

Secondary contact and genomic admixture between rhesus and long-tailed macaques in the Indochina Peninsula

Running title: Genomic admixture in macaques

Tsuyoshi Ito^{1,†}, Sreetharan Kanthaswamy^{2,†}, Srichan Bunlungsup^{3,4}, Robert Oldt², Paul Houghton⁵, Yuzuru Hamada¹, Suchinda Malaivijitnond^{3,4,†}

¹ Department of Evolution and Phylogeny, Primate Research Institute, Kyoto University, Kanrin 41-2, Inuyama, Aichi 484-8506, Japan

² School of Mathematical and Natural Sciences, New College of Interdisciplinary Arts and Sciences, Arizona State University West Campus, 4701 W Thunderbird Road, Glendale, AZ 85306-4908, USA

³ Department of Biology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

⁴ National Primate Research Center of Thailand-Chulalongkorn University, Saraburi 18100, Thailand

⁵ Primate Products, Inc., Immokalee, Florida, USA

[†] Correspondence authors:

TI: ito.tsuyoshi.3a@kyoto-u.ac.jp, +81-568-63-0523 (Ph), +81-568-61-5775 (Fax)

SK: skanthas@asu.edu, +1-602-543-3405 (Ph)

SM: suchinda.m@chula.ac.th, +66-2-2185275 (Ph & Fax)

Author contributions

TI, SK, SB, YH, and SM conceived and designed the research. SK, RO, SB, PH, and SM prepared and provided the samples. TI analyzed the data and drafted the manuscript with contributions by the other authors. All authors approved the final version of this manuscript.

Acknowledgments

We would like to thank the members of the Primate Research Unit of Chulalongkorn University for their help and kindness to TI and the owners and staff of Primate Products for the collection and processing of contributed samples. We also thank the Genomic Sequencing and Analysis Facility at the University of Texas for their support with the library preparation and sequencing and the McDonnell Genome Institute, Washington University, St Louis, MO, for making the Mmul_10 assembly sequences available. This work was partly supported by the Thailand Research Fund-Chinese Academy of Science (DBG6080008), the TSRI Research Team Promotion Grant (RTA6280010), the Keihanshin Consortium for Fostering the Next Generation of Global Leaders in Research (K-CONNEX), established by Human Resource Development Program for Science and Technology, MEXT, and the JSPS Grants-in-Aid for Scientific Research (17K15195 to TI and 19K06865 to YH). US National Institutes of Health (NIH) grants: P51 OD011107, 2U42 OD010990, and 5U42 OD021458 also supported SK's contributions to this study.

Abstract

Understanding the process and consequences of hybridization is one of the major challenges in evolutionary biology. A growing body of literature has reported evidence of ancient hybridization events or natural hybrid zones in primates, including humans; however, we still have relatively limited knowledge about the pattern and history of admixture because there have been little studies that simultaneously achieved genome-scale analysis and a geographically wide sampling of wild populations. Our study applied double-digest restriction site-associated DNA sequencing to samples from the six localities in and around the provisional hybrid zone of rhesus and long-tailed macaques and evaluated population structure, phylogenetic relationships, demographic history, and geographic clines of morphology and allele frequencies. A latitudinal gradient of genetic components was observed, highlighting the transition from rhesus (north) to long-tailed macaque distribution (south) as well as the presence of one northern population of long-tailed macaques exhibiting unique genetic structure. Interspecific gene flow was estimated to have recently occurred after an isolation period, and the migration rate from rhesus to long-tailed macaques was slightly greater than in the opposite direction. Although some rhesus macaque-biased alleles have widely introgressed into long-tailed macaque-populations, the inflection points of allele frequencies have been observed as concentrated around the traditionally recognized interspecific boundary where morphology discontinuously changed; this pattern was more pronounced in the X-chromosome than in autosomes. Thus, due to geographic separation before secondary contact, reproductive isolation could have evolved, contributing to the maintenance of an interspecific boundary and species-specific morphological characteristics.

Keywords

hybridization, Indochina, RAD-seq, reproductive isolation, speciation

1. Introduction

Historically, hybridization has been considered rare in animals, but recent molecular studies have revealed that hybridization is rampant in both captivity and nature (Mallet, 2005; Taylor & Larson, 2019). Primates, including humans, are no exception, and a growing body of

literature has reported evidence of ancient hybridization events and natural hybrid zones associated with various primate taxa (Arnold & Meyer, 2006; Cortés-Ortiz *et al.*, 2007, 2019; Zinner *et al.*, 2009, 2011; Ackermann & Bishop, 2010; Roos *et al.*, 2011; Prüfer *et al.*, 2014; Malukiewicz *et al.*, 2015; Svardal *et al.*, 2017). Hybridization causes genetic introgression, which is not always maladaptive, that could be a fundamental source of evolutionary novelty and phenotypic diversity (Barton, 2001; Seehausen, 2004; Bell & Travis, 2005; Parnell *et al.*, 2008; Parsons *et al.*, 2011; Genner & Turner, 2012; Abbott *et al.*, 2013, 2016; Soltis, 2013; Pereira *et al.*, 2014; Simonti *et al.*, 2016; Svensson *et al.*, 2016; Arnold & Kunte, 2017; Meier *et al.*, 2017; Taylor & Larson, 2019). Thus, hybridization has been recognized as one of the most intriguing topics in evolutionary biology (Abbott *et al.*, 2016).

Rhesus macaques (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*), also known as cynomolgus or crab-eating macaques, are closely-related species that are widely distributed in Asia (Fooden, 2006). These species are also two of the most commonly-used monkey models in experimental and biomedical studies (Sibal & Samson, 2001; Osuna *et al.*, 2016). The natural distribution of rhesus macaques ranges from Afghanistan to China and in the northern part of Indochina (Fooden, 2000), while long-tailed macaques are found in southern Indochina, Sumatra, Borneo, and the Philippines (Fooden, 1995; Malaivijitnond & Hamada, 2008). The two species can be distinguished by their relative tail length (the ratio of tail length and head-body length) and the color pattern on their backs. Rhesus macaques have a relative tail length of approximately 0.4 and tend to have a bipartite back coat that is grayish-brown anteriorly and tawny on the rump; long-tailed macaques have a relative tail length of approximately 1.1 and a back coat that is more or less uniformly colored, ranging from pale brown to dark brown (Fooden, 1964, 2006; Hamada *et al.*, 2005, 2007, 2008). The geographical distributions of the two species meet at approximately 17°N (Fig. 1), where they appear to produce a natural hybridization (Fooden, 2006).

Evidence of the hybridization of the two species has been reported in both morphological and molecular studies. The first reports were morphological studies (Fooden, 1964), where some specimens along the boundary line of distributions showed an intermediate relative tail length between the two species, suggesting that they were hybrids (Fooden, 1997).

Dorsal pelage color, lateral facial crest pattern, and head-body and skull lengths in Indochinese long-tailed macaques are somewhat similar to rhesus macaques, supporting the existence of hybridization between the two species (Fooden, 1995, 1997). Molecular studies revealed that the blood protein frequency of the Indochinese long-tailed macaques was more similar to that of rhesus macaques than that of non-Indochinese (Philippine) long-tailed macaques, although it was initially interpreted as the consequence of symplesiomorphies (Melnick *et al.*, 1985). Tosi *et al.* (2002) compared Y-chromosome data with mitochondrial DNA (mtDNA) markers and found that Indochinese long-tailed macaques clustered more closely with rhesus macaques than with non-Indochinese long-tailed macaques, implying that male-mediated gene flow from rhesus to Indochinese long-tailed macaques had occurred.

Recent studies based on large numbers of genetic markers and/or geographical sampling sites have revealed a more detailed picture of gene introgression between the two species (Street *et al.*, 2007; Kanthaswamy *et al.*, 2008, 2010; Malaivijitnond *et al.*, 2008; Bonhomme *et al.*, 2009; Stevison & Kohn, 2009; Osada *et al.*, 2010; Barr *et al.*, 2011; Yan *et al.*, 2011; Satkoski Trask *et al.*, 2013; Jadejaroen *et al.*, 2016; Bunlungsup *et al.*, 2017b; a; Oldt *et al.*, 2019). Restricted or whole-genome data supported Tosi *et al.*'s (2002) suggestion that gene introgression was biased toward the direction from rhesus to long-tailed macaques (Bonhomme *et al.*, 2009; Stevison & Kohn, 2009; Yan *et al.*, 2011). It was also suggested that ancient gene introgressions occurred far beyond the traditionally recognized area of introgression, i.e., a zone between the (morphology-based) interspecific boundary (ca. 17°N) and Isthmus of Kra (ca. 10°N). Osada *et al.* (2010), analyzing 54 autosomal loci, demonstrated ancient bidirectional gene flow between Indonesian–Malaysian long-tailed and Burmese rhesus macaques. Bunlungsup and her colleagues analyzed widely- and densely-collected samples and revealed that gene introgression from rhesus to long-tailed macaques was beyond the Isthmus of Kra (Bunlungsup *et al.*, 2017a; b), which had traditionally been considered a significant biogeographical barrier. Gene introgression was found to be heterogeneous across the genome; some genes may have experienced adaptive introgression across species, while others may be responsible for reproductive isolation (Osada *et al.*, 2010; Yan *et al.*, 2011; Satkoski Trask *et al.*, 2013). Various approaches have been used to estimate the divergence time between the two species; for instance, the mtDNA molecular clock suggested a divergence time of

approximately 2 MYA (Hayasaka *et al.*, 1996; Blancher *et al.*, 2008; Liedigk *et al.*, 2015; Yao *et al.*, 2017). However, the mtDNA studies may have overestimated the divergence time because they did not consider the effect of ancestral polymorphisms. In fact, demographic analyses considering this have consistently suggested much younger divergence times of approximately 43 KYA (Bonhomme *et al.*, 2009), 0.9 MYA (Osada *et al.*, 2008), 1.3 MYA (Stevison & Kohn, 2009), and 1.5 MYA (Osada *et al.*, 2010). These demographic studies were based on an isolation with migration model that assumed constant migration after divergence (Bonhomme *et al.*, 2009; Stevison & Kohn, 2009; Osada *et al.*, 2010).

As stated above, the pattern and history of hybridization between rhesus and long-tailed macaques have been intensively studied and are relatively well understood. However, there remain critical questions and inscrutable mysteries that should be answered. Firstly, how can we interpret the difference between the geographic clines of morphological and nuclear genomic data? Although intermediate phenotypes were detected at the boundary line of distributions and a latitudinal cline of morphological characteristics were observed in the Indochinese populations, morphological characteristics appear to considerably and discontinuously change at the interspecific boundary (Fooden & Albrecht, 1999; Fooden, 2006; Hamada *et al.*, 2015). In contrast, population genetic analysis using 48 ancestry-informative single nucleotide polymorphisms (SNPs) demonstrated that the global ancestry of autosomes appeared to show a gradual shift from rhesus macaque- to long-tailed macaque-biased allele frequencies along latitude, with no clear abrupt change at the interspecific boundary (Bunlungsup *et al.*, 2017b). The mechanism and process that caused this inconsistency between the morphological characteristics and the nuclear genome remain unelucidated. Secondly, when and how did hybridization between the two species occur? Previous studies have detected evidence of hybridization and evaluated divergence time and migration rates under the assumption of an isolation with migration model; however, more complex demographic models, including the timing of migration, have not been evaluated. Such limitations appear to be partly due to the fact that genome-wide genotyping and wide regional sampling have not been simultaneously achieved.

The present study applied double-digest restriction site-associated DNA sequencing

(ddRAD-seq) (Peterson *et al.*, 2012) to the samples used in Bunlungsup *et al.*'s (2017b) study, which were widely sampled in and around the provisional area of introgression. ddRAD-seq enables low-cost discovery and genotyping of tens or hundreds of thousands of genetic markers (Peterson *et al.*, 2012; Andrews *et al.*, 2016). Using the genome-wide markers of samples that were widely collected geographically, we re-evaluated the genetic structure and phylogenetic relationship of populations in and around the provisional area of introgression. Then, we estimated when and how hybridization occurred based on demographic models that assumed migration and non-migration periods. Finally, we evaluated the geographic clines of morphological characteristics and allele frequencies across the genome.

2. Materials and Methods

2.1. Samples

The 142 blood-extracted DNA samples used in the present study were a part of those used in Bunlungsup *et al.* (2017b). Of them, 95 were obtained from wild individuals from six locations in Thailand (Table 1; Fig. 1). The samples also included 23 rhesus macaques derived from Suzhou/Kunming, China and 24 long-tailed macaques derived from around Palembang, Sumatra Island, Indonesia, all of which were maintained at USA breeding facilities (for details see the footnote in Table 1). The survey in Thailand was permitted by the National Research Council of Thailand and the Department of National Parks, Wildlife and Plant Conservation of Thailand. The experimental protocol was approved by the Institutional Animal Care and Use Committee of the Faculty of Science in accordance with the guidelines for the care and use of laboratory animals prepared by Chulalongkorn University, Thailand (protocol review no. 1423010). Further details regarding the samples can be found in Malaivijitnond *et al.* (2008), Smith *et al.* (Smith *et al.*, 2014), and Bunlungsup *et al.* (2017b).

2.2. Sequencing and SNP calling

The DNA samples were submitted to the Genomic Sequencing and Analysis Facility (GSAF) at The University of Texas at Austin, Texas, USA, where the ddRAD library was prepared and sequenced according to a protocol based on Peterson's original paper (Peterson *et al.*, 2012). Briefly, the restriction enzymes *Nla*III and *Mlu*CI were used to digest the genomic DNA, and fragments of 290–340 bp were selected using the Blue Pippin DNA Size Selection

System (Sage Science, Beverly, MA, USA). The library (pooled with other samples that were not used in this study) was sequenced on seven lanes of an Illumina HiSeq 4000 (Illumina, San Diego, USA) with 2×150 paired-end reads.

The raw reads were demultiplexed and filtered for overall sequence quality using the `process_radtags` program of the Stacks 2.2 software pipeline (Rochette & Catchen, 2017) with the following parameter settings: `-c` (clean data, remove any read with an uncalled base), `-q` (discard reads with low-quality scores), `-r` (rescue barcodes and RAD-tags), `-s 20` (discard reads if the average score within the sliding window drops below this value), `-t 140` (truncate final read length to this value). The filtered reads were mapped to the RefSeq of rhesus macaque [Mmul_10 (GCF_003339765.1)] using Bowtie2 2.3.5 (Langmead & Salzberg, 2012) with `--very-sensitive` option. The mapped reads were filtered to retain uniquely mapped reads with a minimum mapping quality of 20 using SAMtools 1.9 (Li *et al.*, 2009; Li, 2011).

SNP calling was performed using the Stacks 2.5 software pipeline. The reads uniquely mapped to the autosome, X-chromosome, and Y-chromosome in bam format were used as input, and the marukilow model (Maruki & Lynch, 2017) in the `gstacks` program was applied to search variant sites with a relatively stricter criteria than the default setting: `--var-alpha 0.01` (a significant level for calling variant sites) and `--gt-alpha 0.01` (a significant level for calling genotypes). Next, the `populations` program was used for calling SNPs with the following parameter settings: `-R 0.9` (minimum percentage of individuals across populations) and `--write-single-snp` (restrict data analysis to only the first SNP per locus). Because the Stacks software is designed to call SNPs on diploid chromosomes, homogeneous SNPs on the sex chromosome of males were transformed to be haploid using a custom script of Python programming language (Python Software Foundation, <https://www.python.org/>) wherein heterogeneous SNPs (1.5% in the Y-chromosome and 0.9 % in the X-chromosome, respectively) were removed.

For autosomal and X-chromosomes, we removed SNPs with a significant deviation from the Hardy–Weinberg equilibrium (`--hwe`) in any one of the eight populations (a *P*-value threshold was set for each population to 0.05 divided by the number of chromosomes that were surveyed within a population) and a low minor allele frequency (`--maf 0.01`). We then filtered

out individuals with >20% missing data (--mind 0.2). Finally, we removed SNPs in strong linkage disequilibrium (--indep-pairwise 10 3 0.5). This resulted in 109,068 autosomal SNPs in 138 individuals and 3,549 X-chromosome SNPs in 137 individuals. For demographic analysis, minor allele frequency filtering was skipped because it skewed the allele frequency spectrum, resulting in 234,051 autosomal SNPs. For the Y-chromosome, Hardy–Weinberg equilibrium filtering and LD pruning were skipped, resulting in 171 SNPs in 55 individuals.

2.4. Population structure and phylogenetic relationships

Population structure was estimated using a variety of approaches. First, the non-parametric approach was used to visualize the pattern of genetic similarity between populations and between individuals. The pairwise fixation index (F_{ST}) between populations was calculated from the 109,068 autosomal SNPs using the `gl.basic.stats` function of the `dartR` package in R software (R Developmental Core Team, 2019), and multidimensional scaling analysis was performed to visualize the inter-population genetic distances using the `cmdscale` function of the `stats` package in R. Principal component (PC) analysis was also performed based on the autosomal SNPs to visualize inter-individual genetic variations using the `adeigenet` package in R. The hybrid index, the proportion of individual's ancestry belonging to one of the parental populations [Sumatra long-tailed macaques (LT-Sumatra)], and interspecific heterozygosity, the proportion of loci with alleles from both parental populations [China rhesus macaques (RH-China) and LT-Sumatra], were calculated based on the 1,248 autosomal SNPs that showed an allele frequency difference (Δ) between RH-China and LT-Sumatra ≥ 0.8 using the `H1est` package (Fitzpatrick, 2013) in R.

Second, model-based approaches were used to reconstruct historical events more directly. The global ancestry for each individual was estimated using the autosomal and X-chromosome SNPs based on the maximum likelihood estimation of ADMIXTURE 1.3.0 (Alexander & Novembre, 2009), with 10-fold cross-validation for K ranges from 1 to 8. Furthermore, a haplotype-based approach was used to achieve high-resolution inference of recently shared coancestry; this was done based on the `Stacks` output (the haplotype data of autosomes) using `fineRADstructure` software (Malinsky *et al.*, 2018), wherein loci with >20 SNPs and individuals with >20% missing loci were removed. We performed 100,000 Markov chain Monte Carlo (MCMC) sampling steps with 1,000 thin intervals after a burn-in period of

100,000. Because inbreeding may skew the estimation, we repeated these analyses by excluding the samples of captive populations (RH-China and LT-Sumatra).

Finally, the phylogenetic relationship was estimated. The SNP dataset was transformed into PHYLIP format using `vcf2phylip.py` (Ortiz, 2019) with some modifications, and neighbor-joining tree was estimated with 200 bootstrap resampling based on uncorrected P-distance (wherein negative edge length was prohibited), using PAUP* 4.0a166 (<http://phylosolutions.com/paup-test/>). For autosome, phylogenetic network was also estimated based on the uncorrected P-distance matrix, using SplitsTree4 (Huson, 1998; Huson & Bryant, 2006).

2.5. Demographic modeling

The demographic history was assessed using the 234,051 autosomal SNPs (minor allele frequency filtering-skipped dataset). A folded joint site frequency spectrum (SFS) was obtained using the easySFS program (<https://github.com/isaacovercast/easySFS>). To simplify the modeling and based on the results of the population structure and phylogenetic network analysis, we combined the populations of each species into a single population for each species. Because the dataset contained missing values, the sample sizes were projected down to be 20 (rhesus macaques) and 25 (long-tailed macaques). We tested four demographic models using fastsimcoal2 software (Excoffier *et al.*, 2013). The isolation (I) model assumed no migration after the divergence between rhesus and long-tailed macaques, the isolation and migration (IM) model assumed consistent migration after the divergence, the isolation and ancient migration (IAM) model assumed that the migration stopped at some point (T_{MIG}), and the isolation and recent migration (IRM) model assumed there was no migration until some point (T_{MIG}), after which migration continued. All the models allowed asymmetric migration between the two species ($2Nm$), and constant population sizes were assumed. Because the SNP dataset used in this study lacked monomorphic sites, we fixed the effective population size of rhesus macaques at 110,000, the estimate of the effective population size for Burmese rhesus macaques by Osada *et al.* (2010). It is considered to be reasonable as Stevison & Kohn (2009) estimated the effective population size of Chinese and Indochinese rhesus macaques at similar value, 113,000. The effective population size of long-tailed macaques was N_{LT} , and that of an ancestor was N_{ANC} . The two species diverged at T_{DIV} .

For each model, 100 replicate runs were performed with the following settings: -n 100,000 (number of simulations), -m (computes the SFS for minor allele), -M (perform parameter estimation by maximum composite likelihood from the SFS), -L 30 (number of error correction model cycles to be performed when estimating parameters from SFS), -0 (does not consider monomorphic sites in observed SFS for parameter inference), -u (use multidimensional SFS), --nosingleton (ignore singletons in likelihood computation). A run with the highest likelihood was selected for each model, and Akaike's information criterion (AIC; Akaike, 1998) was used to select the best model among the four models. Non-parametric bootstraps were used to assess the credible intervals of the parameter estimate of the best models, wherein 100 pseudo-observed SFSs were generated by resampling SNPs with replacement using the easySFS program with some modifications. For each of the 100 pseudo-observed SFSs, 10 replicate runs were carried out with the same settings as the initial estimate, except for changing -L to 20 and adding the --initValues (containing initial parameter values for parameter estimation) option. The maximum likelihood parameters were used for calculating confidence intervals.

To consider the effects of sampling bias, the projection size of SFS, and the predefined population size of rhesus macaques on model selection and parameter estimation, we repeated the analysis with different settings as follows. First, we tested three types of sample subsets: (1) Wat Tham Pa Mak Ho rhesus macaques (RH-WTPMH), Ban Sang School rhesus macaques (RH-BSS), and Wat Haad Moon long-tailed macaques (LT-WHM) were removed to consider potential bias caused by the difference in the distance from the interspecific boundary; (2) LT-Sumatra population was further removed as it was disproportionately far away from the interspecific boundary; and (3) RH-China and LT-Sumatra were removed to consider the potential bias caused by inbreeding in these captive populations. Then, we tested with two different projection sizes of SFS: the sample sizes of rhesus and long-tailed macaques were projected down to be 10 and 13 (half) and 40 and 50 (two-fold). Finally, we tested with two different population sizes of rhesus macaques: 71,000 (Xue *et al.*, 2016) and 239,704 (Hernandez *et al.*, 2007). Predefined parameters other than these changes are the same as those of the main analysis.

2.6. Geographic cline analysis

Geographic clines along latitude were assessed for the hybrid index of the 1,248 autosomal and 178 X-chromosome's diagnostic SNPs ($\Delta \geq 0.8$) using the hzar package (Derryberry *et al.*, 2014) in R. The hzar package fits molecular and morphological data from hybrid zone to cline models using MCMC algorithm. We also examined the clines of relative tail length, mtDNA, and Y-chromosome ancestries using data from the previous studies (Table 1 of Hamada *et al.*, 2015; Figure 2 of Bunlungsup *et al.*, 2017a). Data from localities with precisely defined geographic positions (reported in Table 1 of Bunlungsup *et al.*, 2017a; b) as well as that from China were used. Because the geographic position of the RH-China was unknown, it was set at 25° latitude. Distances (km) from China were calculated by multiplying degrees latitude by 111 km. Because of the small sample sizes, exponential tails were not implemented at both sides. For the relative tail length, minimum and maximum values were fixed at 0 and 1, respectively. We assessed MCMC convergence by Rhat, confirming that it was <1.1 except for three autosomal SNPs, which were removed from the following analyses. The distribution of cline center and width were visualized using kernel density, and their differences between autosomes and the X-chromosome were evaluated using the kde.test function of the ks package in R. The kde.test is Kernel density-based global two-sample comparison test (Duong *et al.*, 2012). We further tested whether SNPs near genes showed less introgression or vice versa, using Chi-square test. Herein, the SNPs near genes were identified as those in which a range from 10 kb downstream to 10 kb upstream overlaps with any of genes, using GenomicRanges, ChIPpeakAnno, and TxDb.Mmulatta.UCSC.rheMac10.refGene packages in R. Introgression level was defined in two ways by the cline center: 1) south or north of the Isthmus of Kra (10°N); 2) within or without a 100 km north-south range around the interspecific boundary (midway between RH-WTPMH and LT-WHM).

3. Results

3.1. Basic statistics, population structure, and phylogenetic relationships

Our analyses based on genome-wide SNPs present complex pictures of population structure and phylogenetic relationships in Indochinese rhesus and long-tailed macaques.

The samples from RH-China and LT-Sumatra, which were derived from captive

colonies, showed some indication of inbreeding (positive inbreeding coefficient [F_{IS}]). The LT-WHM population showed negative F_{IS} , and the F_{IS} estimate of the other wild populations was relatively close to zero (Table S1). No clear difference in genetic diversity was detected between populations; π ranged from 0.025 (LT-WHM) to 0.046 (LT-Sumatra) at variant sites.

Inter-population and inter-individual variations are presented in Figure 2. Pairwise F_{ST} and its multidimensional scaling analysis showed substantial genetic differences between the two species; the LT-WHM population was considerably differentiated from the other populations of long-tailed macaques. The first two PCs accounted for 26.6% of the total variance. In PC1 (15.0 %), the score tended to be gradually larger at lower latitudes, while intra-specific variation was much larger in long-tailed than rhesus macaques. PC2 (11.6 %) represented a considerable difference in the LT-WHM population compared with the other populations. A triangle plot of the hybrid index and interspecific heterozygosity indicated that there were no young generation hybrids, except for one individual of RH-WTPMH, which appeared to be the backcross of F1 and rhesus macaques. The hybrid index, like PC1, showed a latitudinal cline.

Population structure was also assessed using model-based approaches, namely, ADMIXTURE and fineRADstructure. The cross-validation error in the ADMIXTURE analysis of autosomal SNPs was the smallest when $K = 5$ (Fig. S1a). Rhesus macaques were classified into two clusters (RH-BSS and the other two). Long-tailed macaques constituted three clusters (LT-WHM, Wat Khao Thamon [LT-WKT], and LT-Sumatra), and Suan Somdet Prasrinakharin Chumphon (LT-SSD) and Khao Noi/Khao Tang Kuan (LT-KNKTK) were likely admixed populations between LT-WKT and LT-Sumatra. There was little evidence of current or recent admixture between the two species, except for the one individual in RH-WTPMH. For the X-chromosome, the cross-validation error was the smallest at $K = 8$, representing the independence of local populations (Fig. S1b). The coancestry matrix of RAD-loci inferred by fineRADstructure supported the ADMIXTURE analysis, representing much more shared coancestry within each population than between populations (Fig. 3). The three populations of rhesus macaques shared more coancestry with each other than with the long-tailed macaques, except for the one individual of RH-WTPMH. In long-tailed macaques, LT-WHM shared little

coancestry with the other populations, while the other four shared coancestry relatively well with each other. Even excluding the samples from captive colonies (RH-China and LT-Sumatra), the analyses showed similar patterns (Figs. S2 and S3).

Phylogenetic networks demonstrated a clear division between the two species (Fig. 4). The inter-population (intra-specific) diversity was larger in long-tailed macaques than in rhesus macaques. Rhesus macaques exhibited a polytomic pattern between the three populations, while the phylogenetic relationship in the long-tailed macaques was structured, and LT-WHM was placed outside the other populations. Neighbor-joining trees show a similar pattern with the phylogenetic networks, while, in Y-chromosome tree, LT-WHM, LT-WKT, and LT-SSD are more closely related with rhesus macaques than the other populations of long-tailed macaques (Fig. S4).

3.3. Demographic modeling

The IRM model was strongly selected based on AIC (Tables 2 and S2). In the IRM model, the population size of ancestor and long-tailed macaques were estimated to be 14,850 and 122,658, respectively (Table 3). The divergence between the two species was estimated to have occurred 82,315 generations ago (Table 3). The migration start was estimated to be much younger than this, at 16,922 generations ago. The migration rate ($2Nm$) was slightly larger in the direction from rhesus to long-tailed macaques (1.8) than in the opposite direction (1.6). When excluding the populations close to the interspecific boundary (RH-BSS, RH-WTPMH, and LT-WHM) and/or LT-Sumatra, the asymmetry in the migration rate was strengthened (Table S3). As predefined population size of rhesus macaques increased, divergence time became older, and the population sizes of ancestry and long-tailed macaques became larger. Although halving projection size skewed parameter estimates, doubling it has little influences; therefore, SFS projection seems reasonable unless extremely downsizing.

3.4. Geographic clines

Geographic cline analysis showed that the relative tail length and the type of mtDNA were drastically changed at the traditionally recognized interspecific boundary, approximately 17°N (Fig. 5). The Y-chromosome boundary was located at approximately 10°N (around the Isthmus of Kra). The hybrid index of diagnostic markers showed a gradual change, wherein the

center of a cline was located approximately halfway between the mtDNA and Y-chromosome's boundaries. Although the clinal center of each locus generally tended to shift more southwards than the interspecific boundary, they were concentrated around the interspecific boundary (Figs. 5 and S5). This tendency was more remarkable in the X-chromosome markers than the autosomal markers (kde.test: $Z = 4.89, P = 5.1e - 7$). The cline widths were smaller in the X-chromosome (mean = 650.6, standard deviation [SD] = 709.9) compared with the autosomes (mean = 1037.0, SD = 616.5) (kde.test: $Z = 13.45, P = 1.50e - 41$; t -test: $t_{216.86} = 6.90, P = 5.6e - 11$). Also, with a two-dimensional kernel density, the difference between the X-chromosome and autosomes was significant (kde.test: $Z = 19.73, P = 6.2e - 87$). There was no significant difference between SNPs near genes and the other SNPs in the position of cline center (Table S4): south or north of the Isthmus of Kra ($\chi^2_1 = 0.15, P = 0.69$); within or without the interspecific boundary range ($\chi^2_1 = 3.20, P = 0.07$).

4. Discussion

4.1. Population structure

Our analysis did not reveal the early generation of hybrids between rhesus and long-tailed macaques, except for one individual of RH-WTPMH, which was likely the consequence of backcross between F1 and a rhesus macaque. This interpretation was strongly suggested because, except for that particular individual, coancestry was hardly shared between the two species, and interspecific heterozygosity was relatively low (<0.3). The results of ADMIXTURE also supported the hypothesis that current or recent interspecific admixture was rare. This finding was in accordance with a previous morphological study that demonstrated the rarity of contact-zone specimens in which the relative tail length was intermediate between that of the two species (Fooden, 1997). In contrast, the PC1 score and the hybrid index of diagnostic markers showed latitudinal cline, and the long-tailed macaque's populations north of the Isthmus of Kra, namely, LT-WHM and LT-WKT, showed intermediate scores between their putative parental populations, namely, RH-China and LT-Sumatra, which was in accordance with the results of the previous study that analyzed 48 diagnostic SNPs (Bunlungsup *et al.*, 2017b). This contrast was likely because commonly-used model-based programs, such as ADMIXTURE, are designed for detecting recent or current admixture and cannot necessarily

detect historical admixture if the genetic admixture was pervasive and homogeneous across individuals (Lawson *et al.*, 2018). These findings suggested that the hybrid zone between the two species had been formed by historical admixture and that recent or current hybridization was rare.

The pattern of population structure reveals the direction and sex bias of gene flow. Long-tailed macaques exhibited larger variations in PC1 score and in the hybrid index than rhesus macaques, which supported gene introgression from rhesus to long-tailed macaques being more pervasive than in the opposite direction (Roos & Zinner, 2015; Bunlungsup *et al.*, 2017b). The present study also confirmed that the boundary of Y-chromosome ancestry was located around the Isthmus of Kra, between LT-SSD and LT-KNKTK, supporting genetic introgression from rhesus to long-tailed macaques as being male-induced (Tosi *et al.*, 2002; Bunlungsup *et al.*, 2017a). Male rhesus macaques could be more frequently accepted by female long-tailed macaques than in the opposite situation because rhesus macaques are seasonal breeders and are larger in body size, while long-tailed macaques tend to be continuous breeders and are relatively small in body size (Herndon, 1983; Kavanagh & Laursen, 1984; Weinbauer *et al.*, 2008).

In long-tailed macaques, phylogenetic relationships between populations are structured, and the uniqueness of the most outside lineage, the LT-WHM population, was detected. On the other hand, rhesus macaques showed polytomic phylogeny and relatively homogeneous genetic variations between populations. LT-WHM was largely differentiated in genetic components from the other populations. Considering the negative F_{IS} (excess heterozygosity) in LT-WHM, LT-WHM might have been influenced by an isolate-breaking effect; i.e., gene flow from a genetically differentiated unknown population may have occurred. Although we do not have any clues regarding an unknown source population, it is noteworthy that LT-WHM appears to be located in or close to the area heterogeneous for lateral facial crest pattern (Fig. 9 in Fooden, 1995). In this area, both the transzygomatic lateral facial crest pattern typical to common long-tailed macaques (*M. fascicularis fascicularis*) and infrazygomatic pattern typical to Burmese long-tailed macaques (*M. fascicularis aurea*) (Bunlungsup *et al.*, 2016; Matsudaira *et al.*, 2018; Gumert *et al.*, 2019) have been observed (Fooden, 1995).

Unfortunately, our present study did not examine samples from Burmese long-tailed macaques. Thus, future research is expected to depict a more complex admixture history between the rhesus, common, and Burmese long-tailed macaques (or their relatives).

4.2. Demographic history of hybridization

Our demographic analysis demonstrated that the IRM model more likely better explained the observed data compared with the I, IM, and IAM models, suggesting that rhesus and long-tailed macaques contacted secondarily, resulting in gene flow after long-time isolation. Although the scenario that the hybridization between the two species was likely due to secondary contact has already been suggested in a prior study (Stevison & Kohn, 2009), the significance of the present study is that we directly tested and confirmed this hypothesis. Support for the IRM model remained regardless of differences in pre-settings of the population size of rhesus macaques, the projection size of SFS, and samples. Therefore, it is reasonable to suggest that the gene flow between the two species occurred recently (in a historical sense) after a period of complete isolation or limited gene flow.

The divergence between rhesus and long-tailed macaques was estimated at approximately 82,000 generations ago when using all the samples and assuming rhesus macaque population size at 110,000. Although the generation time of macaques is still not fully understood, population genomic studies have often assumed it as six years (Osada *et al.*, 2010) or 11 years (Xue *et al.*, 2016). The divergence time was approximately 0.49 MYA when assuming a generation time of six years and approximately 0.90 MYA when assuming a generation time of 11 years. These estimates were slightly younger than or comparable to those estimated based on IM models using DNA sequences (approximately 0.9–1.3 MYA) (Osada *et al.*, 2008; Stevison & Kohn, 2009) and were much older than those using microsatellites (approximately 43 KYA) (Bonhomme *et al.*, 2009). Considering the substantial level of homoplasy in microsatellites, the estimation based on microsatellites might be downwardly biased. The slight discrepancy between the estimates of the present study and the previous studies based on DNA sequences might be partly attributed to the differences in the model (between IM and IRM) or different sampling locations.

In contrast with the previous studies, which did not use samples close to the

interspecific boundary (although Stevison & Kohn (2009) used Indochinese long-tailed macaques), the present study included samples from northern Thailand that were close to the interspecific boundary. When including the samples from the Indochinese long-tailed macaques, the estimate of Stevison & Kohn (2009) was approximately 0.45 MYA, close to the estimate of the present study. Also, when the populations living close to the interspecific boundary were removed, the older divergence time (about 120,000 generations ago) was obtained. The drawback of the present study was that we fixed the effective population size of rhesus macaques. When these values changed, the divergence time estimates would covary. Therefore, the absolute values of the estimates of the present study should be interpreted with care because they depend on several uncertain assumptions. Future research using genome-level sequences, instead of only polymorphic SNPs, from samples of various localities, are expected to elucidate this.

Asymmetric gene flow was observed, meaning that the migration rates from rhesus to long-tailed macaques were larger than in the opposite direction. Stevison and Kohn (2009) and Bonhomme *et al.* (2009) also detected unidirectional gene flow from rhesus to long-tailed macaques, although Osada *et al.* (2010) detected symmetric gene flow. The migration rate from rhesus to long-tailed macaques detected in the present study ($2Nm \approx 2$) was smaller than those detected in Bonhomme *et al.* (2009) and Kanthaswamy *et al.* (2008) ($2Nm \approx 10$) and was slightly more significant than those ($2Nm \approx 1$) detected in Osada *et al.* (2010) and Stevison and Kohn (2009). These discrepancies might be attributed to the homoplasy in microsatellite data, which were used in Bonhomme *et al.* (2009) and Kanthaswamy *et al.* (2008), and to the differences in sampling locations. The present study included the two rhesus macaque populations close to the interspecific boundary (RH-BSS and RH-WTPMH) and therefore, likely detected weaker asymmetry than Stevison and Kohn (2009). In fact, the present study showed that when the populations close to the interspecific boundary were removed, the estimated migration rates and the degree of asymmetry increased. When RH-China and LT-Sumatra were removed from the analysis, asymmetry was inverted, probably because the remaining two populations of rhesus macaques are both close to the interspecific boundary. Together, these findings suggested that gene flow from rhesus to long-tailed macaques was more widespread than in the opposite direction.

4.3. Heterogeneity of introgression

Heterogeneity of introgression between genotype and phenotype and between genetic loci were observed. While the types of morphological characteristics (relative tail length) and mtDNA abruptly changed around the traditionally recognized interspecific boundary, the Y-chromosome boundary was far south, around the Isthmus of Kra. In contrast, genomic average ancestry, as inferred by the hybrid index and PC1 scores, gradually changed along latitude across the interspecific boundary and the Isthmus of Kra. These findings were in accordance with previous studies (Tosi *et al.*, 2002; Hamada *et al.*, 2015; Bunlungsup *et al.*, 2017b; a). The significance of the present study is that we detected the heterogeneity of introgression between genetic loci, giving hints to interpret discrepancies between the geographic variations of morphological characteristics and average genomic ancestry. While the center and width (slope) of geographic clines were considerably varied across loci, the centers for some loci were concentrated around the interspecific boundary, and many of them had a small width (steep slope). These loci were probably responsible for reproductive isolation, contributing to the persistence of the interspecific boundary at which morphological characteristics (including relative tail length and pelage color pattern) discontinuously change. Most of the other loci appeared to have experienced genetic introgression of various degrees, likely due to genetic drift, while a portion of the others showed considerable introgression exceptionally far south beyond the Isthmus of Kra and might have experienced adaptive introgression. However, there is no significant difference in the degree of introgression between SNPs near genes and the other SNPs, and thus it still remains unclear whether heterogeneity of introgression is caused by genetic drift or natural selection. Further research using a larger number of markers are expected to elucidate this issue and clarify the properties of genes experiencing reproductive isolation and adaptive introgression if any.

The difference between the introgression patterns of the autosomes and the X-chromosome is also intriguing. The present study revealed that the cline centers were more frequently concentrated around the interspecific boundary, and cline widths were smaller in the loci of X-chromosomes than in autosomes. Such a difference in the introgression pattern between the X-chromosomes and autosomes is commonly observed in the hybrids of mammals, including mice (Tucker *et al.*, 2006) and humans (Sankararaman *et al.*, 2014). This phenomenon

probably represents the larger contribution of the X-chromosomes than autosomes to reproductive isolation (the so-called large X-effect). Like many cases in mammals (including the hybridization between *Homo sapiens* and *H. neanderthalensis*), the X-chromosomes might have contributed more significantly to reproductive isolation than autosomes in the contact zone between rhesus and long-tailed macaques. Alternatively, sex bias in migration also contributes to the discrepancy between the degrees of introgression of the X-chromosomes and autosomes.

4.4. Conclusion

The present study analyzed genome-wide SNPs to elucidate the population structure, demographic history, and geographic clines of morphological characteristics and allele frequencies in the rhesus and long-tailed macaques in the Indochina Peninsula. The genetic structure of the Indochinese long-tailed macaque-populations could not be solely explained by the admixture between Chinese–Indochinese rhesus and Indonesian common long-tailed macaques and might have been influenced by an unknown third lineage. The hybridization between the two species probably occurred by secondary contact after a period of isolation. Although many genes are largely introgressed from rhesus to long-tailed macaques, some genes are likely responsible for reproductive isolation and might have contributed to the maintenance of an interspecific boundary along with species-specific morphological characteristics. This is likely the mechanism underlying the inconsistency that genetic components (on average) gradually changed along latitude while morphological characteristics discontinuously changed at the interspecific boundary. These findings are expected to help in the understanding of hybridization and its consequences as well as speciation in primates, including humans.

References

- Abbott, R.J., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., *et al.* 2013. Hybridization and speciation. *J. Evol. Biol.* **26**: 229–246.
- Abbott, R.J., Barton, N.H. & Good, J.M. 2016. Genomics of hybridization and its evolutionary consequences. *Mol. Ecol.* 2325–2332.
- Ackermann, R.R. & Bishop, J.M. 2010. Morphological and molecular evidence reveals recent hybridization between gorilla taxa. *Evolution (N. Y.)*. **64**: 271–290.
- Akaike, H. 1998. Information theory and an extension of the maximum likelihood principle. In: *Selected Papers of Hirotugu Akaike* (E. Parzen, K. Tanabe, & G. Kitagawa, eds), pp. 199–213. Springer, New York.
- Alexander, D.H. & Novembre, J. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**: 1655–1664.

- Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. & Hohenlohe, P.A. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* **17**: 81–92.
- Arnold, M.L. & Kunte, K. 2017. Adaptive genetic exchange: A tangled history of admixture and evolutionary innovation. *Trends Ecol. Evol.* **32**: 601–611.
- Arnold, M.L. & Meyer, A. 2006. Natural hybridization in primates: one evolutionary mechanism. *Zoology* **109**: 261–276.
- Barr, A., Premasuthan, A., Satkoski, J., Smith, D.G., George, D. & Kanthaswamy, S. 2011. A rapid quantitative real-time PCR-based DNA quantification assay coupled with species-assignment capabilities for two hybridizing macaca species. *Folia Primatol.* **82**: 71–80.
- Barton, N.H. 2001. The role of hybridization in evolution. *Mol. Ecol.* **10**: 551–568.
- Bell, M.A. & Travis, M.P. 2005. Hybridization, transgressive segregation, genetic covariation, and adaptive radiation. *Trends Ecol. Evol.* **20**: 358–361.
- Blancher, A., Bonhomme, M., Crouau-Roy, B., Terao, K., Kitano, T., Saitou, N., *et al.* 2008. Mitochondrial DNA sequence phylogeny of 4 populations of the widely distributed cynomolgus macaque (*Macaca fascicularis fascicularis*). *J. Hered.* **99**: 254–264.
- Bonhomme, M., Cuartero, S., Blancher, A. & Crouau-Roy, B. 2009. Assessing natural introgression in 2 biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*). *J. Hered.* **100**: 158–169.
- Bunlungsup, S., Imai, H., Hamada, Y., Gumert, M.D., San, A.M. & Malaivijitnond, S. 2016. Morphological characteristics and genetic diversity of Burmese long-tailed macaques (*Macaca fascicularis aurea*). *Am. J. Primatol.* **78**: 441–455.
- Bunlungsup, S., Imai, H., Hamada, Y., Matsudaira, K. & Malaivijitnond, S. 2017a. Mitochondrial DNA and two Y-chromosome genes of common long-tailed macaques (*Macaca fascicularis fascicularis*) throughout Thailand and vicinity. *Am. J. Primatol.* **79**: 1–13.
- Bunlungsup, S., Kanthaswamy, S., Oldt, R.F., Smith, D.G., Houghton, P., Hamada, Y., *et al.* 2017b. Genetic analysis of samples from wild populations opens new perspectives on hybridization between long-tailed (*Macaca fascicularis*) and rhesus macaques (*Macaca mulatta*). *Am. J. Primatol.* **79**: e22726.
- Cortés-Ortiz, L., Duda, T.F., Canales-Espinosa, D., García-Orduña, F., Rodríguez-Luna, E. & Bermingham, E. 2007. Hybridization in large-bodied new world primates. *Genetics* **176**: 2421–2425.
- Cortés-Ortiz, L., Roos, C. & Zinner, D. 2019. Introduction to special Issue on primate hybridization and hybrid zones. *Int. J. Primatol.* **40**: 1–8.
- Derryberry, E.P., Derryberry, G.E., Maley, J.M. & Brumfield, R.T. 2014. Hzar: Hybrid zone analysis using an R software package. *Mol. Ecol. Resour.* **14**: 652–663.
- Duong, T., Goud, B. & Schauer, K. 2012. Closed-form density-based framework for automatic detection of cellular morphology changes. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 8382–7. National Academy of Sciences.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C. & Foll, M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**: e1003905.
- Fitzpatrick, B.M. 2013. Alternative forms for genomic clines. *Ecol. Evol.* **3**: 1951–1966.
- Fooden, J. 2006. Comparative review of *fascicularis*-group species of macaques (Primates: *Macaca*). *Fieldiana Zool.* **107**: 1–44.
- Fooden, J. 1964. Rhesus and crab-eating macaques: Intergradation in Thailand. *Science* **143**: 363–364.
- Fooden, J. 1995. Systematic review of southeast Asian longtail macaques, *Macaca fascicularis* (Raffles, 1821). *Fieldiana Zool.* **81**: 1–206.
- Fooden, J. 2000. Systematic review of the rhesus macaques, *Macaca mulatta* (Zimmermann, 1780). *Fieldiana Zool.* **96**: 1–180.
- Fooden, J. 1997. Tail length variation in *Macaca fascicularis* and *M. mulatta*. *Primates* **38**: 221–231.
- Fooden, J. & Albrecht, G.H. 1999. Tail-length evolution in *fascicularis*-group macaques (Cercopithecidae: *Macaca*). *Int. J. Primatol.* **20**: 431–440.
- Genner, M.J. & Turner, G.F. 2012. Ancient hybridization and phenotypic novelty within lake Malawi's cichlid fish radiation. *Mol. Biol. Evol.* **29**: 195–206.
- Gumert, M.D., Tan, A.W.Y., Luncz, L. V., Chua, C.T., Kulik, L., Switzer, A.D., *et al.* 2019. Prevalence of tool behaviour is associated with pelage phenotype in intraspecific hybrid long-

- tailed macaques (*Macaca fascicularis aurea* × *M. f. fascicularis*). *Behaviour*, doi: 10.1163/1568539X-00003557.
- Hamada, Y., Malaivijitnond, S., Kingsada, P. & Bounnam, P. 2007. The distribution and present status of primates in the northern region of Lao PDR. *Nat. Hist. J. Chulalongkorn Univ.* **7**: 161–191.
- Hamada, Y., San, A.M. & Malaivijitnond, S. 2015. Assessment of the hybridization between rhesus (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*) based on morphological characters. *Am. J. Phys. Anthropol.* **159**: 189–198.
- Hamada, Y., Suryobroto, B., Goto, S. & Malaivijitnond, S. 2008. Morphological and body color variation in Thai *Macaca fascicularis fascicularis* north and south of the Isthmus of Kra. *Int. J. Primatol.* **29**: 1271–1294.
- Hamada, Y., Watanabe, T., Chatani, K., Hayakawa, S. & Iwamoto, M. 2005. Morphometrical comparison between Indian- and Chinese-derived rhesus macaques (*Macaca mulatta*). *Anthropol. Sci.* **113**: 183–188.
- Hayasaka, K., Fujii, K. & Horai, S. 1996. Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. *Mol. Biol. Evol.* **13**: 1044–1053.
- Hernandez, R.D., Hubisz, M.J., Wheeler, D.A., Smith, D.G., Ferguson, B., Rogers, J., *et al.* 2007. Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. *Science* **316**: 240–243.
- Herndon, J.G. 1983. Seasonal breeding in rhesus monkeys: Influence of the behavioral environment. *Am. J. Primatol.* **5**: 197–204.
- Huson, D.H. 1998. SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics* **14**: 68–73.
- Narnia.
- Huson, D.H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**: 254–267.
- Jadejaroen, J., Kawamoto, Y., Hamada, Y. & Malaivijitnond, S. 2016. An SNP marker at the STAT6 locus can identify the hybrids between rhesus (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*) in Thailand: a rapid and simple screening method and its application. *Primates* **57**: 93–102.
- Kanthaswamy, S., Satkoski, J., George, D., Kou, A., Erickson, B.J.-A.A. & Smith, D.G. 2008. Hybridization and stratification of nuclear genetic variation in *Macaca mulatta* and *M. fascicularis*. *Int. J. Primatol.* **29**: 1295–1311.
- Kanthaswamy, S., Satkoski, J., Kou, A., Malladi, V. & Glenn Smith, D. 2010. Detecting signatures of inter-regional and inter-specific hybridization among the Chinese rhesus macaque specific pathogen-free (SPF) population using single nucleotide polymorphic (SNP) markers. *J. Med. Primatol.* **39**: 252–265.
- Kavanagh, M. & Laursen, E. 1984. Breeding seasonality among long-tailed macaques, *Macaca fascicularis*, in Peninsular Malaysia. *Int. J. Primatol.* **5**: 17–29.
- Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357–359.
- Lawson, D.J., van Dorp, L. & Falush, D. 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat. Commun.* **9**: 3258.
- Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Liedigk, R., Kolleck, J., Böker, K.O., Meijaard, E., Md-Zain, B.M., Abdul-Latiff, M.A.B., *et al.* 2015. Mitogenomic phylogeny of the common long-tailed macaque (*Macaca fascicularis fascicularis*). *BMC Genomics* **16**: 222.
- Malaivijitnond, S. & Hamada, Y. 2008. Current situation and status of long-tailed macaques (*Macaca fascicularis*) in Thailand. *Nat. Hist. J. Chulalongkorn Univ.* **8**: 185–204.
- Malaivijitnond, S., Sae-Low, W. & Hamada, Y. 2008. The human-ABO blood groups of free-ranging long-tailed macaques (*Macaca fascicularis*) and parapatric rhesus macaques (*M. mulatta*) in Thailand. *J. Med. Primatol.* **37**: 31–37.

- Malinsky, M., Trucchi, E., Lawson, D.J. & Falush, D. 2018. RADpainter and fineRADstructure: Population Inference from RADseq Data. *Mol. Biol. Evol.* **35**: 1284–1290.
- Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**: 229–237.
- Malukiewicz, J., Boere, V., Fuzessy, L.F., Grativol, A.D., De Oliveira E Silva, I., Pereira, L.C.M., *et al.* 2015. Natural and anthropogenic hybridization in two species of eastern Brazilian marmosets (*Callithrix jacchus* and *C. penicillata*). *PLoS One* **10**: 1–22.
- Maruki, T. & Lynch, M. 2017. Genotype calling from population-genomic sequencing data. *G3 & Genes|Genomes|Genetics* **7**: 1393–1404.
- Matsudaira, K., Hamada, Y., Bunlungsup, S., Ishida, T., San, A.M. & Malaivijitnond, S. 2018. Whole mitochondrial genomic and Y-chromosomal phylogenies of Burmese long-tailed macaque (*Macaca fascicularis aurea*) suggest ancient hybridization between *fascicularis* and *sinica* species groups. *J. Hered.* **109**: 360–371.
- Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**: 14363.
- Melnick, D.J., Kidd, K.K., Melnick, D.J. & Kidd, K.K. 1985. Genetic and evolutionary relationships among Asian macaques. *Int. J. Primatol.* **6**: 123–160.
- Oldt, R.F., Kanthaswamy, S., Montes, M., Schumann, L., Grijalva, J., Bunlungsup, S., *et al.* 2019. Population genetics of the ABO locus within the rhesus (*Macaca mulatta*) and cynomolgus (*M. fascicularis*) macaque hybrid zone. *Int. J. Immunogenet.* **46**: 38–48.
- Ortiz, E.M. 2019. vcf2phyliip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. DOI:10.5281/zenodo.2540861.
- Osada, N., Hashimoto, K., Kameoka, Y., Hirata, M., Tanuma, R., Uno, Y., *et al.* 2008. Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M. fascicularis* and *M. mulatta*. *BMC Genomics* **9**: 90.
- Osada, N., Uno, Y., Mineta, K., Kameoka, Y., Takahashi, I. & Terao, K. 2010. Ancient genome-wide admixture extends beyond the current hybrid zone between *Macaca fascicularis* and *M. mulatta*. *Mol. Ecol.* **19**: 2884–2895.
- Osuna, C.E., Lim, S.-Y., Deleage, C., Griffin, B.D., Stein, D., Schroeder, L.T., *et al.* 2016. Zika viral dynamics and shedding in rhesus and cynomolgus macaques. *Nat. Med.* **22**: 1448–1455.
- Parnell, N.F., Hulsey, C.D. & Streelman, J.T. 2008. Hybridization produces novelty when the mapping of form to function is many to one. *BMC Evol. Biol.* **8**: 122.
- Parsons, K.J., Son, Y.H. & Albertson, R.C. 2011. Hybridization promotes evolvability in African cichlids: connections between transgressive segregation and phenotypic integration. *Evol. Biol.* **38**: 306–315.
- Pereira, R.J., Barreto, F.S. & Burton, R.S. 2014. Ecological novelty by hybridization: experimental evidence for increased thermal tolerance by transgressive segregation in *Tigriopus californicus*. *Evolution (N. Y.)* **68**: 204–215.
- Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**: e37135.
- Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., *et al.* 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**: 43–49.
- R Developmental Core Team. 2019. R: A language and environment for statistical computing. Vienna.
- Rochette, N.C. & Catchen, J.M. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* **12**: 2640–2659.
- Roos, C. & Zinner, D. 2015. Diversity and evolutionary history of macaques with special focus on *Macaca mulatta* and *Macaca fascicularis*. *Nonhum. Primate Nonclinical Drug Dev. Saf. Assess.* **3**–16.
- Roos, C., Zinner, D., Kubatko, L.S., Schwarz, C., Yang, M., Meyer, D., *et al.* 2011. Nuclear versus mitochondrial DNA: evidence for hybridization in colobine monkeys. *BMC Evol. Biol.* **11**: 77.
- Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., *et al.* 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**: 354–357.
- Satkoski Trask, J.A., Garnica, W.T., Smith, D.G., Houghton, P., Lerche, N. & Kanthaswamy, S. 2013. Single-nucleotide polymorphisms reveal patterns of allele sharing across the species boundary

- between rhesus (*Macaca mulatta*) and cynomolgus (*M. fascicularis*) macaques. *Am. J. Primatol.* **75**: 135–144.
- Seehausen, O. 2004. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**: 198–207.
- Sibal, L.R. & Samson, K.J. 2001. Nonhuman primates: a critical role in current disease research. *ILAR J.* **42**: 74–84.
- Simonti, C.N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D.S., Chisholm, R.L., *et al.* 2016. The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**: 737–741.
- Smith, D.G., Ng, J., George, D., Trask, J.S., Houghton, P., Singh, B., *et al.* 2014. A genetic comparison of two alleged subspecies of Philippine cynomolgus macaques. *Am. J. Phys. Anthropol.* **155**: 136–148.
- Soltis, P.S. 2013. Hybridization, speciation and novelty. *J. Evol. Biol.* **26**: 291–293.
- Stevison, L.S. & Kohn, M.H. 2009. Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. *Mol. Ecol.* **18**: 2457–75.
- Street, S.L., Kyes, R.C., Grant, R. & Ferguson, B. 2007. Single nucleotide polymorphisms (SNPs) are highly conserved in rhesus (*Macaca mulatta*) and cynomolgus (*Macaca fascicularis*) macaques. *BMC Genomics* **8**: 480.
- Svardal, H., Jasinska, A.J., Apetrei, C., Coppola, G., Huang, Y., Schmitt, C.A., *et al.* 2017. Ancient hybridization and strong adaptation to viruses across African vervet monkey populations. *Nat. Genet.* **49**: 1705–1713.
- Svensson, O., Smith, A., García-alonso, J. & Oosterhout, C. Van. 2016. Hybridization generates a hopeful monster : a hermaphroditic selfing cichlid. *R. Soc. Open Sci.* **3**: 150684.
- Taylor, S.A. & Larson, E.L. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* **3**: 170–177.
- Tosi, A.J., Morales, J.C. & Melnick, D.J. 2002. Y-chromosome and mitochondrial markers in *Macaca fascicularis* indicate introgression with Indochinese *M. mulatta* and a biogeographic barrier in the Isthmus of Kra. *Int. J. Primatol.* **23**: 161–178.
- Tucker, P.K., Sage, R.D., Warner, J., Wilson, A.C. & Eicher, E.M. 2006. Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution (N. Y.)*. **46**: 1146.
- Weinbauer, G.F., Niehoff, M., Niehaus, M., Srivastav, S., Fuchs, A., Van Esch, E., *et al.* 2008. Physiology and endocrinology of the ovarian cycle in macaques. *Toxicol. Pathol.* **36**: 7S-23S.
- Xue, C., Raveendran, M., Alan Harris, R., Fawcett, G.L., Liu, X., White, S., *et al.* 2016. The population genomics of rhesus macaques (*Macaca mulatta*) based on whole-genome sequences. *Genome Res.* **26**: 1–12.
- Yan, G., Zhang, G., Fang, X., Zhang, Y.Y., Li, C., Ling, F., *et al.* 2011. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nat. Biotechnol.* **29**: 1019–1023.
- Yao, L., Li, H., Martin, R.D., Moreau, C.S. & Malhi, R.S. 2017. Tracing the phylogeographic history of Southeast Asian long-tailed macaques through mitogenomes of museum specimens. *Mol. Phylogenet. Evol.* **116**: 227–238.
- Zinner, D., Arnold, M.L. & Roos, C. 2011. The strange blood: Natural hybridization in primates. *Evol. Anthropol.* **20**: 96–103.
- Zinner, D., Groeneveld, L.F., Keller, C. & Roos, C. 2009. Mitochondrial phylogeography of baboons (*Papio* spp.): indication for introgressive hybridization? *BMC Evol. Biol.* **9**: 83.

Data accessibility

The obtained sequencing reads were deposited in the NCBI Sequence Read Archive (PRJNA578019), and the data sets and code used in this study are available from the Dryad public archive (<https://doi.org/10.5061/dryad.1ns1rn8rf>).

Declaration of interests

We declare that we have no competing interests.

Tables

Table 1. Samples used in this study.

Species	Abbreviation	Location	N	Latitude	Longitude
Rhesus	RH-China	Suzhou/Kunming [†]	23	25.0	–
	RH-BSS	Ban Sang School	27	17.9	104.0
	RH-WTPMH	Wat Tham Pa Mak Ho	10	17.2	101.8
Long-tailed	LT-WHM	Wat Haad Moon	29	16.9	100.5
	LT-WKT	Wat Khao Thamon	12	13.0	100.0
	LT-SSD	Suan Somdet	7	9.9	99.0
		Prasrinakharin Chumphon			
	LT-KNKTK	Khao Noi/Khao Tang Kuan	10	7.2	100.6
	LT-Sumatra	near Palembang, Sumatra [‡]	24	-2.9	104.7

[†] The samples were obtained from the California National Primate Research Center, California, USA. These animals are descendants of those imported from Kunming and Suzhou, China.

[‡] The samples were provided by the Primate Products Inc., Immokalee, Florida, USA. These animals are those imported from near Palembang, Sumatra, Indonesia and their descendants.

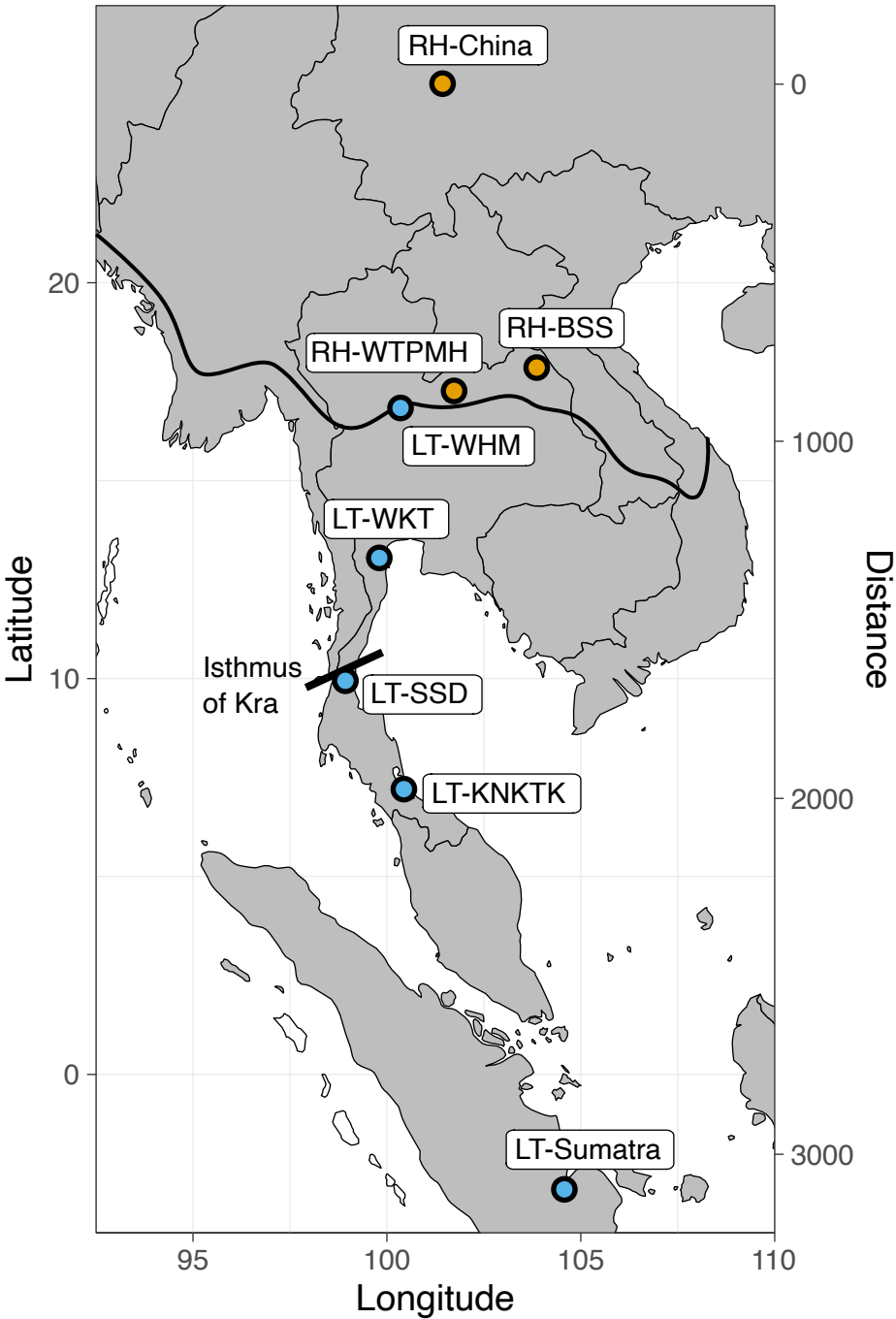
Table 2. Evaluation of demographic models.

Model	MaxEstLhood	Number of parameters (K)	AIC	Δ AIC
I	-74023	3	340895	13166
IM	-71324	5	328468	739
IAM	-71681	6	330114	2385
IRM	-71163	6	327729	0

MaxEstLhood is the maximum composite likelihood estimated according to the model parameters. MaxObsLhood (the maximum possible value for the likelihood if there was a perfect fit of the expected to the observed SFS) is -70576. Note that these values are in log10, while AIC was calculated based on normal logarithm.

Table 3. Parameter estimation for the best demographic model (IRM).

Parameters	Point estimation	95% CI	
		Lower bound	Upper bound
N_{ANC}	14,850	13,022	16,070
N_{LT}	122,658	113,968	143,203
$2Nm$ (from rhesus to long-tailed macaques)	1.8	1.6	1.9
$2Nm$ (from long-tailed to rhesus macaques)	1.6	1.4	1.7
T_{DIV}	82,315	81,452	92,452
T_{MIG}	16,922	16,648	21,425



796

797 **Figure 1.** The locations of eight populations, color-coded by rhesus (orange) and long-tailed

798 macaques (sky blue). A solid line denotes the traditionally recognized (morphology-

799 based) interspecific boundary (Fooden, 2006; adapted from Bunlungsup *et al.*, 2017b;

800 Matsudaira *et al.*, 2018).

801

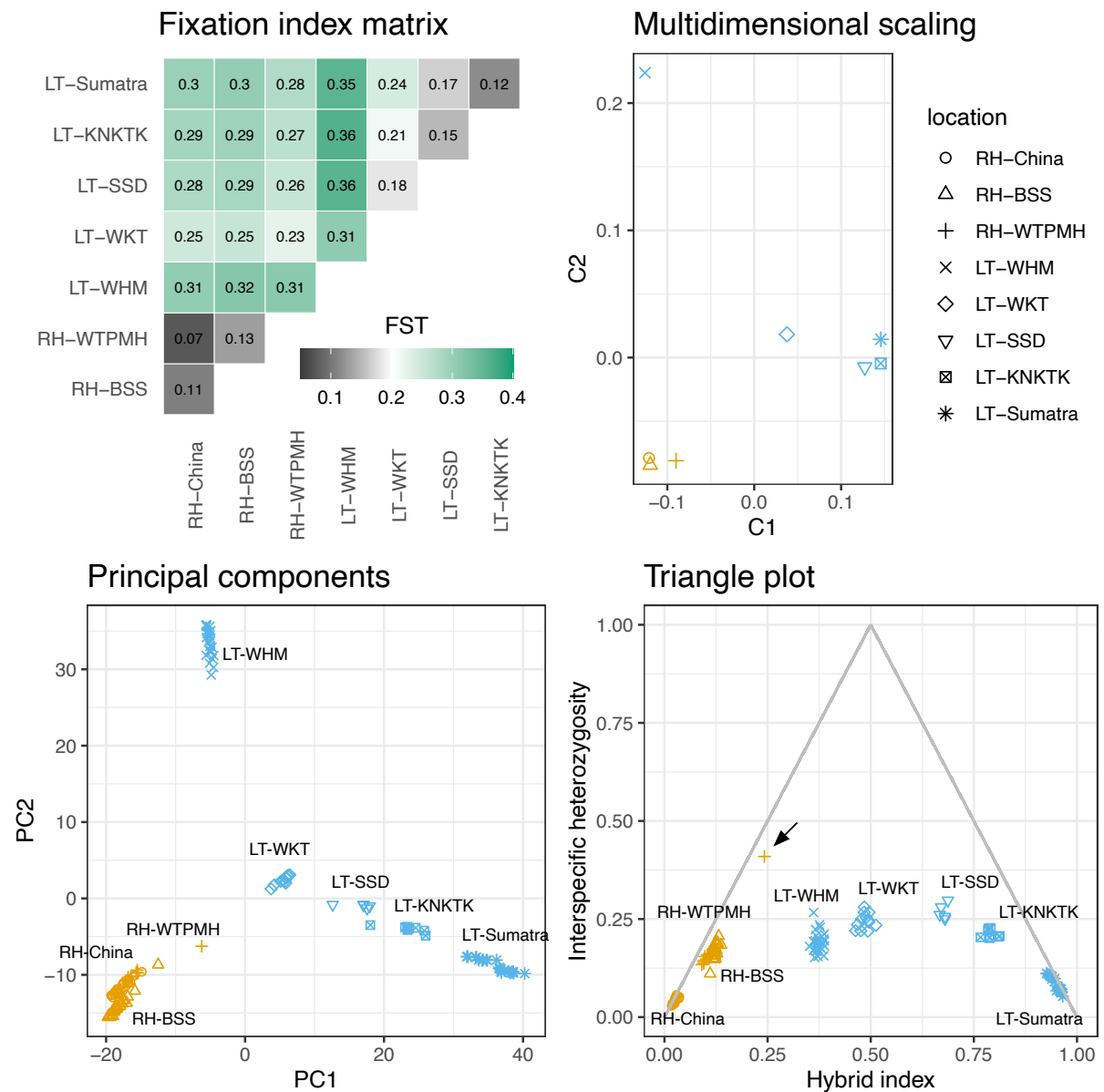


Figure 2. Pairwise F_{ST} between populations (upper left), its multidimensional scaling scores (upper right), PC scores (lower left), and the triangle plot of the hybrid index and interspecific heterozygosity (lower right). For scatter plots, species are coded by color: rhesus (orange) and long-tailed macaques (sky blue); symbols are coded by localities. An arrow in the triangle plot denotes the individual that is likely a backcross generation between F1 and rhesus macaques.

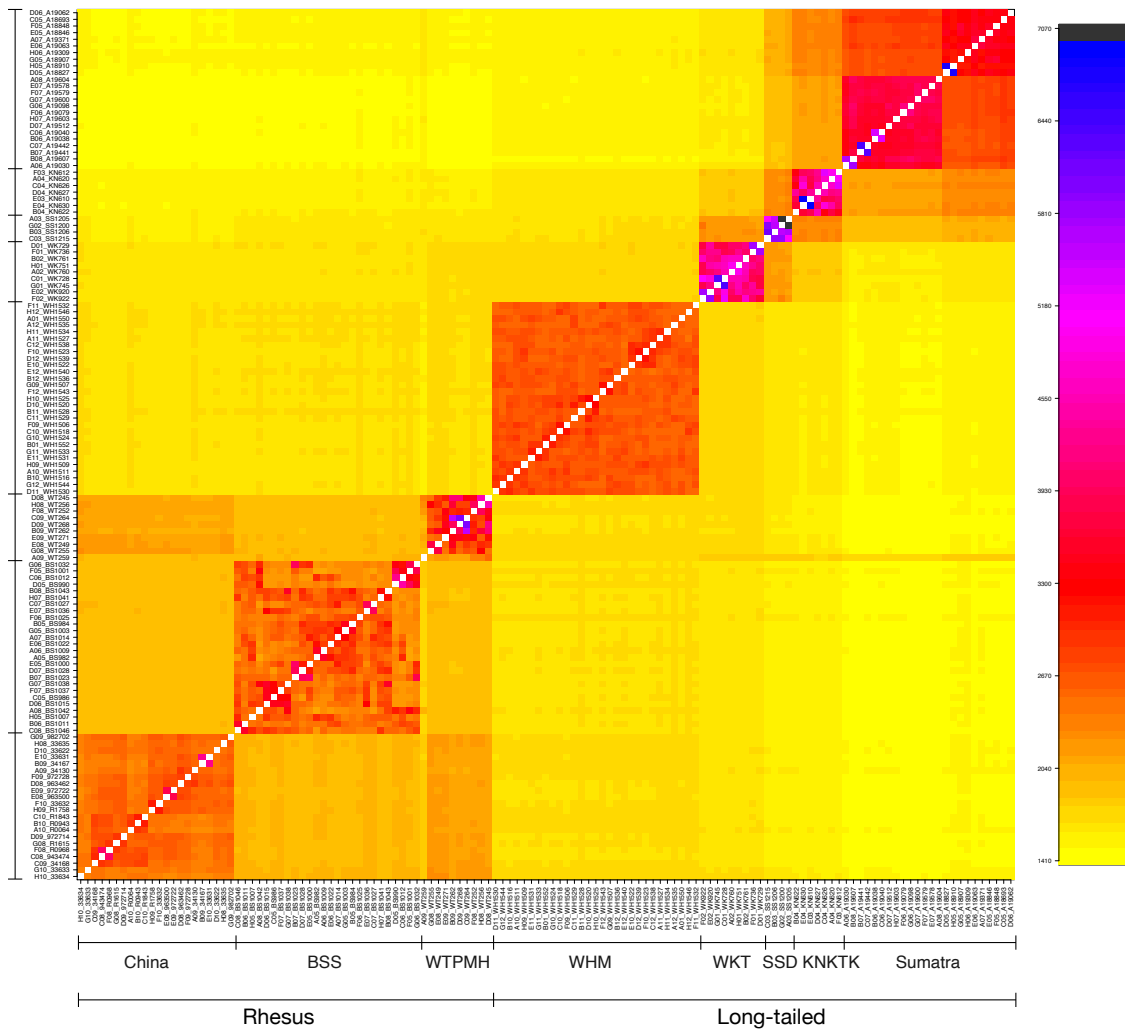


Figure 3. fineRADstructure coancestry matrix. The heatmap of fineRADstructure depicts variation in pairwise coancestry between individuals according to the scale shown on the right.

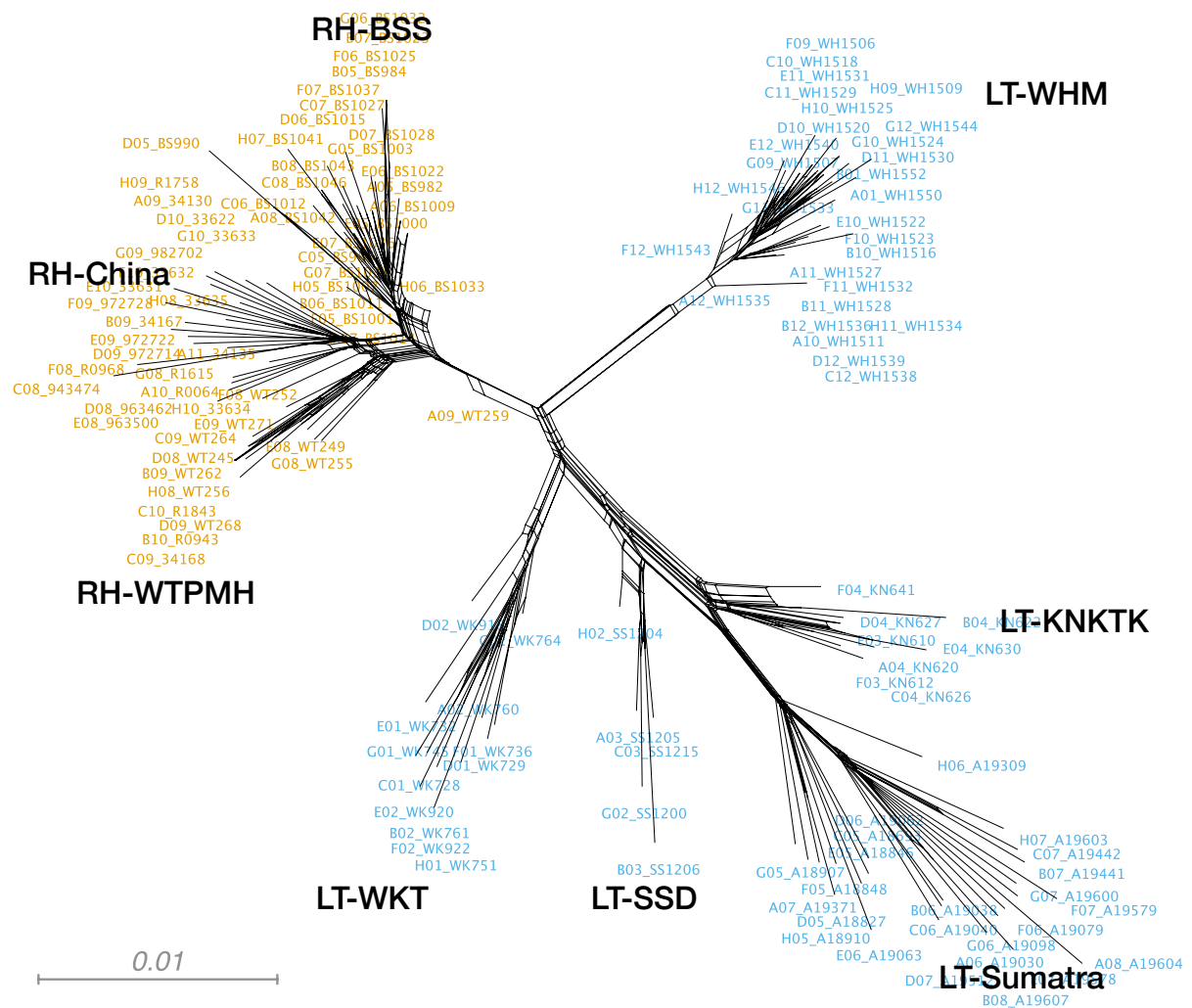


Figure 4. Phylogenetic networks based on the neighbor-net algorithm. The name of the operational taxonomic unit (sample ID) is color-coded by species: rhesus (orange) and long-tailed macaques (sky blue).

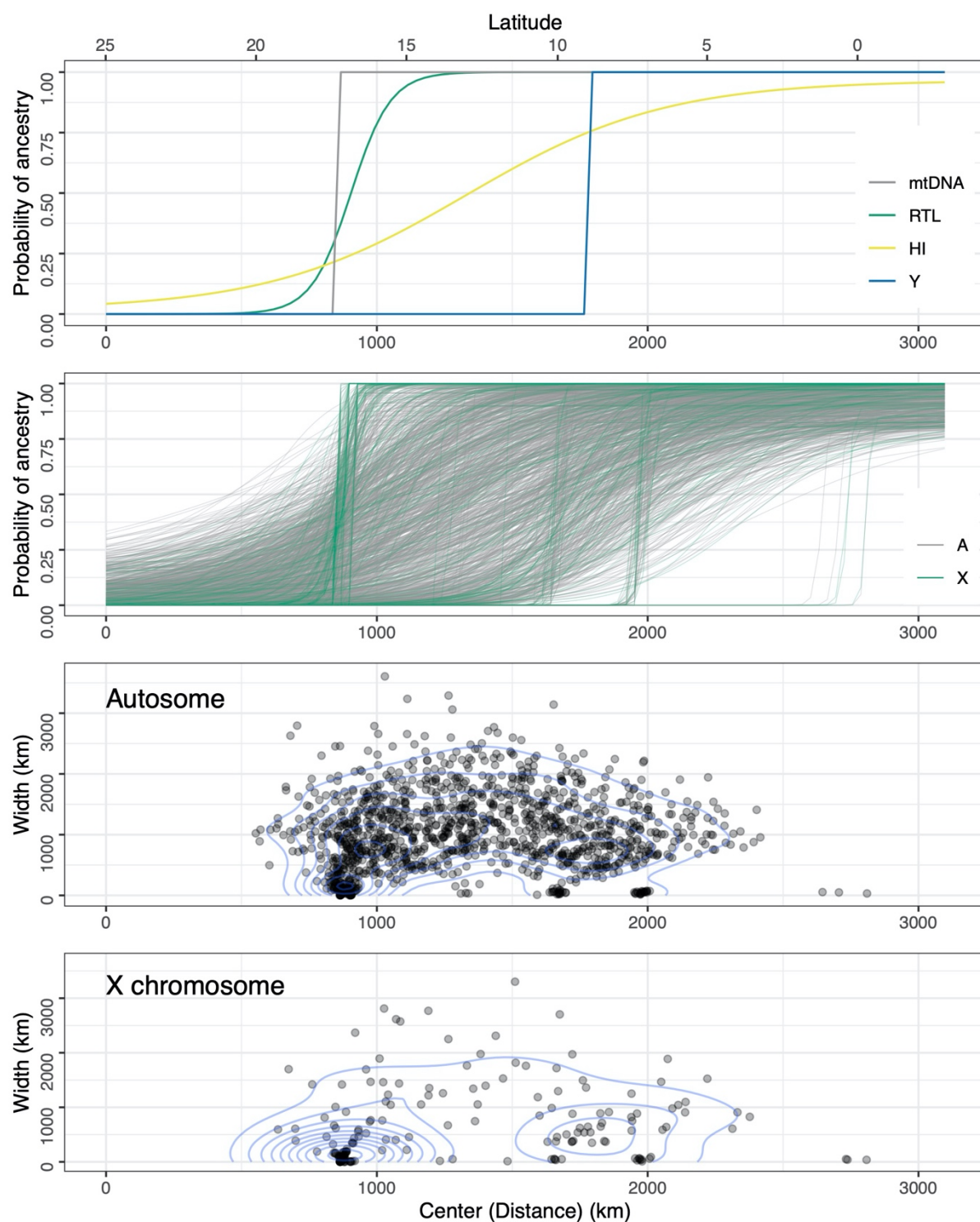


Figure 5. Geographic clines. The top panel denotes the geographic clines for the hybrid index of diagnostic markers (HI; gray), mitochondrial DNA (mtDNA; green), relative tail length (RTL; yellow), and Y-chromosome (Y; blue). The second panel denotes the geographic clines for the allele frequency of each locus in autosomes (gray) and X-chromosomes (green). The third and fourth panels indicate the scatter plots of the cline

822 centers and widths overlaid by their kernel density contours of autosomes and X-
823 chromosomes, respectively.
824

825 **Supplementary materials**

Table S1. Summary statistics of variant sites for autosome.

Population	H_O		H_S		π		F_{IS}	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
RH-China	0.0344	0.0002	0.0367	0.0002	0.0375	0.0002	0.0182	0.0026
RH-BSS	0.0330	0.0002	0.0315	0.0002	0.0321	0.0002	-0.0019	0.0022
RH- WTPMH	0.0399	0.0002	0.0374	0.0002	0.0394	0.0002	-0.0007	0.0010
LT-WHM	0.0264	0.0002	0.0245	0.0002	0.0250	0.0002	-0.0037	0.0023
LT-WKT	0.0415	0.0002	0.0409	0.0002	0.0429	0.0002	0.0042	0.0020
LT-SSD	0.0458	0.0003	0.0410	0.0002	0.0455	0.0002	-0.0004	0.0019
LT-KNKTK	0.0424	0.0002	0.0399	0.0002	0.0428	0.0002	0.0012	0.0025
LT-Sumatra	0.0419	0.0002	0.0452	0.0002	0.0462	0.0002	0.0200	0.0025

H_O , obserbed heterozygosity; H_S , expected heterozygosity; π , nucleotide diversity; F_{IS} , inbreeding coefficient

826

827

Table S2. Evaluation of demographic models in various presettings on the projection size of SFS, the population size of rhesus macaque, and samples.

Projection size of SFS	N_{RH}	Samples	MaxObsLhood	Model	MaxEstLhood	Number of parameters (K)	AIC	ΔAIC
10, 13 (half)	110,000	All	-35418	I	-36560	3	168369	4221
				IM	-35698	5	164407	259
				IAM	-35798	6	164868	720
				IRM	-35642	6	164148	0
40, 50 (2- fod)	110,000	All	-130095	I	-139152	3	640824	33847
				IM	-132246	5	609025	2047
				IAM	-133434	6	614499	7521
				IRM	-131801	6	606978	0
20, 25	71,000	All	-70576	I	-74023	3	340892	13161
				IM	-71322	5	328458	727
				IAM	-71675	6	330086	2355
				IRM	-71163	6	327731	0
20, 25	239,704	All	-70576	I	-74009	3	340832	13102
				IM	-71321	5	328458	728
				IAM	-71684	6	330130	2400
				IRM	-71163	6	327730	0
20, 25	111,000		-74062	I	-78806	3	362919	15407

		Populations close to interspecific boundary (RH-BSS, RH-WTPMH, and LT-WHM) are excluded		IM	-75751	5	348857	1345
				IAM	-76227	6	351050	3539
				IRM	-75459	6	347512	0
20, 25	111,000	Populations close to and disproportionately-far-away from interspecific boundary (RH-BSS, RH-WTPMH, LT-WHM, and LT-Sumatra) are excluded	-75039	I	-79213	3	364797	13817
				IM	-76379	5	351750	769
				IAM	-76735	6	353392	2412
				IRM	-76212	6	350980	0
20, 25	111,000	Captive populations (RH-China and LT-Sumatra) are excluded	-67833	I	-70383	3	324132	10199
				IM	-68237	5	314254	322
				IAM	-68570	6	315789	1856
				IRM	-68167	6	313933	0

MaxEstLhood is the maximum composite likelihood estimated according to the model parameters. MaxObsLhood is the maximum possible value for the likelihood if there was a perfect fit of the expected to the observed SFS. Note that these values are in log10, while AIC was calculated based on normal logarithm.

828

829

Table S3. Parameter estimation for the best demographic model (IRM) in various presettings on the projection size of SFS, the population size of rhesus macaque, and samples.

Projection size of SFS	N_{RH}	Samples	Parameters	Point estimation
10, 13 (half)	110,000	All	N_{ANC}	6,915
			N_{LT}	39,515
			$2Nm$ (from rhesus to long-tailed macaques)	1.3
			$2Nm$ (from long-tailed to rhesus macaques)	1.9
			T_{DIV}	45,179
			T_{MIG}	10,924
40, 50 (2-fod)	110,000	All	N_{ANC}	14,429
			N_{LT}	150,694
			$2Nm$ (from rhesus to long-tailed macaques)	2.1
			$2Nm$ (from long-tailed to rhesus macaques)	1.8
			T_{DIV}	77,715
			T_{MIG}	15,456
20, 25	71,000	All	N_{ANC}	9,016
			N_{LT}	78,944
			$2Nm$ (from rhesus to long-tailed macaques)	1.6
			$2Nm$ (from long-tailed to rhesus macaques)	1.4
			T_{DIV}	54,823
			T_{MIG}	12,712

20, 25	239,704	All	N_{ANC}	31,667
			N_{LT}	273,305
			$2Nm$ (from rhesus to long-tailed macaques)	1.6
			$2Nm$ (from long-tailed to rhesus macaques)	1.4
			T_{DIV}	189,147
			T_{MIG}	44,173
20, 25	111,000	Populations close to interspecific boundary (RH-BSS, RH-WTPMH, and LT-WHM) are excluded	N_{ANC}	17,205
			N_{LT}	206,371
			$2Nm$ (from rhesus to long-tailed macaques)	2.4
			$2Nm$ (from long-tailed to rhesus macaques)	0.6
			T_{DIV}	123,459
			T_{MIG}	15,217
20, 25	111,000	Populations close to and disproportionately-far-away from interspecific boundary (RH-BSS, RH-WTPMH, LT-WHM, and LT-Sumatra) are excluded	N_{ANC}	11,787
			N_{LT}	139,091
			$2Nm$ (from rhesus to long-tailed macaques)	1.5
			$2Nm$ (from long-tailed to rhesus macaques)	0.7
			T_{DIV}	97,376
			T_{MIG}	23,795
20, 25	111,000	Captive populations (RH-China and LT-Sumatra) are excluded	N_{ANC}	17,259
			N_{LT}	103,768
			$2Nm$ (from rhesus to long-tailed macaques)	1.1
			$2Nm$ (from long-tailed to rhesus macaques)	1.2

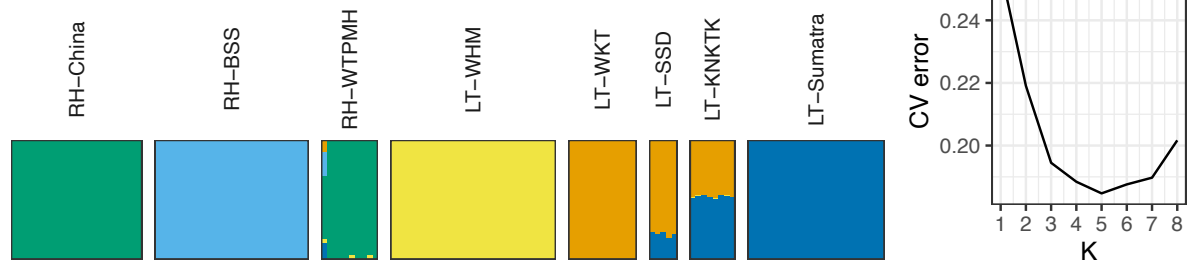
		T_{DIV}	100,270
		T_{MIG}	33,796
830			
831			

Table S4. Cross-tabulation table of SNPs.

	SNPs near genes (< 10kb)	SNPs not near genes
<i>Cline center is south of the Isthmus of Kra (10° N)</i>		
Yes	54	334
No	134	901
<i>Cline center is around the interspecific boundary (100 km north–south range)</i>		
Yes	46	230
No	142	1005

832

Autosome ($K = 5$)



X chromosome ($K = 8$)

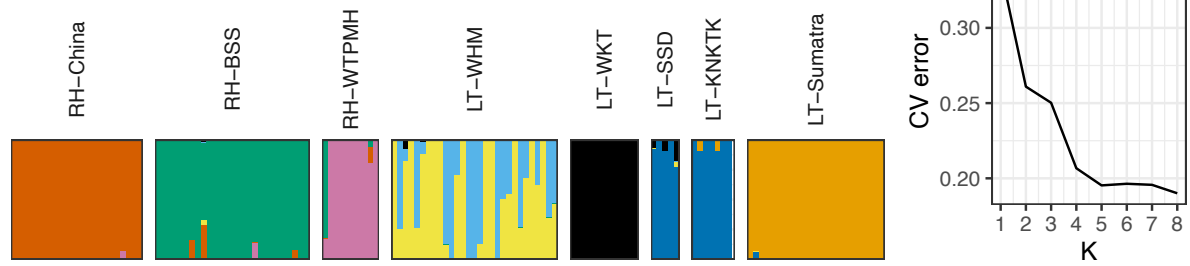


Figure S1 ADMIXTURE barplot of autosome (a) and X-chromosome (b). The cross-validation error is the smallest when $K = 5$ for autosome and $K = 8$ for X-chromosome.

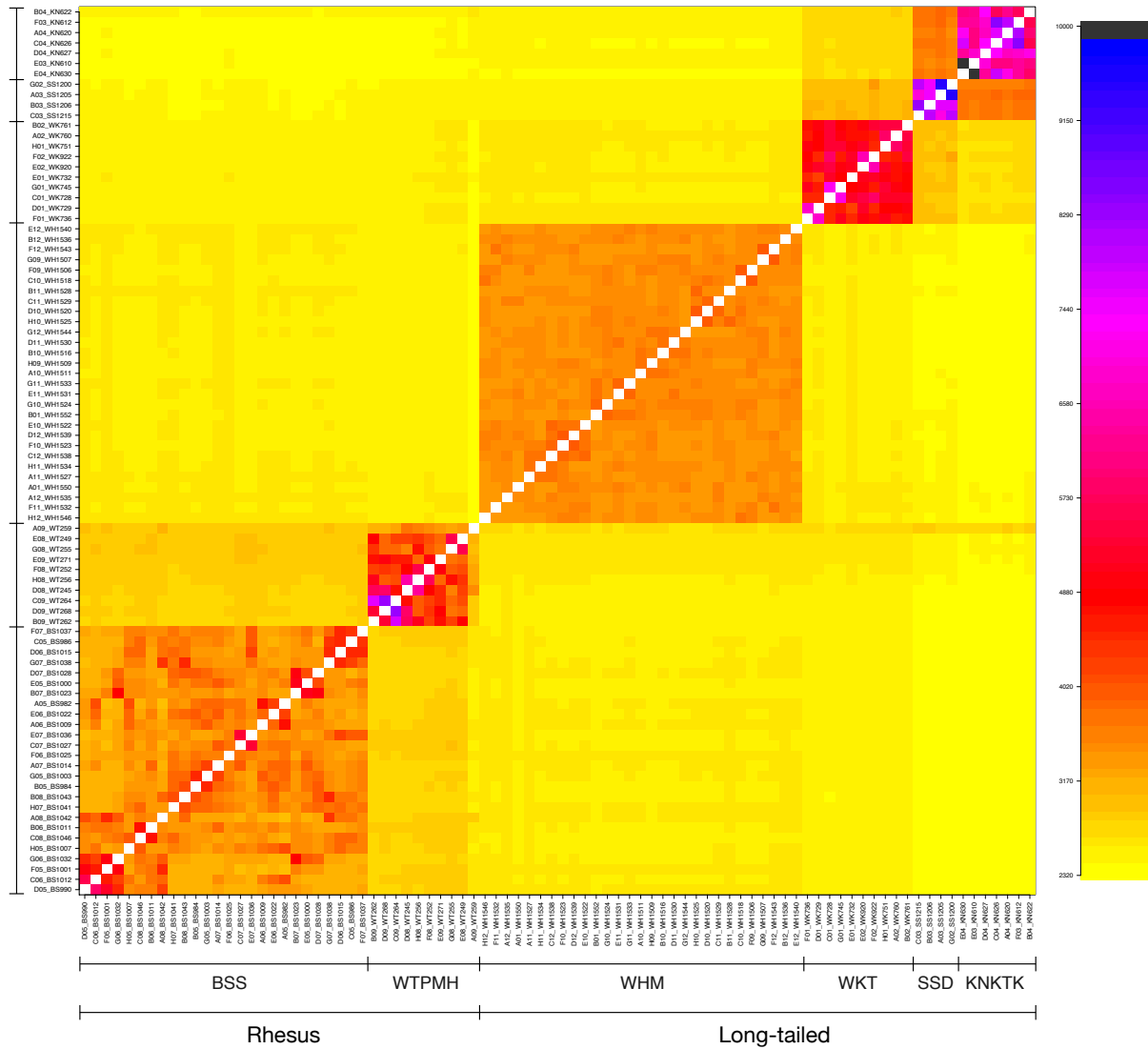


Figure S2 fineRADstructure coancestry matrix for wild-derived samples only. The heatmap of fineRADstructure depicts variation in pairwise coancestry between individuals according to the scale shown on the right.

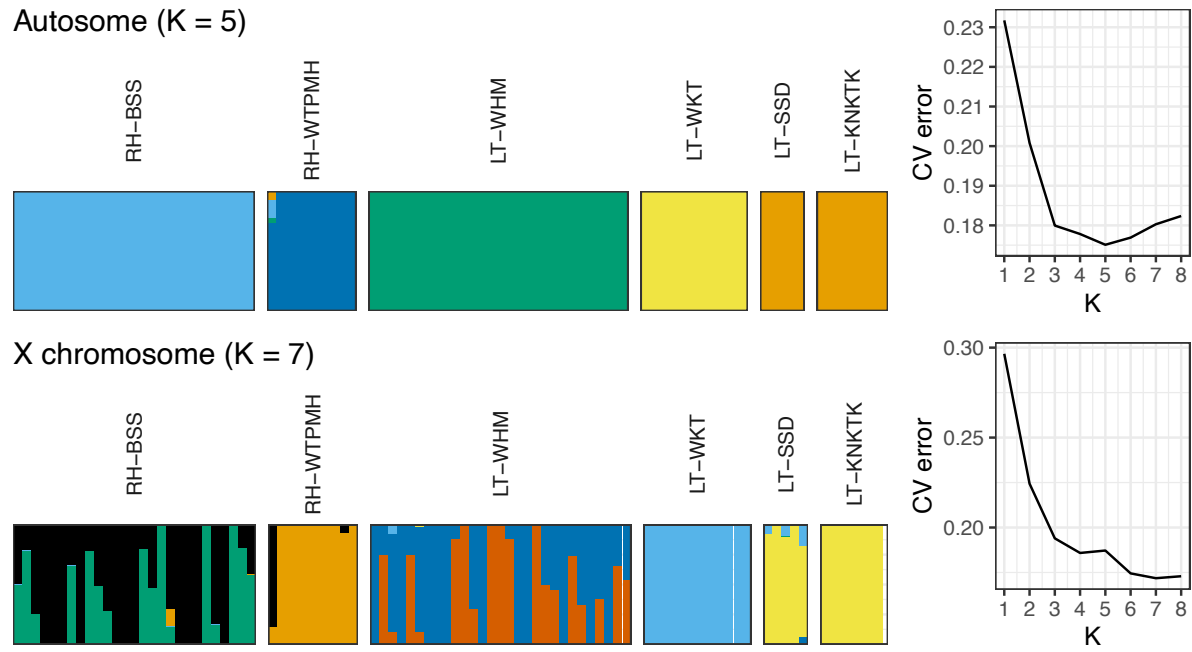


Figure S3 ADMIXTURE barplot of autosomes (a) and X-chromosome (b) for wild-derived samples only. The cross-validation error is the smallest when $K = 5$ for autosomes and $K = 7$ for X-chromosome.

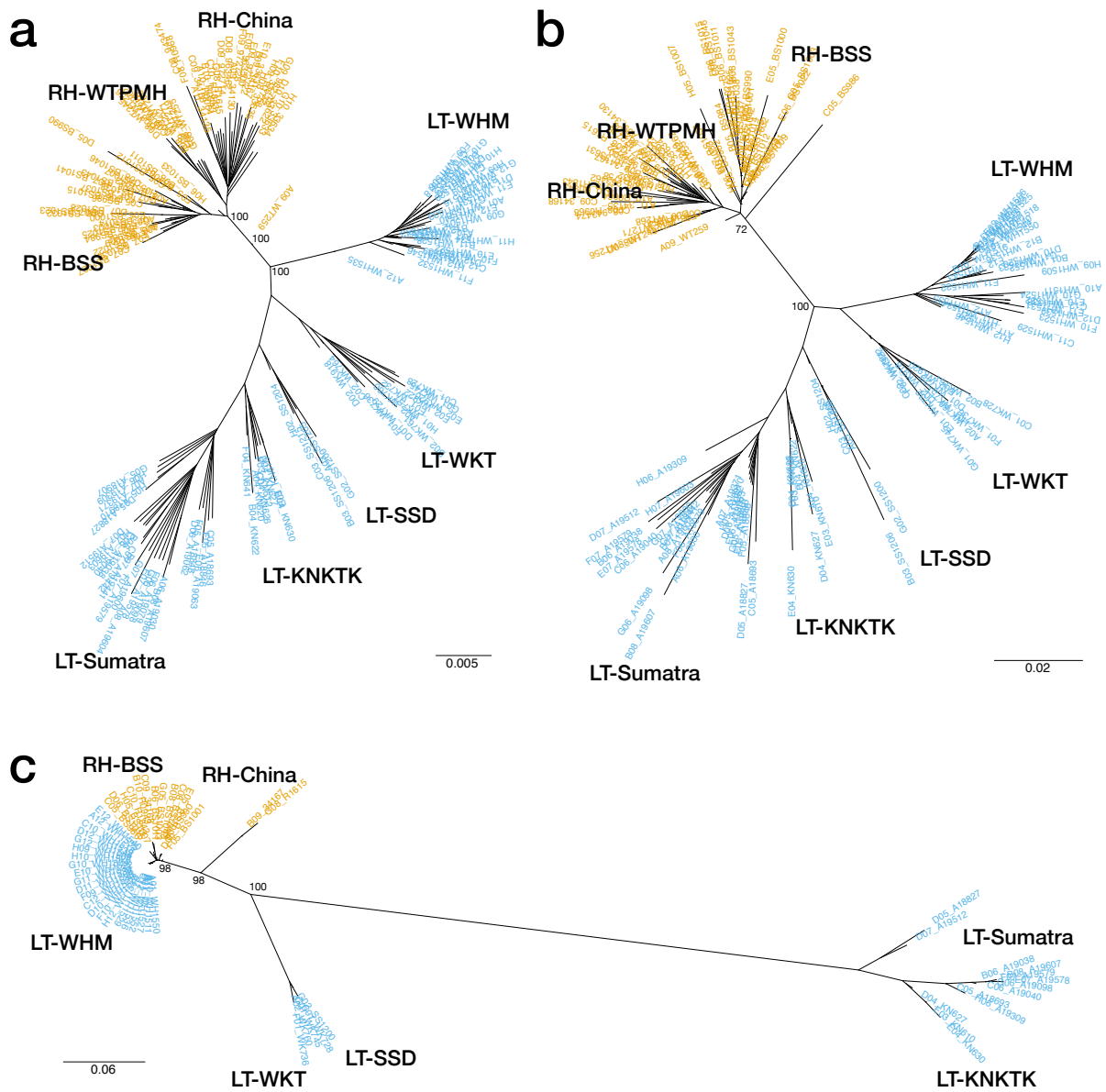


Figure S4 Neighbor-joining tree of autosome (a), X-chromosome (b), and Y-chromosome (c). Bootstrap support values are shown on the nodes of major clades.

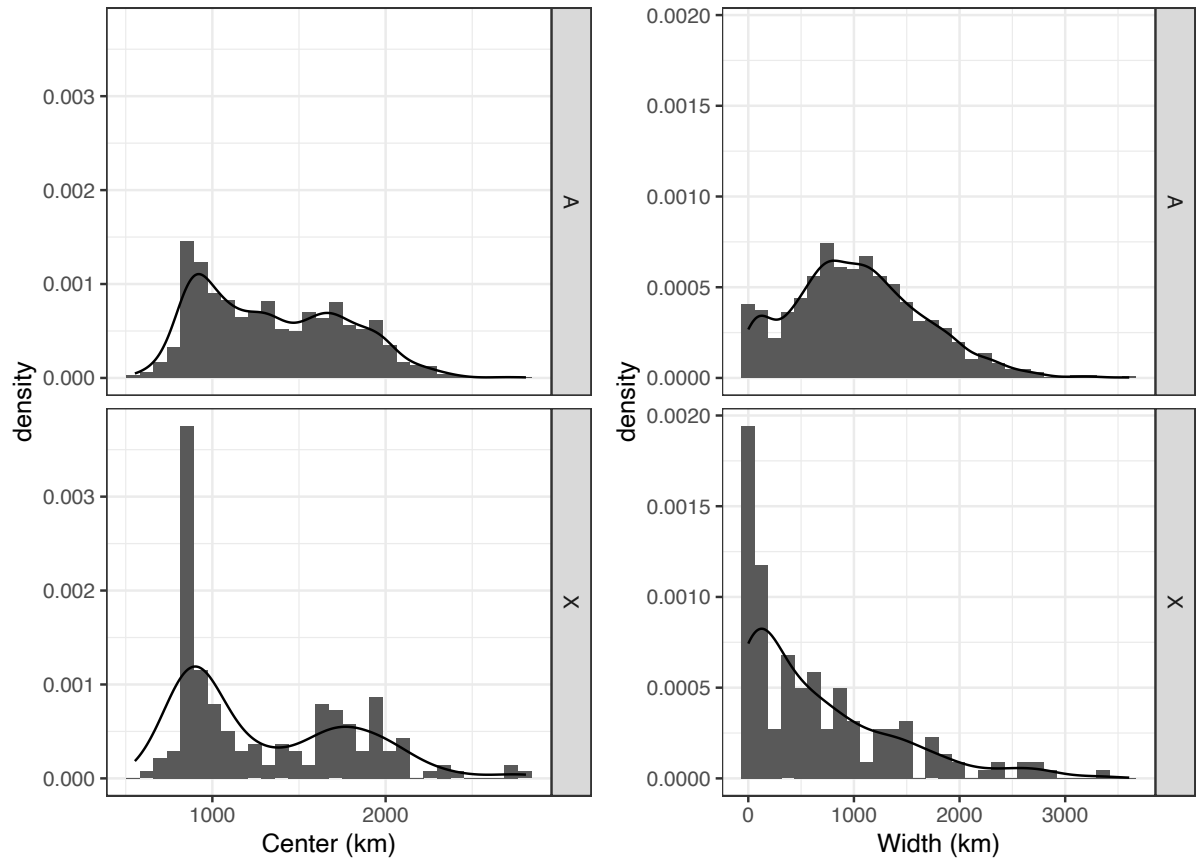


Figure S5 Density histograms of cline centers (left) and widths (right) overlaid by their kernel density profiles of autosomes (A, upper) and X-chromosomes (X, lower), respectively.