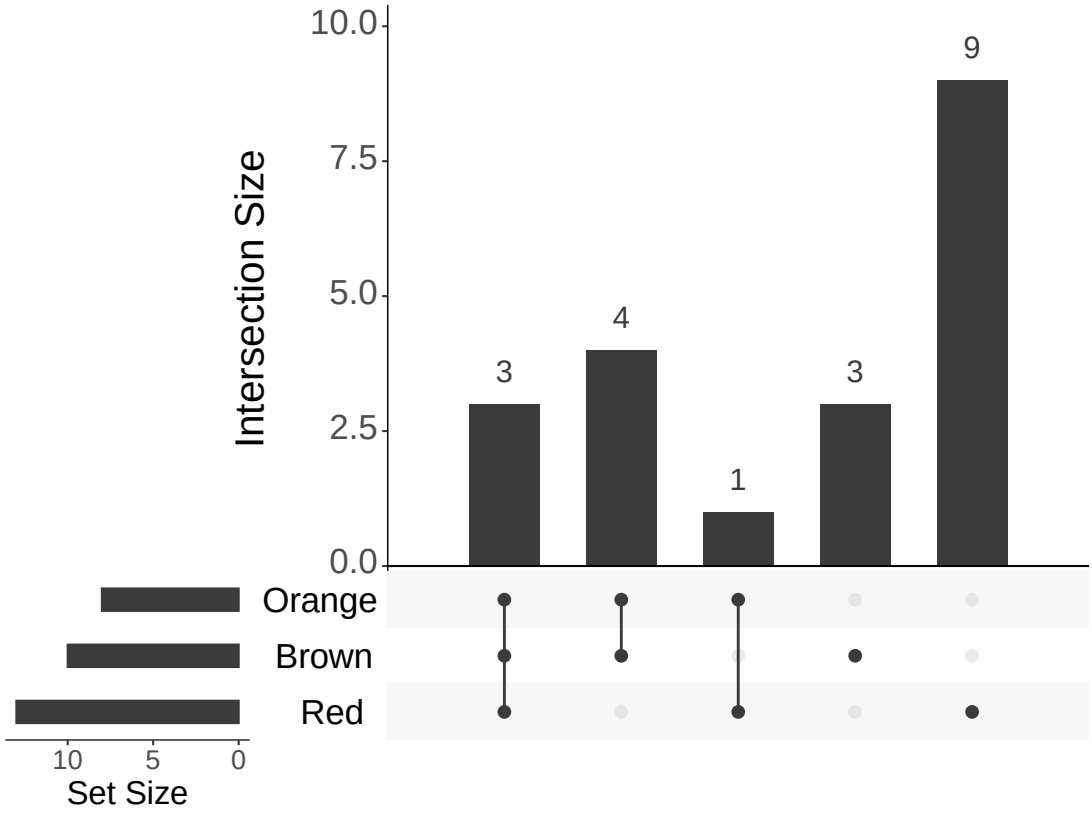

FIG. S28. Upset plot illustrating common outliers from the 30kb window π analysis of the *M. mulatta* from India dataset. Labeling follows Fig. S23.

A3.18 π outlier intersection: *M. arctoides* dataset

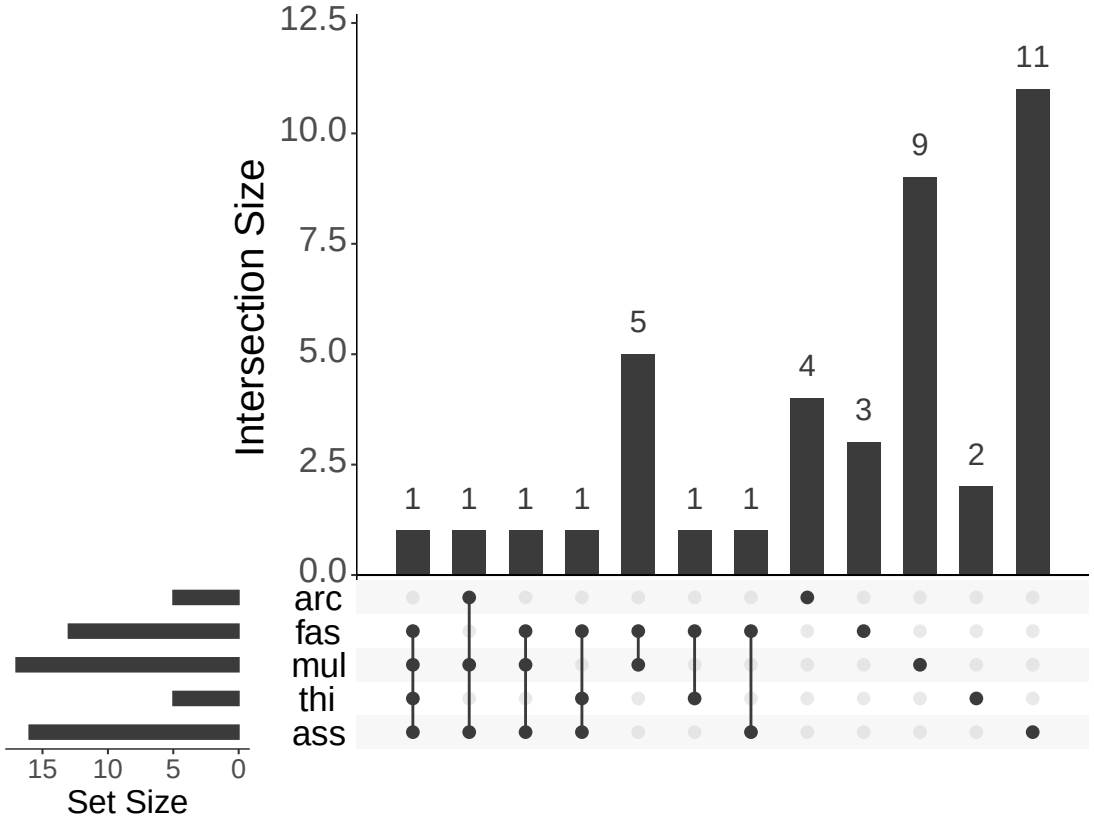

FIG. S29. Upset plot illustrating common outliers from the 30kb window π analysis of the *M. arctoides* dataset. Labeling follows Fig. S23.

A3.19 π outlier intersection: *M. f. aurea* dataset

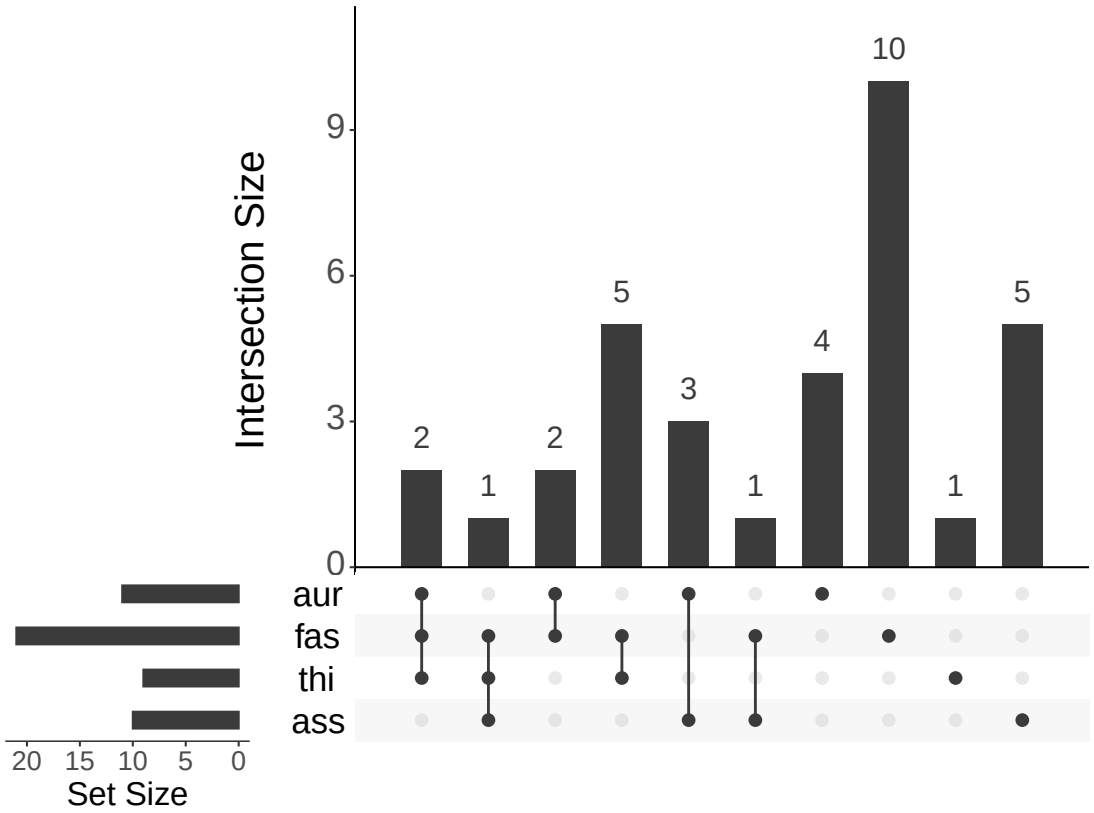

FIG. S30. Upset plot illustrating common outliers from the 30kb window π analysis of the *M. f. aurea* dataset. Labeling follows Fig. S23.

A3.20 Introgression analysis

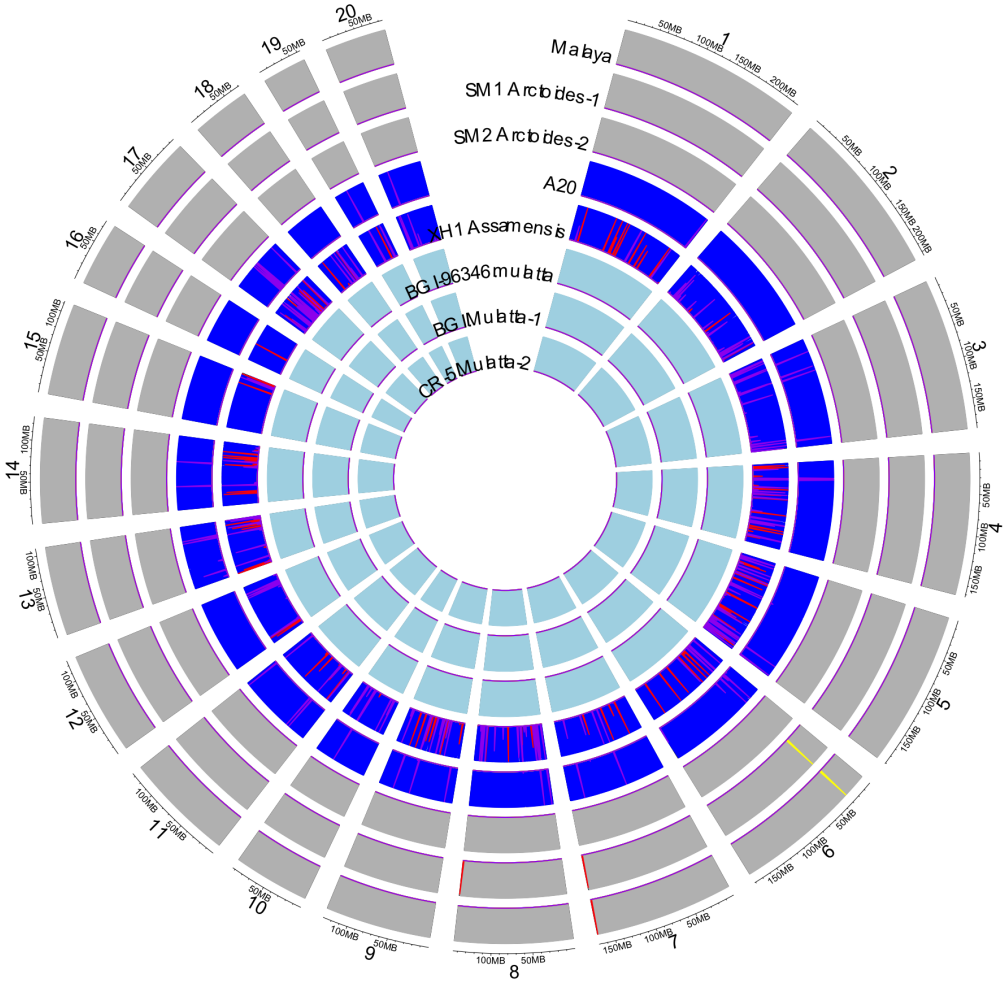

FIG. S31. Admixfrog analysis of *M. arctoides*, *M. assamensis*, and *M. mulatta*. Gray, blue, light blue, red, yellow, and purple indicate homozygous *M. arctoides*, homozygous *M. assamensis*, homozygous *M. mulatta*, heterozygous *M. arctoides*/*M. assamensis*, heterozygous *M. arctoides*/*M. mulatta* and heterozygous *M. assamensis*/*M. mulatta* respectively. Numbers refer to chromosomes and individual genomes are in each layer. As detailed in the main text, most individuals do not have much evidence of recent gene flow with the exception of one *M. assamensis* individual where extensive gene flow with *M. arctoides* is evidenced by an abundance of heterozygous *M. arctoides*/*M. assamensis* blocks.

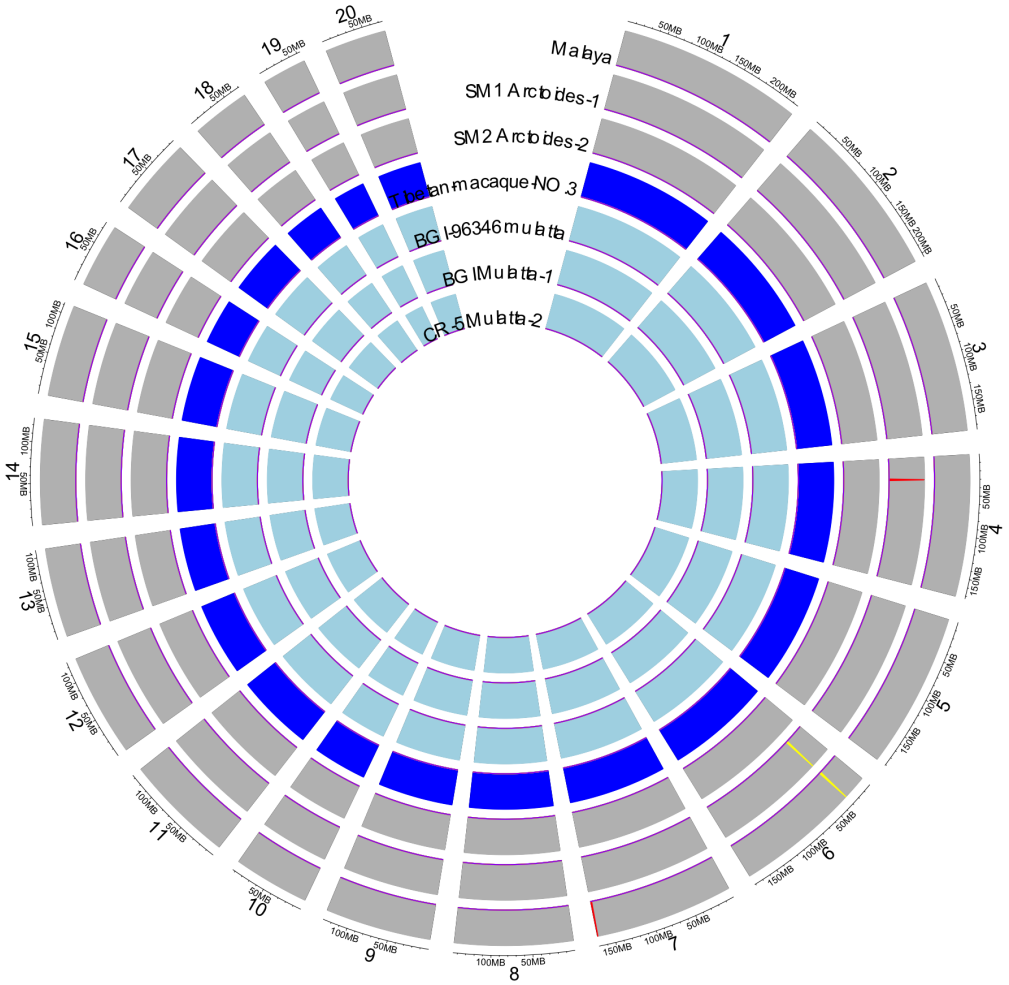

FIG. S32. Admixfrog analysis of *M. arctoides*, *M. thibetana*, and *M. mulatta*. Gray, blue, light blue, red, yellow, and purple indicate homozygous *M. arctoides*, homozygous *M. assamensis*, homozygous *M. mulatta*, heterozygous *M. arctoides*/*M. thibetana*, heterozygous *M. arctoides*/*M. mulatta* and heterozygous *M. thibetana* /*M. mulatta* respectively. Numbers refer to chromosomes and individual genomes are in each layer.

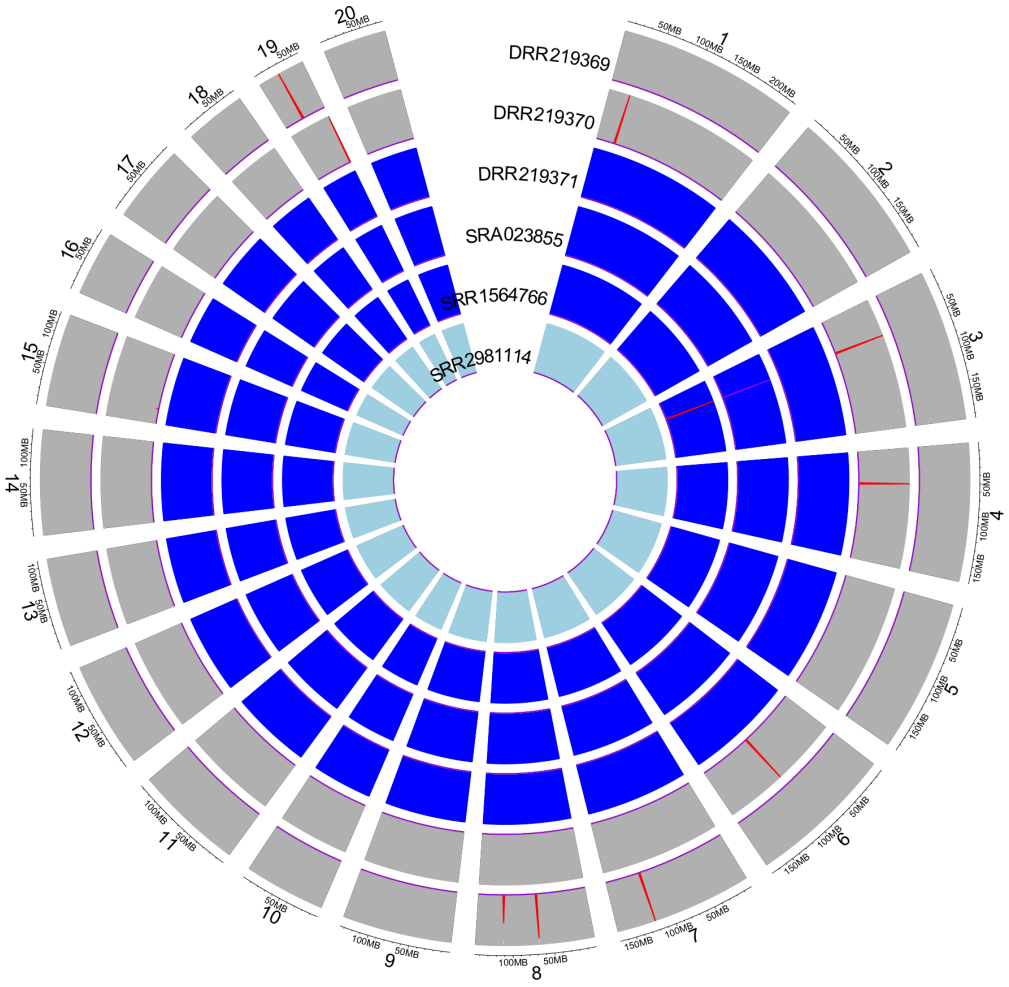

FIG. S33. Admixfrog analysis of *M. f. aurea*, *M. fascicularis*, and *M. assamensis*. Gray, blue, light blue, red, yellow, and purple indicate homozygous *M. f. aurea*, homozygous *M. fascicularis*, homozygous *M. assamensis*, heterozygous *M. f. aurea* /*M. fascicularis*, heterozygous *M. f. aurea* /*M. assamensis* and heterozygous *M. fascicularis* /*M. assamensis* respectively. Numbers refer to chromosomes and individual genomes are in each layer.

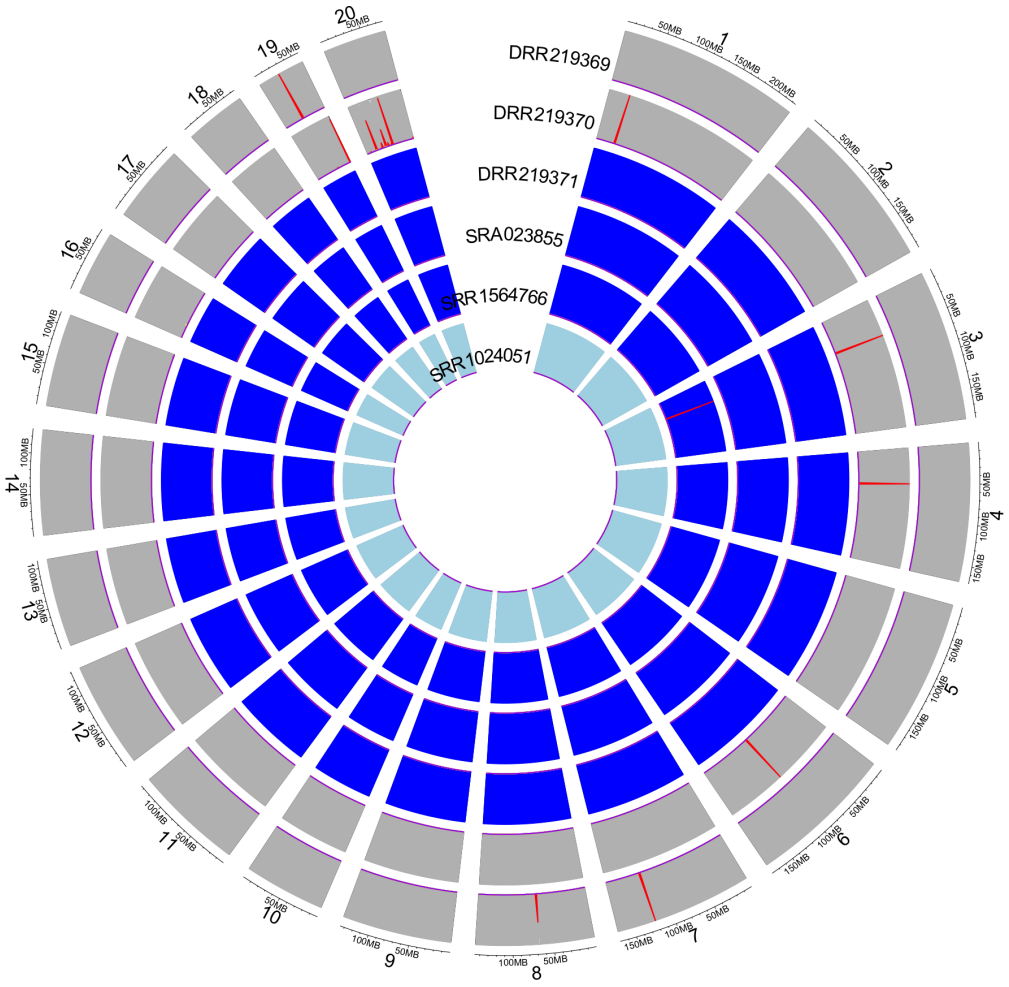

FIG. S34. Admixfrog analysis of *M. f. aurea*, *M. fascicularis*, and *M. thibetana*. Gray, blue, light blue, red, yellow, and purple indicate homozygous *M. f. aurea*, homozygous *M. fascicularis*, homozygous *M. thibetana*, heterozygous *M. f. aurea* /*M. fascicularis*, heterozygous *M. f. aurea* /*M. thibetana* and heterozygous *M. fascicularis* /*M. thibetana* respectively. Numbers refer to chromosomes and individual genomes are in each layer.

A4 Comparison to *silenus* group macaques

Signatures of natural selection were frequently concordant in independent comparisons involving different sets of species or populations in this study and also comparisons in the *silenus* and Sulawesi group species and populations (Evans et al. 2021). For example, NDUFA2, which is part of the N-module of OXPHOS complex I, was an upper

F_{ST} outliers in at least four comparisons (F_{ST}) and a lower π outliers in at least four populations/species studied here (Tables S7,S8,S10), and this gene was also an upper F_{ST} outlier in ten of ten pairwise comparisons in the *silenus* group and a lower π outliers in four of five species or populations in the *silenus* group (Evans et al. 2021). HARS2, which catalyzes the ATP-dependent ligation of histidine to the mitochondrial tRNA for this amino acid, was an outlier in many comparisons in this study (Tables S7,S8,S10) and also was an upper F_{ST} outlier in nine of ten pairwise comparisons in the *silenus* group and a lower π outlier in two of five species or populations in the *silenus* group (Evans et al. 2021). Similarly, MRPL55, which contributes to mitochondrial translation as a component of the large subunit of the mitochondrial ribosome, was an outlier in many comparisons in this study (Tables S7,S8,S10), and was an upper F_{ST} outlier in seven of ten pairwise comparisons in the *silenus* group and a lower π outlier in three of five species or populations examined (Evans et al. 2021). Although a lower π outlier in fewer than four or more species studied here, the REP gene POLRMT, which is involved with mitochondrial RNA polymerase and the initiation of mitochondrial replication, was an upper F_{ST} outlier in five comparisons in this study and also was an outlier for several metrics in *silenus* group macaques (Evans et al. 2021).

A5 Examples of mitonuclear discordance in macaques

There are many examples of mitonuclear phylogenomic discordance in macaques. For example, mitochondrial DNA from *M. mulatta* is paraphyletic with respect to *M. fuscata* and *M. cyclopis*, that of *M. nemestrina* is paraphyletic with respect to *M. leonina*, *M. silenus* and the Sulawesi macaques, and mitochondrial relationships among Sulawesi macaques are poorly resolved whereas nuclear variation strongly supports reciprocal monophyly of species on the northern peninsula and species from the rest of Sulawesi (Fig. 1; Evans et al. 2017; Evans et al. 2020). Within *M. fascicularis*, Y chromosome variation has a sharp genetic discontinuity at the Isthmus of Kra whereas mitochondrial chromosome variation in the same individuals does not (Tosi et al. 2002). As detailed in the main text, there are also striking examples of introgression of diverged mitochondrial in *M. f. aurea* and *M. arctoides* (Tosi et al. 2003; Fan et al. 2018; Matsudaira et al. 2018) that also render evolutionary relationships among mitochondrial and nuclear genomes discordant.

TABLE S1. Results of linear models and permutation tests for runs of homozygosity. Coefficients and 95% confidence intervals in parentheses for the parameters of the linear model (see Methods for definitions of parameters) are mostly significant (*) but permutation test P values are generally not individually significant. Ninteract ROHs are significantly longer than non-Ninteract ROHs across all linear models and across all permutation tests (see main text).

Population	Ninteract	Number_of_genes	Interaction	Permutation P value
M. mulatta from China				
Blue	0.730 (±0.034)*	0.228 (±0.002)*	-0.147 (±0.003)*	0.793
Red	0.512 (±0.047)*	0.283 (±0.004)*	-0.170 (±0.008)*	0.31
Purple	0.696 (±0.041)*	0.243 (±0.002)*	-0.149 (±0.004)*	0.966
M. mulatta from India				
Orange	0.834 (±0.025)*	0.244 (±0.001)*	-0.174(±0.002)*	0.784
Brown	0.722 (±0.079)*	0.261 (±0.005)*	-0.167 (±0.009)*	0.102
Red	0.274 (±0.262)*	0.351 (±0.018)*	-0.081 (±0.062)*	0.236
M. arctoides dataset				
M. arctoides	0.637 (±0.131)*	0.183 (±0.005)*	-0.097 (±0.012)*	0.055
M. thibetana	0.594 (±0.224)*	0.097 (±0.006)*	-0.040 (±0.013)*	0.307
M. mulatta	1.113 (±0.118)*	0.198 (±0.005)*	-0.155 (±0.007)*	0.000*
M. fascicularis	0.509 (±0.264)*	0.279 (±0.016)*	-0.120 (±0.044)*	0.746
M. f. aurea dataset				
M. f. aureus	0.426 (±0.234)*	0.266 (±0.010)*	-0.097 (±0.030)*	0.873
M. fascicularis	0.563 (±0.182)*	0.339 (±0.011)*	-0.185 (±0.025)*	0.771
M. thibetana	0.543 (±0.237)*	0.197 (±0.010)*	-0.097 (±0.027)*	0.734
M. assamensis	0.371 (±0.296)*	0.393 (±0.024)*	-0.128 (±0.082)*	0.057

*Individually significant ($P < 0.05$)

TABLE S2. Results of linear models and permutation tests for F_{ST} analysis in 30kb genomic windows. For the linear model, coefficients are followed by block bootstrap 95% confidence intervals in parentheses (see Methods for definitions of parameters)

	Comparison	Ninteract	gene_number	Interaction	Permutation P value
M. mulatta from China					
	Blue : Purple	-0.001(±0.005)	0.000(±0.000)*	0.002(±0.003)	0.000*
	Blue : Red	0.001(±0.005)	0.000(±0.000)	0.000(±0.003)	0.007*
	Purple : Red	-0.002(±0.004)	0.000(±0.000)*	0.002(±0.002)	0.037*
M. mulatta from India					
	Brown : Orange	-0.000(±0.001)	0.000(±0.000)	0.000(±0.001)	0.641
	Brown : Red	-0.007(±0.013)	0.001(±0.002)	0.001(±0.008)	0.907
	Orange : Red	0.000(±0.002)	0.000(±0.000)	-0.000(±0.001)	0.925
M. arctoides dataset					
	M. arctoides : *M. assamensis*	0.084(±0.054)*	0.020(±0.005)*	-0.032(±0.020)*	0.010*
	M. arctoides : *M. thibetana*	0.082(±0.064)*	0.012(±0.006)*	-0.028(±0.032)	0.015*
	M. arctoides : *M. mulatta*	0.047(±0.050)	0.018(±0.005)*	-0.013(±0.027)	0.004*
	M. arctoides : *M. fascicularis*	0.049(±0.064)	0.025(±0.006)*	-0.016(±0.029)	0.041*
	M. assamensis : *M. thibetana*	-0.026(±0.048)	0.002(±0.005)	0.017(±0.028)	0.31
	M. assamensis : *M. mulatta*	0.034(±0.056)	0.021(±0.006)*	-0.004(±0.028)	0.000*
	M. assamensis : *M. fascicularis*	0.025(±0.053)	0.030(±0.006)*	-0.005(±0.029)	0.037*
	M. thibetana : *M. mulatta*	0.022(±0.047)	0.008(±0.005)*	0.003(±0.029)	0.004*
	M. mulatta : *M. fascicularis*	0.020(±0.063)	0.019(±0.005)*	-0.004(±0.029)	0.111
M. f. aurea dataset					
	M. f. aureus : *M. assamensis*	0.060(±0.049)*	0.028(±0.006)*	-0.015(±0.025)	0.000*
	M. f. aureus : *M. thibetana*	0.040(±0.046)	0.017(±0.005)*	-0.006(±0.028)	0.001*
	M. f. aureus : *M. fascicularis*	0.047(±0.034)*	0.014(±0.004)*	-0.014(±0.020)	0.003*
	M. fascicularis : *M. assamensis*	0.035(±0.035)*	0.022(±0.004)*	-0.010(±0.016)	0.001*
	M. fascicularis : *M. thibetana*	0.019(±0.032)	0.014(±0.004)*	-0.003(±0.016)	0.012*
	M. assamensis : *M. thibetana*	-0.030(±0.058)	0.004(±0.005)*	0.011(±0.032)	0.752

*Individually signficant ($P < 0.05$)

TABLE S3. Results of linear models and permutation tests for F_{ST} analysis in 100kb genomic windows. For the linear model, coefficients are followed by block bootstrap 95% confidence intervals in parentheses (see Methods for definitions of parameters).

	Comparison	Ninteract	gene_number	Interaction	Permutation P value
M. mulatta from China					
	Blue : Purple	0.002(±0.003)	0.000(±0.000)*	-0.000(±0.001)	0.006*
	Blue : Red	0.002(±0.002)	0.000(±0.000)	-0.000(±0.000)	0.011*
	Purple : Red	0.002(±0.002)*	0.000(±0.000)*	-0.000(±0.001)	0.009*
M. mulatta from India					
	Brown : Orange	-0.000(±0.001)	0.000(±0.000)	-0.000(±0.000)	0.978
	Brown : Red	-0.001(±0.009)	0.001(±0.001)	-0.001(±0.004)	0.9
	Orange : Red	0.000(±0.001)	0.000(±0.000)	-0.000(±0.000)	0.938
	M. arctoides : *M. assamensis*	0.037(±0.040)	0.008(±0.002)*	-0.002(±0.011)	0.000*
	M. arctoides : *M. thibetana*	0.042(±0.039)	0.005(±0.003)*	-0.001(±0.012)	0.000*
	M. arctoides : *M. mulatta*	0.047(±0.034)*	0.006(±0.003)*	-0.005(±0.009)	0.000*
	M. arctoides : *M. fascicularis*	0.063(±0.046)*	0.010(±0.003)*	-0.008(±0.010)	0.000*
	M. assamensis : *M. thibetana*	0.005(±0.031)	0.003(±0.002)*	0.001(±0.008)	0.114
	M. assamensis : *M. mulatta*	0.041(±0.037)*	0.009(±0.002)*	-0.004(±0.010)	0.000*
	M. assamensis : *M. fascicularis*	0.049(±0.035)*	0.013(±0.003)*	-0.006(±0.012)	0.002*
	M. thibetana : *M. mulatta*	0.031(±0.031)	0.004(±0.002)*	-0.001(±0.011)	0.001*
	M. thibetana : *M. fascicularis*	0.044(±0.042)	0.007(±0.003)*	-0.003(±0.013)	0.001*
	M. mulatta : *M. fascicularis*	0.044(±0.047)	0.008(±0.002)*	-0.004(±0.011)	0.002*
M. f. aurea dataset					
	M. f. aureus : *M. assamensis*	0.047(±0.038)*	0.012(±0.002)*	-0.006(±0.009)	0.000*
	M. f. aureus : *M. thibetana*	0.026(±0.036)	0.007(±0.002)*	-0.001(±0.009)	0.003*
	M. f. aureus : *M. fascicularis*	0.042(±0.027)*	0.006(±0.002)*	-0.007(±0.008)	0.008*
	M. fascicularis : *M. assamensis*	0.032(±0.028)*	0.009(±0.002)*	-0.005(±0.006)	0.003*
	M. fascicularis : *M. thibetana*	0.017(±0.023)	0.006(±0.002)*	-0.001(±0.006)	0.005*
	M. assamensis : *M. thibetana*	-0.019(±0.045)	0.003(±0.002)*	0.007(±0.008)	0.355

*Individually signficant (P < 0.05)

TABLE S4. Results of linear models and permutation tests for π analysis in 30kb genomic windows. For the linear model, coefficients are followed by block bootstrap 95% confidence intervals in parentheses (see Methods for definitions of parameters).

	Population	Ninteract	gene_number	Interaction	Permutation P value
M. mulatta from China					
	Red	-0.003(±0.008)	-0.001(±0.001)*	0.000(±0.003)	0.058
	Purple	-0.001(±0.009)	-0.001(±0.001)*	-0.001(±0.004)	0.074
M. mulatta from India					
	Orange	-0.005 (0.003)	-0.002 (0.000)*	0.002 (0.002)	0.12
	Brown	-0.006 (0.004)	-0.002 (0.000)*	0.002 (0.002)	0.093
	Red	-0.007 (0.007)	-0.002 (0.001)*	0.003 (0.004)	0.354
M. arctoides dataset					
	M. arctoides	-0.011(±0.012)	-0.003(±0.001)*	0.003(±0.005)	0.332
	M. assamensis	-0.004(±0.010)	-0.006(±0.001)*	-0.001(±0.005)	0.071
	M. thibetana	0.003(±0.014)	0.001(±0.001)	-0.004(±0.006)	0.049*
	M. mulatta	-0.006(±0.008)	-0.000(±0.001)	0.001(±0.004)	0.062
	M. fascicularis	-0.005(±0.018)	-0.003(±0.002)*	0.001(±0.008)	0.013*
M. f. aurea dataset					
	M. f. aureus	-0.022(±0.020)*	-0.004(±0.002)*	0.005(±0.007)	0.001*
	M. fascicularis	-0.012(±0.012)	-0.003(±0.001)*	0.003(±0.005)	0.001*
	M. thibetana	0.008(±0.014)	0.002(±0.002)	-0.006(±0.008)	0.254
	M. assamensis	-0.006(±0.016)	-0.011(±0.002)*	0.002(±0.009)	0.071

*Individually signficant (P < 0.05)

TABLE S5. Results of linear models and permutation tests for π analysis in 100kb genomic windows. For linear models, coefficients are followed by 95% confidence intervals in parentheses (see Methods for definitions of parameters). π is significantly higher in Ninteract windows compared to non-Ninteract windows across all linear models and across all permutation tests (See main text).

	Population	Ninteract	gene_number	Interaction	Permutation P value
M. mulatta from China					
	Red	-0.002(±0.004)	0.000(±0.000)*	-0.000(±0.001)	0.154
	Purple	-0.003(±0.005)	0.000(±0.000)*	-0.000(±0.001)	0.429
M. mulatta from India					
	Orange	-0.003(±0.005)	-0.000(±0.000)	0.001(±0.001)	0.036*
	Brown	-0.003(±0.005)	-0.000(±0.000)	0.001(±0.001)	0.010*
	Red	-0.008(±0.009)	-0.001(±0.001)*	0.003(±0.002)*	0.015*
M. arctoides dataset					
	M. arctoides	-0.008(±0.006)*	-0.000(±0.000)	0.000(±0.002)	0.001*
	M. assamensis	-0.004(±0.008)	-0.002(±0.001)*	-0.000(±0.002)	0.003*
	M. thibetana	-0.000(±0.010)	0.000(±0.001)	-0.001(±0.002)	0.074
	M. mulatta	-0.004(±0.006)	0.000(±0.000)	0.000(±0.001)	0.004*
	M. fascicularis	-0.013(±0.012)*	-0.001(±0.001)*	0.001(±0.003)	0.003*
M. f. aurea dataset					
	M. f. aureus	-0.016(±0.015)	-0.001(±0.001)	0.002(±0.003)	0.001*
	M. fascicularis	-0.011(±0.008)*	-0.000(±0.001)	0.002(±0.002)	0.016*
	M. thibetana	0.008(±0.010)	0.001(±0.001)*	-0.003(±0.003)*	0.342
	M. assamensis	-0.007(±0.012)	-0.005(±0.001)*	0.000(±0.004)	0.020*

*Individually signficant ($P < 0.05$)

TABLE S6. Results of analysis of Tajima's D in 30kb and 100 kb genomic windows for each *M. mulatta* dataset. Population names match mitochondrial clades as indicated in the main text. Asterisks indicate individually significant P values based on permutation tests.

	Population	Ninteract	non-Ninteract	Difference	Permutation P value
30 kb windows					
	Chinese				
	Blue	-0.973898	-0.909971	-0.063927	0.022*
	Purple	-0.789795	-0.74444	-0.045356	0.088
	Red	-1.010326	-0.989135	-0.021191	0.239
	Indian				
	Brown	-0.017077	0.042884	-0.059961	0.091
	Green	0.297418	0.288634	0.008784	0.549
100 kb windows					
	Chinese				
	Blue	-0.955705	-0.906766	-0.048939	0.014*
	Purple	-0.784149	-0.739575	-0.044574	0.015*
	Red	-1.016996	-0.99312	-0.023876	0.07
	Indian				
	Brown	-0.027035	0.052608	-0.079643	0.015*
	Green	0.273335	0.302412	-0.029077	0.276

TABLE S7. Upper F_{ST} outliers in each pairwise comparison in each of the four datasets.

M. mulatta from China	
Red : Blue	TARS2, TIMMDC1, ATP5O, ATP5J2, MRPS36, MRPL52, NDUFB8, MRPL51, ACP1, MRPL11, MRPL49, SURF1, MRPL10, MRPL57, MRPL54, PET100, MRPL4, NDUFA13, MRPS34, COX4I1
Red : Purple	MRPL24, ATP5O, MRPS36, CD14,NDUFA2, HARS2, TMEM70, TYMP, BCS1L, SURF1, MRPL10, POLRMT, MRPL54, SARS2,MRPS12, MRPL28, MRPS34, EARS2
Blue : Purple	DARS2, ATP5O, ATP5J2, NDUFB2, MRPS18A, ATP5I, LYRM7, CD14,NDUFA2 COX5A, MRPL52, TFAM, NDUFB8, MRPL43, TYMP, NDUFS1, ACP1, UQCC3, ATP5L, SURF1, MRPS7, UQCR, MRPL54, NDUFA7, MRPL4, MRPL28, MRPS34, COX4I1 , ATP5O, NDUFB2, MRPS18A, ATP5I, LYRM7, CD14, NDUFA2, NDUFAF1, COX5A, MRPL52, TFAM, NDUFB8, MRPL43, TYMP, ACP1, UQCC3, ATP5L, SURF1, MRPS7, UQCR, MRPL54, NDUFA7, MRPL4, MRPS34, COX4I1
M. mulatta from India	
Brown : Orange	MRPL3, UQCRQ, NDUFB8, MRPL44, MRPL30, MRPL49, FOXRED1, MRPL54, SARS2,MRPS12
Orange : Red	LARS2, ATP5MF, TMEM70, CYC1, NDUFB8, SARS2,MRPS12, MRPL28, MRPS34, COX4I1
Brown : Red	LARS2, ACAD9, ATP5MF, NDUFB8, SARS2,MRPS12, MRPL28, MRPS34, COX4I1
M. arctoides dataset	
M arctoides : *M. assamensis*	DARS2, NDUFAF3, ATP5J2, ATP5I, TMEM70, MRPS16, MRPL40, TYMP, ATP5B, NDUFB3, UQCC3, TTC19, POLRMT, NDUFA13, MRPL28
M. arctoides : *M. fascicularis*	TARS2, MRPL55, NDUFAF3, ATP5J, AARS2, RARS2, CD14, NDUFA2, HARS2, CYC1, MRPS16, TYMP, MRPL51, ATP5B, UQCC3, NDUFA13
M. arctoides : *M. mulatta*	MRPL55, ATP5J2, CD14, NDUFA2, HARS2, NDUFB8, MRPL43, MRPL51, ATP5B, UQCC3, TMEM126B, MRPL34
M. arctoides : *M. thibetana*	ATP5I, HARS2, MRPL51, UQCC3, MRPS7, NDUFA7, NDUFA13, MRPL28, MRPS34

M. assamensis : *M. fascicularis*	MRPL55, NDUFAF3, MRPL2, CD14, NDUFA2, HARS2, MRPS16, MRPL43, USMG5, ATP5B, UQCC3, MRPL10, MRPL57, POLRMT
M. assamensis : *M. mulatta*	ATP5F1, MRPL55, DARS2, NDUFAF3, MRPL2, MRPS36, UQCRQ, HARS2, NDUFB1, MRPS16, MRPL43, USMG5, ATP5B, UQCC3, TMEM126B, MRPL45
M. assamensis : *M. thibetana*	MRPL2, ATP5I, MRPS18C, NDUFAF2, HARS2, TMEM70, MRPS16, MRPL43, MRPL40, UQCC3, MRPS2, CCDC56, MRPS7, MRPL38, MRPL57, POLRMT, MRPL20, MRPS34
M. mulatta : *M. fascicularis*	MRPL37, NDUFAF3, ATP5J2, MRPL2, MRPL1, NDUFC1, MRPS36, LYRM7, CD14, NDUFA2, MRPL22, MRPS16, NDUFB8, MRPL43, USMG5, MRPL51, ATP5B, BCS1L, COX7C, TMEM126B, MRPL10, NDUFA13
M. thibetana : *M. fascicularis*	MRPL37, NDUFAF3, MRPL2, CD14, NDUFA2, HARS2, CYC1, MRPL43, USMG5, ATP5B, MRPL49, UQCC3, MRPL10, MRPS7, POLRMT, NDUFA13
M. thibetana : *M. mulatta*	ATP5F1, MRPL9, MRPL55, MRPL2, CD14, NDUFA2, HARS2, NDUFB8, MRPL43, NDUFV1, MRPL49, UQCC3, TMEM126B, MRPL34, NDUFA13

M. f. aurea dataset	
M. f. aureus : *M. thibetana*	MRPS21, NDUFAF3, ATP5MF, MRPL14, UQCRQ, MRPL52, BCS1L, UQCC3, MRPL10, CCDC56
M. f. aureus : *M. fascicularis*	MRPS21, HIGD1A, NDUFAF3, RARS2, UQCRQ, COX16, NDUFB8, MARS2, NDUFB3, BCS1L, MRPL53, MRPL23, NARS2, TMEM126B, TTC19, ATPAF2, MRPL10, CCDC56, TACO1
M. f. aureus : *M. assamensis*	MRPS21, NDUFAF3, ATP5MF, TFB1M, RARS2, MRPS18A, UQCRQ, BCS1L, UQCC3, TTC19, MRPL10, MRPL12, CARS2, NDUFA13
M. assamensis : *M. thibetana*	MRPL20, NDUFAF3, COX19, ATP5MF, RARS2, ATP5ME, MRPL52, CYC1, MRPL21, CARS2, UQCR, NDUFA7, NDUFA13, MRPL28, MRPS34
M. fascicularis : *M. thibetana*	NDUFAF3, MRPL2, UQCRQ, NDUFA2, CD14, HARS2, CYC1, BCS1L, COX7C, MRPL49, MRPL10, CCDC56, NDUFA13

M. fascicularis : *M. assamensis*	PDC, NDUFAF3, RARS2, MRPL2, UQCRQ, NDUFA2, CD14, HARS2, COX6C, CYC1, BCS1L, COX7C, MRPL49, UQCC3, NARS2, MRPL10, MRPL45, TACO1, NDUFA13

Table S8. Lower π outliers for each population or species in each of the four datasets.

M. mulatta from China	
Red	MRPL9, MRPL55, ATP5J2, UQCRQ, ATP5S, NDUFB8, MRPL43, MRPL53, MRPL49, TMEM126B, MRPL10, NDUFA7
Blue	MRPL55, ATP5J2, HARS2, NDUFB8, MRPL43, MRPL53, MRPL49, TMEM126B, NDUFA7
Purple	MRPL55, ATP5J2, UQCRQ, HARS2, NDUFB8, MRPL43, MRPL53, NDUFV1, MRPL49, TMEM126B, NDUFA7
M. mulatta from India	
Orange	MRPS21, MRPL18, MRPL2, UQCRQ, ATP5MPL, NDUFB8, MRPL43, MRPL10
Brown	MRPS21, MRPL18, MRPL2, UQCRQ, ATP5MPL, MRPL43, NDUFS1, UQCC3, MRPL10, MRPS7
Red	ATP5MF, MRPL2, UQCRQ, ATP5MPL, TMEM70, NDUFB8, LOC716161, SARS2, MRPS12, MRPL28, MRPS34, EARS2, COX4I1
M. arctoides dataset	
M. arctoides	MRPL55, ATP5I, TYMP, MRPL51, MRPS34
M. fascicularis	MRPL37, MRPL2, CD14, NDUFA2, HARS2, CYC1, MRPL43, USMG5, MRPL49, UQCC3, MRPL10, POLRMT, NDUFA13
M. mulatta	MRPL37, ATP5F1, MRPL9, MRPL55, ATP5J2, MRPL2, CD14, NDUFA2, HARS2, NDUFB1, NDUFB8, MRPL43, USMG5, NDUFV1, TMEM126B, MRPL50, MRPL34
M. thibetana	MRPL2, MRPL49, UQCC3, SURF1, CCDC56
M. assamensis	NDUFS5, MRPS21, MRPL55, DARS2, MRPL2, UQCRQ, MRPL15, TMEM70, MRPS16, MRPL43, MRPS26, NDUFB3, BCS1L, UQCC3, TTC19, POLRMT

M. f. aurea dataset	
M. f. aurea	MRPS21, NDUFAF3, ATP5MF, UQCRQ, MRPL43, MRPL51, BCS1L, UQCC3, MRPL10, CCDC56, CARS2
M. fascicularis	NDUFAF3, MRPL18, RARS2, MRPL2, NDUFAF2, MRPS36, MRPS27, NDUFA2, CD14, NDUFAF6, CYC1, BCS1L, MRPL53, COX7C, MRPL49, NARS2, TMEM126B, MRPL10, CCDC56, NDUFA13, NDUFB10
M. thibetana	MRPL18, MRPL2, NDUFA2, CD14, MRPL49, LOC722212, MRPL10, CCDC56, NDUFB10
M. assamensis	ATP5MF, MRPL36, NDUFS6, UQCC3, MRPL12, CARS2, POLRMT, NDUFA13, MRPS34, NDUFB10

TABLE S9. The rate ratio of nonsynonymous to synonymous substitutions per site was individually significantly higher in Ninteract than non-Ninteract in two of four focal taxa. The observed difference in the mean dN/dS ratio of Ninteract and non-Ninteract genes (Observed) is positive for all comparisons. P is the proportion of permutations that had a difference between randomized means that was greater than the observed, with asterisks indicating individual significance. n(Ninteract) and n(non-Ninteract) are the numbers of genes that were included in the analysis after quality control.

Taxon	Observed	P	n(Ninteract)	n(non-Ninteract)
M. mulatta from China	0.046	0.115	156	2722
M. mulatta from India	0.042	0.093	157	2651
M. arctoides dataset	0.226	0.001*	134	2566
M. f. aurea dataset	0.128	0.016*	162	2699

TABLE S10. Lower π outliers for each population or species in each of the four datasets.

Dataset/F_{ST} comparison	p population	# p outliers	# F_{ST} outliers	# shared outliers	p proportion shared	F_{ST} proportion shared	Gene acronyms of shared outliers
M. mulatta from China							
Red : Blue	Red	12	19	3	0.25	0.16	MRPL49; NDUFB8; MRPL10
Red : Blue	Blue	9	19	2	0.22	0.11	MRPL49; NDUFB8
Red : Purple	Red	12	19	1	0.08	0.05	MRPL10
Red : Purple	Purple	11	19	1	0.09	0.05	HARS2
Blue : Purple	Blue	9	26	3	0.33	0.12	MRPL43; NDUFA7; NDUFB8
Blue : Purple	Purple	11	26	3	0.27	0.12	MRPL43; NDUFA7; NDUFB8
M. mulatta from India							
Brown : Orange	Brown	10	11	1	0.1	0.09	UQCRQ
Brown : Orange	Orange	8	11	2	0.25	0.18	NDUFB8; UQCRQ
Brown : Red	Brown	10	10	0	0	0	

Brown : Red	Red	13	10	8	0.62	0.8	SARS2; ATP5MF; MRPS12; LOC716161; MRPS34; MRPL28; NDUFB8; COX4I1
Orange : Red	Orange	8	11	1	0.12	0.09	NDUFB8
Orange : Red	Red	13	11	8	0.62	0.73	TMEM70; SARS2; ATP5MF; MRPS12; MRPS34; MRPL28; NDUFB8; COX4I1
M. arctoides dataset							
M arctoides : *M. assamensis*	*M. arctoides*	5	15	2	0.4	0.13	TYMP; ATP5I
M. arctoides : *M. assamensis*	*M. assamensis*	16	15	7	0.44	0.47	TTC19; TMEM70; POLRMT; NDUFB3; UQCC3; MRPS16; DARS2
M. arctoides : *M. fascicularis*	*M. arctoides*	5	16	3	0.6	0.19	MRPL51; TYMP; MRPL55
M. arctoides : *M. fascicularis*	*M. fascicularis*	13	16	6	0.46	0.38	NDUFA13; CYC1; UQCC3; CD14; HARS2; NDUFA2
M. arctoides : *M. mulatta*	*M. arctoides*	5	12	2	0.4	0.17	MRPL51; MRPL55
M. arctoides : *M. mulatta*	*M. mulatta*	17	12	9	0.53	0.75	MRPL34; ATP5J2; NDUFB8; MRPL55; CD14; MRPL43; TMEM126B; HARS2; NDUFA2
M. arctoides : *M. thibetana*	*M. arctoides*	5	9	3	0.6	0.33	MRPL51; MRPS34; ATP5I
M. arctoides : *M. thibetana*	*M. thibetana*	5	9	1	0.2	0.11	UQCC3
M. assamensis : *M. fascicularis*	*M. assamensis*	16	14	6	0.38	0.43	POLRMT; MRPL2; UQCC3; MRPL55; MRPS16; MRPL43
M. assamensis : *M. fascicularis*	*M. fascicularis*	13	14	9	0.69	0.64	MRPL10; POLRMT; USMG5; MRPL2; UQCC3; CD14; MRPL43; HARS2; NDUFA2
M. assamensis : *M. mulatta*	*M. assamensis*	16	16	7	0.44	0.44	UQCRQ; MRPL2; UQCC3; MRPL55; MRPS16; MRPL43; DARS2
M. assamensis : *M. mulatta*	*M. mulatta*	17	16	8	0.47	0.5	ATP5F1; USMG5; MRPL2; MRPL55; NDUFB1; MRPL43; TMEM126B; HARS2
M. assamensis : *M. thibetana*	*M. assamensis*	16	18	6	0.38	0.33	TMEM70; POLRMT; MRPL2; UQCC3; MRPS16; MRPL43
M. assamensis : *M. thibetana*	*M. thibetana*	5	18	3	0.6	0.17	MRPL2; CCDC56; UQCC3
M. mulatta : *M. fascicularis*	*M. mulatta*	17	22	9	0.53	0.41	MRPL37; ATP5J2; USMG5; MRPL2; NDUFB8; CD14; MRPL43; TMEM126B; NDUFA2
M. mulatta : *M. fascicularis*	*M. fascicularis*	13	22	8	0.62	0.36	MRPL10; MRPL37; NDUFA13; USMG5; MRPL2; CD14; MRPL43; NDUFA2
M. thibetana : *M. fascicularis*	*M. thibetana*	5	16	3	0.6	0.19	MRPL49; MRPL2; UQCC3
M. thibetana : *M. fascicularis*	*M. fascicularis*	13	16	13	1	0.81	MRPL49; MRPL10; POLRMT; MRPL37; NDUFA13; USMG5; MRPL2; CYC1; UQCC3; CD14; MRPL43; HARS2; NDUFA2
M. thibetana : *M. mulatta*	*M. thibetana*	5	15	3	0.6	0.2	MRPL49; MRPL2; UQCC3
M. thibetana : *M. mulatta*	*M. mulatta*	17	15	12	0.71	0.8	ATP5F1; MRPL34; NDUFV1; MRPL2; MRPL9; NDUFB8; MRPL55; CD14; MRPL43; TMEM126B; HARS2; NDUFA2
M. f. aurea dataset							
M. f. aureus : *M. thibetana*	*M. f. aureus*	11	10	8	0.73	0.8	MRPL10; MRPS21; ATP5MF; CCDC56; NDUFAF3; UQCRQ; UQCC3; BCS1L
M. f. aureus : *M. thibetana*	*M. thibetana*	9	10	2	0.22	0.2	CCDC56; MRPL10
M. f. aureus : *M. fascicularis*	*M. f. aureus*	11	19	6	0.55	0.32	MRPL10; MRPS21; CCDC56; NDUFAF3; UQCRQ; BCS1L
M. f. aureus : *M. fascicularis*	*M. fascicularis*	21	19	8	0.38	0.42	MRPL10; CCDC56; RARS2; NDUFAF3; NARS2; TMEM126B; BCS1L; MRPL53
M. f. aureus : *M. assamensis*	*M. f. aureus*	11	14	8	0.73	0.57	CARS2; MRPL10; MRPS21; ATP5MF; NDUFAF3; UQCRQ; UQCC3; BCS1L
M. f. aureus : *M. assamensis*	*M. assamensis*	10	14	5	0.5	0.36	CARS2; ATP5MF; NDUFA13; UQCC3; MRPL12
M. assamensis : *M. thibetana*	*M. assamensis*	10	15	4	0.4	0.27	CARS2; NDUFA13; MRPS34; ATP5MF
M. assamensis : *M. thibetana*	*M. thibetana*	9	15	0	0	0	

M. fascicularis : *M. thibetana*	*M. fascicularis*	21	13	11	0.52	0.85	MRPL49; MRPL10; COX7C; NDUFA13; CCDC56; NDUFAF3; MRPL2; CYC1; CD14; BCS1L; NDUFA2
M. fascicularis : *M. thibetana*	*M. thibetana*	9	13	6	0.67	0.46	MRPL49; MRPL10; CCDC56; MRPL2; CD14; NDUFA2
M. fascicularis : *M. assamensis*	*M. fascicularis*	21	18	12	0.57	0.67	MRPL49; MRPL10; COX7C; RARS2; NDUFA13; NDUFAF3; MRPL2; CYC1; CD14; NARS2; BCS1L; NDUFA2
M. fascicularis : *M. assamensis*	*M. assamensis*	10	18	2	0.2	0.11	NDUFA13; UQCC3

TABLE S11. Results of analyses of introgression using Patterson's *D* statistic. Analyses were performed for the *M. arctoides* and the *M. f. aurea* datasets and using 30 or 100kb genomic windows. Values of Patterson's *D* are presented for the entire genome (Genome-wide), for Ninteract and non-Ninteract windows, and for non-genic windows based on configurations of taxa (P1:P2:P3:O) with species abbreviated using the first three letters of the species name. The test statistic for permutation tests is the difference between the mean Patterson's *D* statistic in Ninteract and non-Ninteract windows (Difference) and an asterisk indicates an individually significant P values (P) based on a permutation test.

M. arctoides dataset							
Window size	P1: P2: P3: O	Genome-wide	Ninteract windows	non-Ninteract windows	Non-genic windows	Difference	P
30k	arc:ass:mul:nem	-0.10435	-0.038	-0.1	-0.105	0.062371	0.907
	arc:thi:mul:nem	-0.11456	0.016656	-0.021486	-0.115	0.132571	0.989
100k	arc:ass:mul:nem	-0.124274	-0.122	-0.119	-0.128	-0.003	0.428
	arc:thi:mul:nem	-0.113505	-0.074	-0.112	-0.115	0.038	0.857
M. f. aureus dataset							
30k	fas:aur:ass:nem	0.084386	0.057042	0.078528	0.085693	-0.021	0.758
	fas:aur:thi:nem	0.083	0.071107	0.078071	0.084022	-0.007	0.599
100k	fas:aur:ass:nem	0.102535	0.078892	0.096135	0.107086	-0.017	0.692
	fas:aur:thi:nem	0.101548	0.073516	0.09603	0.105563	-0.023	0.758

References

Bouckaert, R., Vaughan, T., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., Maio, N., Matschiner, M., Mendes, F., Müller, N., Ogilvie, H., Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., and Drummond, A. (2019). BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology* 15, e1006650.

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty:de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45, gkw955.

Dobson, F. (Nov. 1982). Competition for mates and predominant juvenile male dispersal in mammals. *Animal Behaviour* 30(4), 1183–1192.

Evans, B., Tosi, A., Zeng, K., Dushoff, J., Corvelo, A., and Melnick, D. (Oct. 2017). Speciation over the edge: gene flow among non-human primate species across a formidable biogeographic barrier. *Royal Society Open Science* 4(10), 170351.

Fischer, J., Higham, J., Alberts, S., Barrett, L., Beehner, J., Bergman, T., Carter, A., Collins, A., Elton, S., Fagot, J., Ferreira Da Silva, M., Hammerschmidt, K., Henzi, P., Jolly, C., Knauf, S., Kopp, G., Rogers, J., Roos, C., Ross, C., Seyfarth, R., Silk, J., Snyder-Mackler, N., Staedele, V., Swedell, L., Wilson, M., and Zinner, D. (Nov. 12, 2019). Insights into the evolution of social systems and species from baboon studies. *eLife* 8, e50989.

Greenwood, P. (Nov. 1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour* 28(4), 1140–1162.

Li, X.-Y. and Kokko, H. (2019). Sex-biased dispersal: a review of the theory. *Biological Reviews* 94(2), 721–736.

Minh, B., Nguyen, M., and Von Haeseler, A. (Feb. 15, 2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30(5), 1188–1195.

Mootha, V., Bunkenborg, J., Olsen, J., Hjerrild, M., Wisniewski, J., Stahl, E., Bolouri, M., Ray, H., Sihag, S., Kamal, M., Patterson, N., Lander, E., and Mann, M. (Nov. 2003). Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* 115(5), 629–640.

Nguyen, L.-T., Schmidt, H., Von Haeseler, A., and Minh, B. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32(1), 268–274.

Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H., Schönfisch, B., Perschil, I., Chacinska, A., Guiard, B., Rehling, P., Pfanner, N., and Meisinger, C. (Oct. 23, 2003). The proteome of Saccharomyces cerevisiae mitochondria. *Proceedings of the National Academy of Sciences* 100(23), 13207–13212.

Smits, P., Smeitink, J., and Van Den Heuvel, L. (2010). Mitochondrial translation and beyond: Processes implicated in combined oxidative phosphorylation deficiencies. *Journal of Biomedicine and Biotechnology* 2010, 1–24.

Tosi, A. and Coke, C. (2002). Comparative phylogenetics offer new insights into the biogeographic history of Macaca fascicularis and the origin of the Mauritian macaques. *Molecular Phylogenetics and Evolution* 42(2), 498–504.

Tosi, A., Morales, J., and Melnick, D. (2003). Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution* 57(6), 1419.

Appendix B

Subgenome evolution of an allotetraploid *X. mellotropicalis*

This appendix summarizes progress on an as yet unfinished project.

B1 Introduction

Whole genome duplication (WGD) is a relatively common genomic transformation that occurred at key points during diversification of many groups, including animals, plants, and fungi (Song et al. 2012). WGD can increase genome complexity and catalyze phenotypic evolution, and is considered an important factor in speciation (Moriyama and Koshiba-Takeuchi, 2018). WGD happens via two mechanisms. The first is autopolyploidization wherein genome duplication happens within a single ancestral species, and the second is allotetraploidization, wherein genome duplication happens in association with hybridization among species (Chenuil et al. 1999).

WGD occurred remarkably frequently in African clawed frog of the genus *Xenopus* (Tymowska and Fischberg, 1973). *Xenopus* consists of two subgenera *Xenopus* and *Silurana*. Among 29 extant known species in the genus *Xenopus*, *Xenopus tropicalis* (Gray, 1864) is the only diploid derived from a single ancestor (2n = 20, where n refers to the number of chromosomes in a gamete). The other species are all either allotetraploid, allooctoploid, or allododecaploid (Evans et al. 2015). *Xenopus tropicalis* is a member of the subgenera *Silurana*, together with three other allotetraploid species (Evans et al. 2015; Fischberg et al. 1982).

Subgenomes – the compartments of an allopolyploid genome that are derieved from each ancestral species – may evolve asymmetrically. Assymetric subgenome evolution is evidenced in *Xenopus* by genomic analysis of the allotetraploid species *X. laevis* (Session et al. 2016). For example, compared to the L subgenome, the S subgenome of *X. laevis* underwent more rearrangements, more pseudogenization, and is under less stringent purifying selection (Session et al. 2016, Furman et al. 2018).

Xenopus mellotropicalis is one of the allotetraploid species in subgenus *Silurana*. Cytogenetic comparative analysis reveals that one subgenome of *X. mellotropicalis* is

closely related to *X. tropicalis* (Knytl et al. 2017). Phylogenetic analysis based on mitochondria sequences also indicates that these two species have a close relationship (Evans et al. 2019). However, whether, in what way, and how the subgenomes of *X. mellotropicalis* may evolve assymetrically remains unknown. To better understand allopolyploid evolution in general, we are studying the genome of *X. mellotropicalis* with an aim of testing whether this genome has a pattern of asymmetric subgenome evolution that is similar to *X. laevis*.

B2 Objectives

We aim to explore whether subgenome evolution in *X. mellotropicalis* is asymmetric in terms of protein evolution (dN/dS), gene silencing, and transposible element mobility and content. To achieve these goals, we attempted to generate i) a high quality *de novo* assembly of this species and then use this ii) to determine which contigs belongs to which subgenome. To facilitate the delineation of subgenomes of *X. mellotropicalis*, we will also evaluate the species status of *X. "Liberia"*, which is a diploid taxon from Liberia that either is *X. tropicalis* or is closely related to this species.

B3 Methods and Results

B3.1 *X. mellotropicalis* genome assembly

We used an unpublished approach to generate paired-end reads with a large insert sizes to facilitate assembling. A *de novo* assembly of was performed using 10X Supernova software (Weisenfeld et al. 2017). The assembly contains 770,810 contigs and the longest one is 653,197 bp. The N50 statistic is 17,380 bp and the GC proportion is 39.89%.

In the following analyses, we exclude contigs whose length is less than 5,000 bp, leaving a total of 88,599 contigs. We use MUMmer v3.2.3 (Kurtz et al. 2004) to map *X. mellotropicalis* contigs to the *X. tropicalis* genome assembly version 10 (GCA_000004195.4) and we then used samtools v1.15.1 (Danecek et al. 2021) to calculate coverage statistics. In non repetitive regions and barring deletions, we expected each region of the *X. tropicalis* genome to match orthologous sequences from each subgenome of *X. mellotropicalis*. Figure 1 shows the depth by site plot in chromosome 7, which is the sex chromosome in *X. tropicalis*. This plot illustrates that some regions have far higher coverage than the genome-wide mean, which suggests that they are highly repetitive in *X. mellotropicalis*.

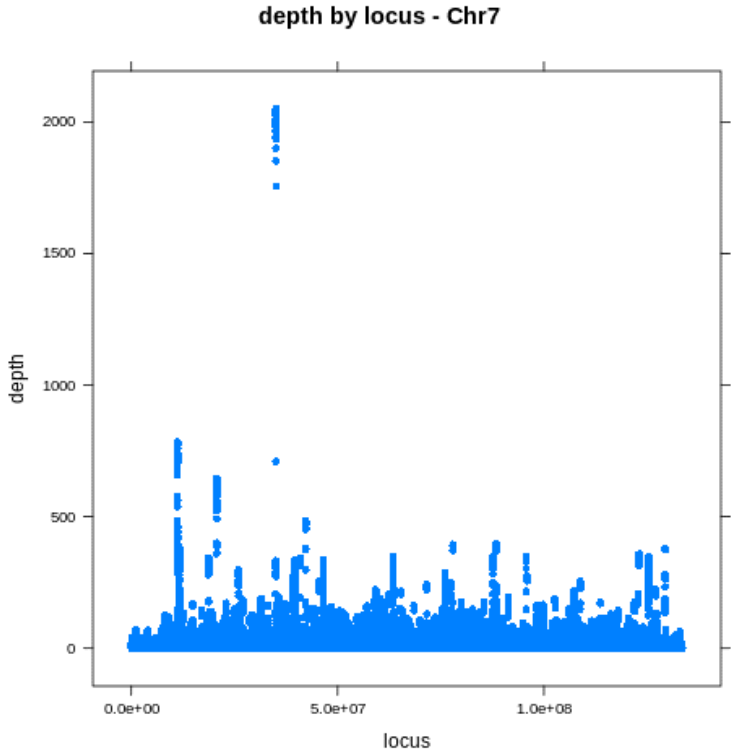

FIG. S2.1. Depth by locus plot in *X. tropicalis* chromosome 7. The x-axis is the site of bases along chromosome 7 and the y-axis represents the number of contigs covering this site.

To further evaluate the effects of highly repetitive sequences, we tabulated how many hits each *X. mellotropicalis* contig had when it was mapped to the *X. tropicalis* genome assembly. Figure 2 shows the distribution of these hits. Most *X. mellotropicalis* contigs have only one hit, but some have multiple hits; those with multiple hits presumably contain sequences that are repetitive in the *X. tropicalis* genome.

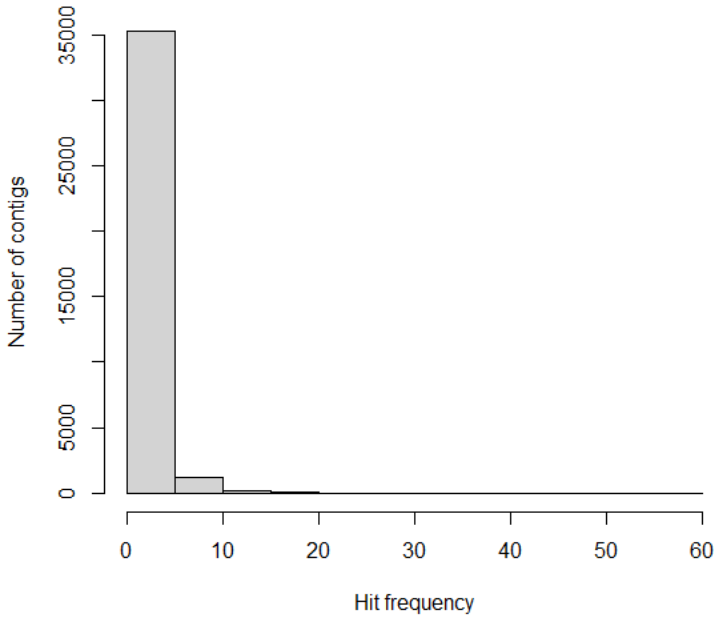

FIG. S2.2. A histogram of the distribution of of hits to the *X. tropicalis* genome for *X. mellotropicalis* contigs.

For subsequent analyses described below, we decided to use contigs larger than 50,000 bp. This set contains 8,809 contigs in total.

B3.2 *X. "Liberia"* assembly

Because our attempts to construct a chromosome-scale genome assembly were not successful, we instead attempted to use phylogenetic information to determine from which subgenome each *X. mellotropicalis* contig originated. We leveraged information from a putatively undescribed diploid *Xenopus* species, hereafter *X. "Liberia"* (Evans et al., 2019) along with genomic data from *X. tropicalis* and *X. laevis*.

To test whether *X. "Liberia"* is indeed a different species, we first used RRGS data to determine if *X. "Liberia"* is substantially diverged from *X. tropicalis*. We combined this *X. "Liberia"* sample with 31 *X. tropicalis* samples and performed PCA analysis with R package SNPRelate v1.6.4 (Zheng et al., 2012). Fig 3 shows the result of PCA, we can see that *X. "Liberia"* does not cluster with other *X. tropicalis* samples, which is preliminarily consistent with it belonging to a separate species.

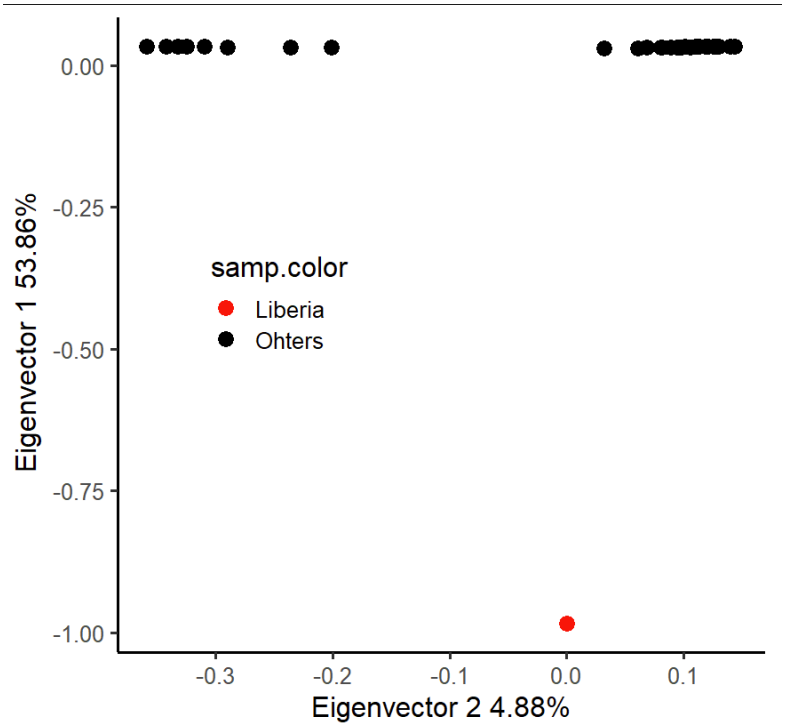

FIG. S2.3. PCA results for the combined data. The Liberia sample is marked red.

The result of PCA analysis motivated us to perform a larger scale data collection and analysis. The whole genome sequencing (WGS) data was generated by Illumina NovaSeq 6000 S4 flowcell, with pair-end 150 bp reads. We use fastqc v0.11.9 (Andrews, 2010) to do the quality control. Total number of reads is 315,832,366 and the proportion of GC contents is 44%. The overall sequencing quality was excellent.

We trimmed the raw sequences with Trimmomatic v0.39 (Bolger et al. 2014) with Illumina Tru-Seq2 adaptor. DNA sequences were mapped by BWA-mem v0.7.17 (Li, 2013). PCR duplicates were removed by Picard v2.26.3 (Broad Institute, 2019) and variants were called with GATK v4.2.4.0 (Van and O'Connor, 2020). We also filtered variants with GATK toolkits to keep only biallelic SNPs with mapping quality >= 30.

To facilitate assembly of *X. "Liberia"* short read data, we use bbnorm v38.86 (Bushnell, 2014) to normalize coverage. We then used SPAdes v3.15.4 (Bankevich et al., 2012) to assemble the genome. K-mers was set to 21,33,49 and all other settings were kept default. The total number of contigs is 692,866. The proportion of GC contents is 39.9% and N50 is 127,313. We then used Ragtag v2.1.0 (Alonge et.al 2021) to improve the assembly quality. Due to a close evolutionary relationship, the *de novo* assembly of X. "Liberia" was considerably improved. In the new assembly, the total number of contigs is 130,000 and N50 is 87,795,884 (n=5). We calculated the average nucleotide identity (ANI) between assembled *X. "Liberia"* chromosomes and the *X. tropicalis* genome with

fastANI v1.33 (Jain et al., 2018). Individuals who have ANI >= 95% are typically from the same species (Goris et al., 2007). Table 1 shows the result. We can see some evidence to support *X. "Liberia"* is a new species.

TABLE S2.1. ANI between *X. "Liberia"* chromosomes and the *X. tropicalis* genome. The value here is shown in percentage.

	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10
Similarity	94.9	94.8	94.9	95	94.8	94.8	94.9	95	95	95.2

To evaluate whether our reference-based assembly might exaggerates the estimates of similarity to *X. tropicalis*, we used WGS data from another *X. tropicalis* individual that was sampled in Ghana (hereafter Trop_Ghana). The *de novo* assembly of this individual contains 1,579,735 contigs and N50 is 2,655. After reference-based correction, the total number of contigs is 861,558 and N50 is 125,138,899 (n=5).

TABLE S2.2. ANI between Trop_Ghana chromosomes and the *X. tropicalis* reference genome. The value here is shown in percentage.

	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10
Similarity	98.2	98.2	98.2	98.2	98.1	98.2	98	98.1	98	97.9

TABLE S2.3. Statistics of the *X. "Liberia"* assemblies

Statistics without reference	Liberia raw assembly	Liberia corrected assembly
# contigs	493526	34365
# contigs (>= 10000 bp)	912	28
# contigs (>= 50000 bp)	0	10
Largest contig	25300	124025852
Total length	753962296	843349271
N50	1884	87795884
N75	1090	76004098
L50	115119	5
L75	247605	7
GC (%)	39.92	39.78

TABLE S2.4. Statistics of the Trop_Ghana assemblies

Statistics without reference	Trop_Ghana raw assembly	Trop_Ghana corrected assembly
# contigs	483032	29690
# contigs (>= 10000 bp)	6254	28
# contigs (>= 50000 bp)	2	11
Largest contig	84627	180210507
Total length	1046382448	1186201617
N50	3215	125138899
N75	1638	105494856
L50	92853	5
L75	207005	7
GC (%)	39.16	39.27

B3.3 Comparisons between *X. tropicalis*, *X. "Liberia"*, *X. mellotropicalis*

We attempted to evaluate evolutionary relationships among each of the *X. mellotropicalis* contigs with respect to the *X. tropicalis* reference genome, and the Trop_Ghana and *X. "Liberia"* genome assemblies. We used the L subgenome of *X. laevis* reference genome v10.1 (GCA_017654675.1) as an outgroup. We first used minimap2 v2.24 (Li , 2018) to map *X. mellotropicalis* contigs to each of the rest 4 assemblies. We only kept hits with the highest mapping quality. For each *X. mellotropicalis* contig, we gathered all coordinates from its alignment against the other 4 genomes, and then we split those segments with samtools v1.15.1 (Danecek et al., 2021). Multiple alignment between these five segments was done by mafft v7.471 (Katoh et al., 2002). We used Gblocks v0.91b (Castresana, 2000) to trim alignments with low quality. Finally, we used FastTree v2.1.11 (Price et al., 2010) to calculate neighbour-joining trees and Emboss v6.6.0 (Rice et al., 2000) to calculate the distance matrixes.

Not all *X. mellotropicalis* contigs were able to be mapped to *X. laevis* with high confidence. We only retrieved the phylogeny of 3271 contigs. We summarized the topologies of those trees and the result is shown below:

((((Trop,TropGhana),Liberia),mello),LaevisL); number: 2534;

((((Trop,TropGhana),mello),Liberia),LaevisL); number: 371;

(((mello,Liberia),(Trop,TropGhana)),LaevisL); number: 178;

((((Trop,Liberia),TropGhana),mello),LaevisL); number: 22;

((((Trop,Liberia),mello),TropGhana),LaevisL); number: 7;

We can see that *X. "Liberia"* is usually closer to *X. tropicalis* in most cases. This argues against the utility of the *X. "Liberia"* for determining from which subgenome each *X. mellotropicalis* contig is derived. However, this summary hides the information about branch length. To retrieve this information, we summarize the distribution of

each comparison from the distance matrix. The result was shown in Fig 4. The distance was measured in estimated substitution per 100 bases corrected by Jukes-Cantor method (Jukes and Cantor, 1992).

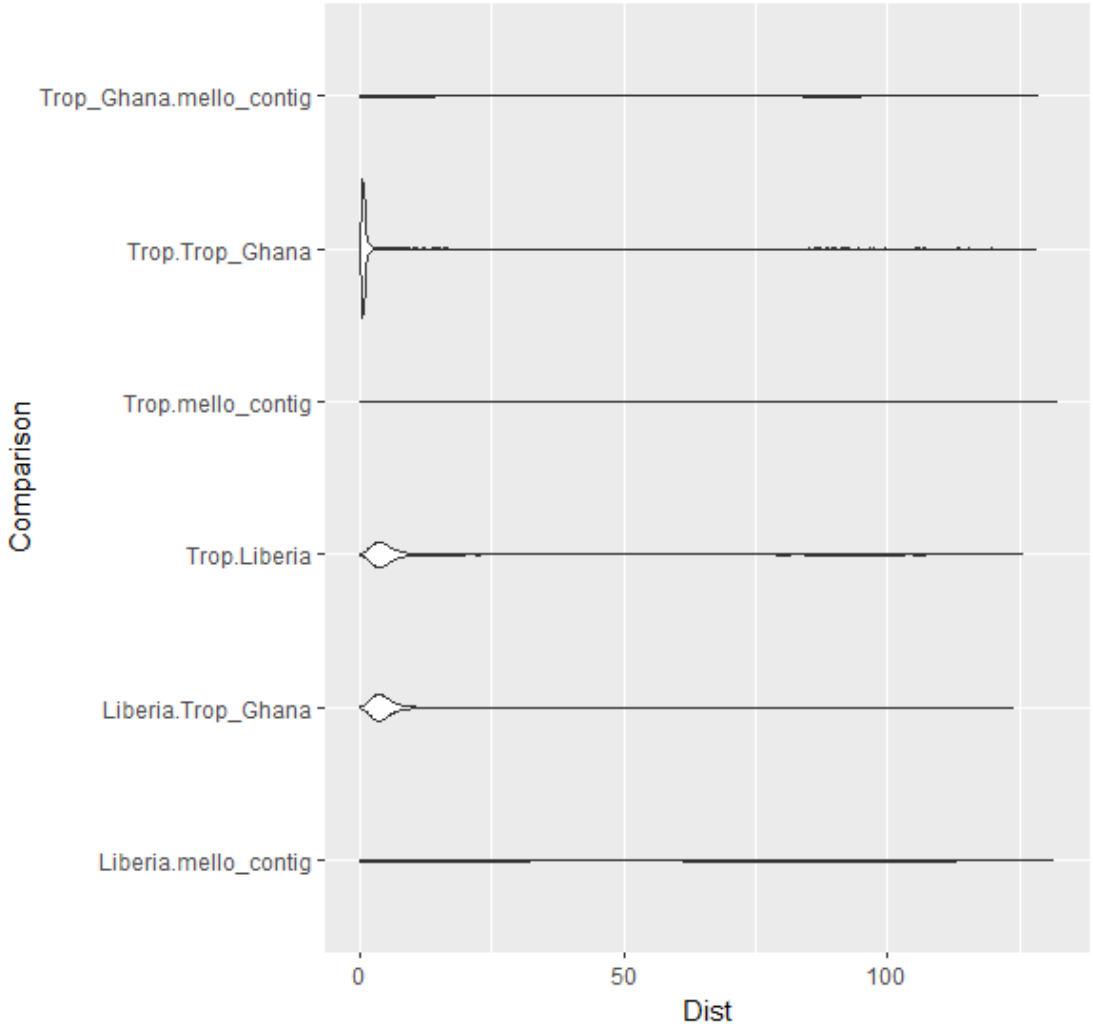

FIG. S2.4. Distribution of pair-wise distance between all comparisons. The name of comparisons are listed in the y-axis and assemblies were separated by a dot sign. The distance was shown in the x-axis.

Although all of those comparisons have long tails, these distributions indicate the following phylogenies: i) the reference genome and Trop_Ghana assembly of *X. tropicalis* are closely related to each other. ii) *X. "Liberia"* is closely related to *X. tropicalis*, iii) distances between *X. mellotropicalis* contigs and *X. "Liberia"* assembly have 2 peaks whereas this pattern is not obvious in comparisons between *X. tropicalis*. The first two points are what we expected to see but point 3 is hard to interpret. We expect to see a

similar double-peak distribution between the *X. tropicalis* genome and *X. mellotropicalis* contigs. It is possible that the double peak we see in *X. "Liberia"* is due to error and we need further analysis to explore this possibility.

B4 Future direction

Results shown in the previous section provide preliminary support that *X. "Liberia"* is a new species. Future work will focus on:

1. Additional analyses of genomic data from more geographical isolates of *X. tropicalis* and *X. "Liberia"*.

2. Further exploration of the use of genomic data from *X. tropicalis* and *X. "Liberia"* to assign *X. mellotropicalis* contigs to different subgenomes

3. Depending on the success of point 2 above, we aim to characterize and compare several aspects of subgenome evolution, including transposible element composition, protein evolution, and expression levels.

References

Alonge, M., Lebeigle, L., Kirsche, M., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., and Soyk, S. (Nov. 2021). *Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing*. en. Tech. rep. Type: article. bioRxiv. Chap. New Results, 2021.11.18.469135.

Andrews, S. (2010). *FASTQC. A quality control tool for high throughput sequence data*.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (May 2012). SPAdes: A neww genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19(5), 455–477.

Bolger, A. M., Lohse, M., and Usadel, B. (Aug. 2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15), 2114–2120. ISSN: 1367-4803.

BroadInstitute (2019). *Picard*.

Bushnell, B. (Mar. 2014). *BBMap: A fast, accurate, splice-aware aligner*. English. Tech. rep.

Castresana, J. (Apr. 2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17(4), 540–552. ISSN: 0737-4038.

Cauret, C. M., Gansauge, M.-T., Tupper, A. S., Furman, B. L., Knytl, M., Song, X.-Y., Greenbaum, E., Meyer, M., and Evans, B. J. (n.d.). Developmental systems drift and the drivers of sexChromosome evolution ().

Chenuil, A., Galtier, N., and Berrebi, P. (Apr. 1999). A test of the hypothesis of an autopolyploid vs. allopolyploid origin for a tetraploid lineage: application to the genus *Barbus* (*Cyprinidae*). en. *Heredity* 82(4), 373–380. ISSN: 1365-2540.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., and Li, H. (Feb. 2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10(2), giab008. ISSN: 2047-217X.

Evans, B. J., Carter, T. F., Greenbaum, E., Gvoždík, V., Kelley, D. B., McLaughlin, P. J., Pauwels, O. S. G., Portik, D. M., Stanley, E. L., Tinsley, R. C., Tobias, M. L., and Blackburn, D. C. (Dec. 2015). Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African clawed frog (*Xenopus Pipidae*) from west and central Africa. en. *PLOS ONE* 10(12), e0142823. ISSN: 1932-6203.

Evans, B. J., Gansauge, M.-T., Stanley, E. L., Furman, B. L. S., Cauret, C. M. S., Ofori-Boateng, C., Gvoždík, V., Streicher, J. W., Greenbaum, E., Tinsley, R. C., Meyer,

M., and Blackburn, D. C. (Sept. 2019). *Xenopus fraseri*: Mr. Fraser, where did your frog come from? en. *PLOS ONE* 14(9), e0220892. ISSN: 1932-6203.

Gray, J. (Oct. 1864). Notice of a new genus (Silurana) of frogs from west Africa. *Annals and Magazine of Natural History* 14(82), 315–316. ISSN: 0374-5481.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (Nov. 2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. en. *Nature Communications* 9(1), 5114. ISSN: 2041-1723.

Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T. (July 2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30(14), 3059–3066. ISSN: 0305-1048.

Knytl, M., Smolík, O., Kubíčková, S., Tlapáková, T., Evans, B. J., and Krylov, V. (May 2017). Chromosome divergence during evolution of the tetraploid clawed frogs, *Xenopus mellotropicalis* and *Xenopus epitropicalis* as revealed by Zoo-FISH. en. *PLOS ONE* 12(5), e0177087. ISSN: 1932-6203.

Li, H. (May 2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. Tech. rep. arXiv:1303.3997 [q-bio] type: article. arXiv.

Li, H. (Sept. 2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18), 3094–3100. ISSN: 1367-4803.

Moriyama, Y. and Koshiba-Takeuchi, K. (Sept. 2018). Significance of whole-genome duplications on the emergence of evolutionary novelties. *Briefings in Functional Genomics* 17(5), 329–338. ISSN: 2041-2657.

Price, M. N., Dehal, P. S., and Arkin, A. P. (Mar. 2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. en. *PLOS ONE* 5(3), e9490. ISSN: 1932-6203.

Rice, P., Longden, I., and Bleasby, A. (June 2000). EMBOSS: The European molecular biology open software suite. English. *Trends in Genetics* 16(6), 276–277. ISSN: 0168-9525.

Segerman, B. (2020). The most frequently used sequencing technologies and assembly methods in different time segments of the bacterial surveillance and RefSeq genome databases. *Frontiers in Cellular and Infection Microbiology* 10. ISSN: 2235-2988.

Session, A. M., Uno, Y., Kwon, T., Chapman, J. A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., Heeringen, S. J. van, Quigley, I., Heinz, S., Ogino, H., Ochi, H., Hellsten, U., Lyons, J. B., Simakov, O., Putnam, N., Stites, J., Kuroki, Y., Tanaka, T., Michiue, T., Watanabe, M., Bogdanovic, O., Lister, R., Georgiou, G., Paranjpe, S. S., Kruijsbergen, I. van, Shu, S., Carlson, J., Kinoshita, T., Ohta, Y., Mawaribuchi, S., Jenkins, J., Grimwood, J., Schmutz, J., Mitros, T., Mozaffari, S. V., Suzuki, Y., Haramoto, Y., Yamamoto, T. S., Takagi, C., Heald, R., Miller, K., Haudenschild, C., Kitzman, J., Nakayama, T., Izutsu, Y., Robert, J., Fortriede, J., Burns, K., Lotay, V., Karimi, K., Yasuoka, Y., Dichmann, D. S., Flajnik, M. F., Houston, D. W., Shendure, J., DuPasquier, L., Vize, P. D., Zorn, A. M., Ito, M., Marcotte, E. M., Wallingford, J. B., Ito, Y., Asashima, M., Ueno, N., Matsuda, Y., Veenstra, G. J. C., Fujiyama, A., Harland, R. M., Taira, M., and Rokhsar, D. S. (Oct. 2016). Genome evolution in the allotetraploid frog *Xenopus laevis*. en. *Nature* 538(7625), 336–343. ISSN: 1476-4687.

Song, C., Liu, S., Xiao, J., He, W., Zhou, Y., Qin, Q., Zhang, C., and Liu, Y. (Apr. 2012). Polyploid organisms. en. *Science China Life Sciences* 55(4), 301–311. ISSN: 1869-1889.

Tamura, K. (July 1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution* 9(4), 678–687. ISSN: 0737-4038.

Tymowska, J. and Fischberg, M. (Sept. 1973). Chromosome complements of the genus *Xenopus*. en. *Chromosoma* 44(3), 335–342. ISSN: 1432-0886.

Van der Auwera GA, O. B. (2020). *Genomics in the cloud: Using Docker, GATK, and WDL in Terra (1st Edition)*.

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., and Jaffe, D. B. (Jan. 2017). Direct determination of diploid genome sequences. en. *Genome Research* 27(5), 757–767. ISSN: 1088-9051, 1549-5469.