1   **Secondary contact and genomic admixture between rhesus and long-tailed macaques in**

2   **the Indochina Peninsula**

3   **Running title: Genomic admixture in macaques**

4

5   Tsuyoshi Ito[1, †], Sreetharan Kanthaswamy[2, †], Srichan Bunlungsup[3, 4], Robert Oldt[2], Paul

6   Houghton[5], Yuzuru Hamada[1], Suchinda Malaivijitnond[3, 4, †]

7

8   [1] Department of Evolution and Phylogeny, Primate Research Institute, Kyoto University,

9   Kanrin 41-2, Inuyama, Aichi 484-8506, Japan

10   [2] School of Mathematical and Natural Sciences, New College of Interdisciplinary Arts and

11   Sciences, Arizona State University West Campus, 4701 W Thunderbird Road, Glendale, AZ

12   85306-4908, USA

13   [3] Department of Biology, Faculty of Science, Chulalongkorn University, Bangkok 10330,

14   Thailand

15   [4] National Primate Research Center of Thailand-Chulalongkorn University, Saraburi 18100,

16   Thailand

17   [5] Primate Products, Inc., Immokalee, Florida, USA

18

19   [†] Correspondence authors:

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

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

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

23

24 **Author contributions**

25 TI, SK, SB, YH, and SM conceived and designed the research. SK, RO, SB, PH,

26 and SM prepared and provided the samples. TI analyzed the data and drafted the manuscript

27 with contributions by the other authors. All authors approved the final version of this

28 manuscript.

43

44   **Abstract**

45   Understanding the process and consequences of hybridization is one of the major

46   challenges in evolutionary biology. A growing body of literature has reported evidence of

47   ancient hybridization events or natural hybrid zones in primates, including humans; however,

48   we still have relatively limited knowledge about the pattern and history of admixture because

49   there have been little studies that simultaneously achieved genome-scale analysis and a

50   geographically wide sampling of wild populations. Our study applied double-digest restriction

51   site-associated DNA sequencing to samples from the six localities in and around the provisional

52   hybrid zone of rhesus and long-tailed macaques and evaluated population structure,

53   phylogenetic relationships, demographic history, and geographic clines of morphology and

54   allele frequencies. A latitudinal gradient of genetic components was observed, highlighting the

55   transition from rhesus (north) to long-tailed macaque distribution (south) as well as the presence

56   of one northern population of long-tailed macaques exhibiting unique genetic structure.

57   Interspecific gene flow was estimated to have recently occurred after an isolation period, and

58   the migration rate from rhesus to long-tailed macaques was slightly greater than in the opposite

59   direction. Although some rhesus macaque-biased alleles have widely introgressed into long-

60   tailed macaque-populations, the inflection points of allele frequencies have been observed as

61   concentrated around the traditionally recognized interspecific boundary where morphology

62   discontinuously changed; this pattern was more pronounced in the X-chromosome than in

63   autosomes. Thus, due to geographic separation before secondary contact, reproductive isolation

64   could have evolved, contributing to the maintenance of an interspecific boundary and species-

65   specific morphological characteristics.

66   **Keywords**

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

68   **1. Introduction**

69   Historically, hybridization has been considered rare in animals, but recent molecular

70   studies have revealed that hybridization is rampant in both captivity and nature (Mallet, 2005;

71   Taylor & Larson, 2019). Primates, including humans, are no exception, and a growing body of

72 literature has reported evidence of ancient hybridization events and natural hybrid zones

73 associated with various primate taxa (Arnold & Meyer, 2006; Cortés-Ortiz *et al.*, 2007, 2019;

74 Zinner *et al.*, 2009, 2011; Ackermann & Bishop, 2010; Roos *et al.*, 2011; Prüfer *et al.*, 2014;

75 Malukiewicz *et al.*, 2015; Svardal *et al.*, 2017). Hybridization causes genetic introgression,

76 which is not always maladaptive, that could be a fundamental source of evolutionary novelty

77 and phenotypic diversity (Barton, 2001; Seehausen, 2004; Bell & Travis, 2005; Parnell *et al.*,

78 2008; Parsons *et al.*, 2011; Genner & Turner, 2012; Abbott *et al.*, 2013, 2016; Soltis, 2013;

79 Pereira *et al.*, 2014; Simonti *et al.*, 2016; Svensson *et al.*, 2016; Arnold & Kunte, 2017; Meier

80 *et al.*, 2017; Taylor & Larson, 2019). Thus, hybridization has been recognized as one of the

81 most intriguing topics in evolutionary biology (Abbott *et al.*, 2016).

82 Rhesus macaques (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*), also

83 known as cynomolgus or crab-eating macaques, are closely-related species that are widely

84 distributed in Asia (Fooden, 2006). These species are also two of the most commonly-used

85 monkey models in experimental and biomedical studies (Sibal & Samson, 2001; Osuna *et al.*,

86 2016). The natural distribution of rhesus macaques ranges from Afghanistan to China and in

87 the northern part of Indochina (Fooden, 2000), while long-tailed macaques are found in

88 southern Indochina, Sumatra, Borneo, and the Philippines (Fooden, 1995; Malaivijitnond &

89 Hamada, 2008). The two species can be distinguished by their relative tail length (the ratio of

90 tail length and head-body length) and the color pattern on their backs. Rhesus macaques have

91 a relative tail length of approximately 0.4 and tend to have a bipartite back coat that is grayish-

92 brown anteriorly and tawny on the rump; long-tailed macaques have a relative tail length of

93 approximately 1.1 and a back coat that is more or less uniformly colored, ranging from pale

94 brown to dark brown (Fooden, 1964, 2006; Hamada *et al.*, 2005, 2007, 2008). The geographical

95 distributions of the two species meet at approximately 17°N (Fig. 1), where they appear to

96 produce a natural hybridization (Fooden, 2006).

97 Evidence of the hybridization of the two species has been reported in both

98 morphological and molecular studies. The first reports were morphological studies (Fooden,

99 1964), where some specimens along the boundary line of distributions showed an intermediate

100 relative tail length between the two species, suggesting that they were hybrids (Fooden, 1997).

101   Dorsal pelage color, lateral facial crest pattern, and head-body and skull lengths in Indochinese

102   long-tailed macaques are somewhat similar to rhesus macaques, supporting the existence of

103   hybridization between the two species (Fooden, 1995, 1997). Molecular studies revealed that

104   the blood protein frequency of the Indochinese long-tailed macaques was more similar to that

105   of rhesus macaques than that of non-Indochinese (Philippine) long-tailed macaques, although

106   it was initially interpreted as the consequence of symplesiomorphies (Melnick *et al.*, 1985).

107   Tosi *et al*. (2002) compared Y-chromosome data with mitochondrial DNA (mtDNA) markers

108   and found that Indochinese long-tailed macaques clustered more closely with rhesus macaques

109   than with non-Indochinese long-tailed macaques, implying that male-mediated gene flow from

110   rhesus to Indochinese long-tailed macaques had occurred.

111   Recent studies based on large numbers of genetic markers and/or geographical

112   sampling sites have revealed a more detailed picture of gene introgression between the two

113   species (Street *et al.*, 2007; Kanthaswamy *et al.*, 2008, 2010; Malaivijitnond *et al.*, 2008;

114   Bonhomme *et al.*, 2009; Stevison & Kohn, 2009; Osada *et al.*, 2010; Barr *et al.*, 2011; Yan *et

115   al.*, 2011; Satkoski Trask *et al.*, 2013; Jadejaroen *et al.*, 2016; Bunlungsup *et al.*, 2017b; a; Oldt

116   *et al.*, 2019). Restricted or whole-genome data supported Tosi *et al.*'s (2002) suggestion that

117   gene introgression was biased toward the direction from rhesus to long-tailed macaques

118   (Bonhomme *et al.*, 2009; Stevison & Kohn, 2009; Yan *et al.*, 2011). It was also suggested that

119   ancient gene introgressions occurred far beyond the traditionally recognized area of

120   introgression, i.e., a zone between the (morphology-based) interspecific boundary (ca. 17°N)

121   and Isthmus of Kra (ca. 10°N). Osada *et al*. (2010), analyzing 54 autosomal loci, demonstrated

122   ancient bidirectional gene flow between Indonesian–Malaysian long-tailed and Burmese rhesus

123   macaques. Bunlungsup and her colleagues analyzed widely- and densely-collected samples and

124   revealed that gene introgression from rhesus to long-tailed macaques was beyond the Isthmus

125   of Kra (Bunlungsup *et al.*, 2017a; b), which had traditionally been considered a significant

126   biogeographical barrier. Gene introgression was found to be heterogeneous across the genome;

127   some genes may have experienced adaptive introgression across species, while others may be

128   responsible for reproductive isolation (Osada *et al.*, 2010; Yan *et al.*, 2011; Satkoski Trask *et

129   al.*, 2013). Various approaches have been used to estimate the divergence time between the two

130   species; for instance, the mtDNA molecular clock suggested a divergence time of

5

131 approximately 2 MYA (Hayasaka *et al.*, 1996; Blancher *et al.*, 2008; Liedigk *et al.*, 2015; Yao

132 *et al.*, 2017). However, the mtDNA studies may have overestimated the divergence time

133 because they did not consider the effect of ancestral polymorphisms. In fact, demographic

134 analyses considering this have consistently suggested much younger divergence times of

135 approximately 43 KYA (Bonhomme *et al.*, 2009), 0.9 MYA (Osada *et al.*, 2008), 1.3 MYA

136 (Stevison & Kohn, 2009), and 1.5 MYA (Osada *et al.*, 2010). These demographic studies were

137 based on an isolation with migration model that assumed constant migration after divergence

138 (Bonhomme *et al.*, 2009; Stevison & Kohn, 2009; Osada *et al.*, 2010).

139 As stated above, the pattern and history of hybridization between rhesus and long-

140 tailed macaques have been intensively studied and are relatively well understood. However,

141 there remain critical questions and inscrutable mysteries that should be answered. Firstly, how

142 can we interpret the difference between the geographic clines of morphological and nuclear

143 genomic data? Although intermediate phenotypes were detected at the boundary line of

144 distributions and a latitudinal cline of morphological characteristics were observed in the

145 Indochinese populations, morphological characteristics appear to considerably and

146 discontinuously change at the interspecific boundary (Fooden & Albrecht, 1999; Fooden, 2006;

147 Hamada *et al.*, 2015). In contrast, population genetic analysis using 48 ancestry-informative

148 single nucleotide polymorphisms (SNPs) demonstrated that the global ancestry of autosomes

149 appeared to show a gradual shift from rhesus macaque- to long-tailed macaque-biased allele

150 frequencies along latitude, with no clear abrupt change at the interspecific boundary

151 (Bunlungsup *et al.*, 2017b). The mechanism and process that caused this inconsistency between

152 the morphological characteristics and the nuclear genome remain unelucidated. Secondly, when

153 and how did hybridization between the two species occur? Previous studies have detected

154 evidence of hybridization and evaluated divergence time and migration rates under the

155 assumption of an isolation with migration model; however, more complex demographic models,

156 including the timing of migration, have not been evaluated. Such limitations appear to be partly

157 due to the fact that genome-wide genotyping and wide regional sampling have not been

158 simultaneously achieved.

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

160 (ddRAD-seq) (Peterson *et al.*, 2012) to the samples used in Bunlungsup *et al.*'s (2017b) study,

161 which were widely sampled in and around the provisional area of introgression. ddRAD-seq

162 enables low-cost discovery and genotyping of tens or hundreds of thousands of genetic markers

163 (Peterson *et al.*, 2012; Andrews *et al.*, 2016). Using the genome-wide markers of samples that

164 were widely collected geographically, we re-evaluated the genetic structure and phylogenetic

165 relationship of populations in and around the provisional area of introgression. Then, we

166 estimated when and how hybridization occurred based on demographic models that assumed

167 migration and non-migration periods. Finally, we evaluated the geographic clines of

168 morphological characteristics and allele frequencies across the genome.

169 **2. Materials and Methods**

170 **2.1. Samples**

171 The 142 blood-extracted DNA samples used in the present study were a part of those

172 used in Bunlungsup *et al.* (2017b). Of them, 95 were obtained from wild individuals from six

173 locations in Thailand (Table 1; Fig. 1). The samples also included 23 rhesus macaques derived

174 from Suzhou/Kunming, China and 24 long-tailed macaques derived from around Palembang,

175 Sumatra Island, Indonesia, all of which were maintained at USA breeding facilities (for details

176 see the footnote in Table 1). The survey in Thailand was permitted by the National Research

177 Council of Thailand and the Department of National Parks, Wildlife and Plant Conservation of

178 Thailand. The experimental protocol was approved by the Institutional Animal Care and Use

179 Committee of the Faculty of Science in accordance with the guidelines for the care and use of

180 laboratory animals prepared by Chulalongkorn University, Thailand (protocol review no.

181 1423010). Further details regarding the samples can be found in Malaivijitnond et al. (2008),

182 Smith et al. (Smith *et al.*, 2014), and Bunlungsup et al. (2017b).

183 **2.2. Sequencing and SNP calling**

184 The DNA samples were submitted to the Genomic Sequencing and Analysis Facility

185 (GSAF) at The University of Texas at Austin, Texas, USA, where the ddRAD library was

186 prepared and sequenced according to a protocol based on Peterson's original paper (Peterson *et*

187 *al.*, 2012). Briefly, the restriction enzymes *Nla*III and *Mlu*CI were used to digest the genomic

188 DNA, and fragments of 290–340 bp were selected using the Blue Pippin DNA Size Selection

189   System (Sage Science, Beverly, MA, USA). The library (pooled with other samples that were

190   not used in this study) was sequenced on seven lanes of an Illumina HiSeq 4000 (Illumina, San

191   Diego, USA) with 2×150 paired-end reads.

192   The raw reads were demultiplexed and filtered for overall sequence quality using the

193   process_radtags program of the Stacks 2.2 software pipeline (Rochette & Catchen, 2017) with

194   the following parameter settings: -c (clean data, remove any read with an uncalled base), -q

195   (discard reads with low-quality scores), -r (rescue barcodes and RAD-tags), -s 20 (discard reads

196   if the average score within the sliding window drops below this value), -t 140 (truncate final

197   read length to this value). The filtered reads were mapped to the RefSeq of rhesus macaque

198   [Mmul_10 (GCF_003339765.1)] using Bowtie2 2.3.5 (Langmead & Salzberg, 2012) with --

199   very-sensitive option. The mapped reads were filtered to retain uniquely mapped reads with a

200   minimum mapping quality of 20 using SAMtools 1.9 (Li *et al.*, 2009; Li, 2011).

201   SNP calling was performed using the Stacks 2.5 software pipeline. The reads

202   uniquely mapped to the autosome, X-chromosome, and Y-chromosome in bam format were

203   used as input, and the marukilow model (Maruki & Lynch, 2017) in the gstacks program was

204   applied to search variant sites with a relatively stricter criteria than the default setting: --var-

205   alpha 0.01 (a significant level for calling variant sites) and --gt-alpha 0.01 (a significant level

206   for calling genotypes). Next, the populations program was used for calling SNPs with the

207   following parameter settings: -R 0.9 (minimum percentage of individuals across populations)

208   and --write-single-snp (restrict data analysis to only the first SNP per locus). Because the Stacks

209   software is designed to call SNPs on diploid chromosomes, homogeneous SNPs on the sex

210   chromosome of males were transformed to be haploid using a custom script of Python

211   programming language (Python Software Foundation, https://www.python.org/) wherein

212   heterogeneous SNPs (1.5% in the Y-chromosome and 0.9 % in the X-chromosome,

213   respectively) were removed.

214   For autosomal and X-chromosomes, we removed SNPs with a significant deviation

215   from the Hardy–Weinberg equilibrium (--hwe) in any one of the eight populations (a *P*-value

216   threshold was set for each population to 0.05 divided by the number of chromosomes that were

217   surveyed within a population) and a low minor allele frequency (--maf 0.01). We then filtered

8

218 out individuals with >20% missing data (--mind 0.2). Finally, we removed SNPs in strong

219 linkage disequilibrium (--indep-pairwise 10 3 0.5). This resulted in 109,068 autosomal SNPs in

220 138 individuals and 3,549 X-chromosome SNPs in 137 individuals. For demographic analysis,

221 minor allele frequency filtering was skipped because it skewed the allele frequency spectrum,

222 resulting in 234,051 autosomal SNPs. For the Y-chromosome, Hardy–Weinberg equilibrium

223 filtering and LD pruning were skipped, resulting in 171 SNPs in 55 individuals.

2.4. Population structure and phylogenetic relationships

225 Population structure was estimated using a variety of approaches. First, the non-

226 parametric approach was used to visualize the pattern of genetic similarity between populations

227 and between individuals. The pairwise fixation index (F_{ST}) between populations was calculated

228 from the 109,068 autosomal SNPs using the gl.basic.stats function of the dartR package in R

229 software (R Developmental Core Team, 2019), and multidimensional scaling analysis was

230 performed to visualize the inter-population genetic distances using the cmdscale function of the

231 stats package in R. Principal component (PC) analysis was also performed based on the

232 autosomal SNPs to visualize inter-individual genetic variations using the adegenet package in

233 R. The hybrid index, the proportion of individual's ancestry belonging to one of the parental

234 populations [Sumatra long-tailed macaques (LT-Sumatra)], and interspecific heterozygosity, the

235 proportion of loci with alleles from both parental populations [China rhesus macaques (RH-

236 China) and LT-Sumatra], were calculated based on the 1,248 autosomal SNPs that showed an

237 allele frequency difference (Δ) between RH-China and LT-Sumatra ≥ 0.8 using the HIest

238 package (Fitzpatrick, 2013) in R.

239 Second, model-based approaches were used to reconstruct historical events more

240 directly. The global ancestry for each individual was estimated using the autosomal and X-

241 chromosome SNPs based on the maximum likelihood estimation of ADMIXTURE 1.3.0

242 (Alexander & Novembre, 2009), with 10-fold cross-validation for *K* ranges from 1 to 8.

243 Furthermore, a haplotype-based approach was used to achieve high-resolution inference of

244 recently shared coancestry; this was done based on the Stacks output (the haplotype data of

245 autosomes) using fineRADstructure software (Malinsky *et al.*, 2018), wherein loci with >20

246 SNPs and individuals with >20% missing loci were removed. We performed 100,000 Markov

247 chain Monte Carlo (MCMC) sampling steps with 1,000 thin intervals after a burn-in period of

248   100,000. Because inbreeding may skew the estimation, we repeated these analyses by excluding

249   the samples of captive populations (RH-China and LT-Sumatra).

250   Finally, the phylogenetic relationship was estimated. The SNP dataset was

251   transformed into PHYLIP format using vcf2phylip.py (Ortiz, 2019) with some modifications,

252   and neighbor-joining tree was estimated with 200 bootstrap resampling based on uncorrected

253   P-distance (wherein negative edge length was prohibited), using PAUP* 4.0a166

254   (http://phylosolutions.com/paup-test/). For autosome, phylogenetic network was also estimated

255   based on the uncorrected P-distance matrix, using SplitsTree4 (Huson, 1998; Huson & Bryant,

256   2006).

2.5. Demographic modeling

258   The demographic history was assessed using the 234,051 autosomal SNPs (minor

259   allele frequency filtering-skipped dataset). A folded joint site frequency spectrum (SFS) was

260   obtained using the easySFS program (https://github.com/isaacovercast/easySFS). To simplify

261   the modeling and based on the results of the population structure and phylogenetic network

262   analysis, we combined the populations of each species into a single population for each species.

263   Because the dataset contained missing values, the sample sizes were projected down to be 20

264   (rhesus macaques) and 25 (long-tailed macaques). We tested four demographic models using

265   fastsimcoal2 software (Excoffier *et al.*, 2013). The isolation (I) model assumed no migration

266   after the divergence between rhesus and long-tailed macaques, the isolation and migration (IM)

267   model assumed consistent migration after the divergence, the isolation and ancient migration

268   (IAM) model assumed that the migration stopped at some point (T_{MIG}), and the isolation and

269   recent migration (IRM) model assumed there was no migration until some point (T_{MIG}), after

270   which migration continued. All the models allowed asymmetric migration between the two

271   species ($2Nm$), and constant population sizes were assumed. Because the SNP dataset used in

272   this study lacked monomorphic sites, we fixed the effective population size of rhesus macaques

273   at 110,000, the estimate of the effective population size for Burmese rhesus macaques by Osada

274   et al. (2010). It is considered to be reasonable as Stevison & Kohn (2009) estimated the effective

275   population size of Chinese and Indochinese rhesus macaques at similar value, 113,000. The

276   effective population size of long-tailed macaques was N_{LT}, and that of an ancestor was N_{ANC}.

277   The two species diverged at T_{DIV}.

278   For each model, 100 replicate runs were performed with the following settings: -n

279   100,000 (number of simulations), -m (computes the SFS for minor allele), -M (perform

280   parameter estimation by maximum composite likelihood from the SFS), -L 30 (number of error

281   correction model cycles to be performed when estimating parameters from SFS), -0 (does not

282   consider monomorphic sites in observed SFS for parameter inference), -u (use

283   multidimensional SFS), --nosingleton (ignore singletons in likelihood computation). A run with

284   the highest likelihood was selected for each model, and Akaike's information criterion (AIC;

285   Akaike, 1998) was used to select the best model among the four models. Non-parametric

286   bootstraps were used to assess the credible intervals of the parameter estimate of the best models,

287   wherein 100 pseudo-observed SFSs were generated by resampling SNPs with replacement

288   using the easySFS program with some modifications. For each of the 100 pseudo-observed

289   SFSs, 10 replicate runs were carried out with the same settings as the initial estimate, except

290   for changing -L to 20 and adding the --initValues (containing initial parameter values for

291   parameter estimation) option. The maximum likelihood parameters were used for calculating

292   confidence intervals.

293   To consider the effects of sampling bias, the projection size of SFS, and the pre-

294   defined population size of rhesus macaques on model selection and parameter estimation, we

295   repeated the analysis with different settings as follows. First, we tested three types of sample

296   subsets: (1) Wat Tham Pa Mak Ho rhesus macaques (RH-WTPMH), Ban Sang School rhesus

297   macaques (RH-BSS), and Wat Haad Moon long-tailed macaques (LT-WHM) were removed to

298   consider potential bias caused by the difference in the distance from the interspecific boundary;

299   (2) LT-Sumatra population was further removed as it was disproportionately far away from the

300   interspecific boundary; and (3) RH-China and LT-Sumatra were removed to consider the

301   potential bias caused by inbreeding in these captive populations. Then, we tested with two

302   different projection sizes of SFS: the sample sizes of rhesus and long-tailed macaques were

303   projected down to be 10 and 13 (half) and 40 and 50 (two-fold). Finally, we tested with two

304   different population sizes of rhesus macaques: 71,000 (Xue *et al.*, 2016) and 239,704

305   (Hernandez *et al.*, 2007). Predefined parameters other than these changes are the same as those

306   of the main analysis.

11

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

12

336   colonies, showed some indication of inbreeding (positive inbreeding coefficient [F_{IS}]). The LT-

337   WHM population showed negative F_{IS}, and the F_{IS} estimate of the other wild populations was

338   relatively close to zero (Table S1). No clear difference in genetic diversity was detected between

339   populations; π ranged from 0.025 (LT-WHM) to 0.046 (LT-Sumatra) at variant sites.

340   Inter-population and inter-individual variations are presented in Figure 2. Pairwise

341   F_{ST} and its multidimensional scaling analysis showed substantial genetic differences between

342   the two species; the LT-WHM population was considerably differentiated from the other

343   populations of long-tailed macaques. The first two PCs accounted for 26.6% of the total

344   variance. In PC1 (15.0 %), the score tended to be gradually larger at lower latitudes, while intra-

345   specific variation was much larger in long-tailed than rhesus macaques. PC2 (11.6 %)

346   represented a considerable difference in the LT-WHM population compared with the other

347   populations. A triangle plot of the hybrid index and interspecific heterozygosity indicated that

348   there were no young generation hybrids, except for one individual of RH-WTPMH, which

349   appeared to be the backcross of F1 and rhesus macaques. The hybrid index, like PC1, showed

350   a latitudinal cline.

351   Population structure was also assessed using model-based approaches, namely,

352   ADMIXTURE and fineRADstructure. The cross-validation error in the ADMIXTURE analysis

353   of autosomal SNPs was the smallest when $K = 5$ (Fig. S1a). Rhesus macaques were classified

354   into two clusters (RH-BSS and the other two). Long-tailed macaques constituted three clusters

355   (LT-WHM, Wat Khao Thamon [LT-WKT], and LT-Sumatra), and Suan Somdet Prasrinakharin

356   Chumphon (LT-SSD) and Khao Noi/Khao Tang Kuan (LT-KNKTK) were likely admixed

357   populations between LT-WKT and LT-Sumatra. There was little evidence of current or recent

358   admixture between the two species, except for the one individual in RH-WTPMH. For the X-

359   chromosome, the cross-validation error was the smallest at $K = 8$, representing the

360   independence of local populations (Fig. S1b). The coancestry matrix of RAD-loci inferred by

361   fineRADstructure supported the ADMIXTURE analysis, representing much more shared

362   coancestry within each population than between populations (Fig. 3). The three populations of

363   rhesus macaques shared more coancestry with each other than with the long-tailed macaques,

364   except for the one individual of RH-WTPMH. In long-tailed macaques, LT-WHM shared little

13

365   coancestry with the other populations, while the other four shared coancestry relatively well

366   with each other. Even excluding the samples from captive colonies (RH-China and LT-Sumatra),

367   the analyses showed similar patterns (Figs. S2 and S3).

368   Phylogenetic networks demonstrated a clear division between the two species (Fig.

369   4). The inter-population (intra-specific) diversity was larger in long-tailed macaques than in

370   rhesus macaques. Rhesus macaques exhibited a polytomic pattern between the three

371   populations, while the phylogenetic relationship in the long-tailed macaques was structured,

372   and LT-WHM was placed outside the other populations. Neighbor-joining trees show a similar

373   pattern with the phylogenetic networks, while, in Y-chromosome tree, LT-WHM, LT-WKT, and

374   LT-SSD are more closely related with rhesus macaques than the other populations of long-tailed

375   macaques (Fig. S4).

376   **3.3. Demographic modeling**

377   The IRM model was strongly selected based on AIC (Tables 2 and S2). In the IRM

378   model, the population size of ancestor and long-tailed macaques were estimated to be 14,850

379   and 122,658, respectively (Table 3). The divergence between the two species was estimated to

380   have occurred 82,315 generations ago (Table 3). The migration start was estimated to be much

381   younger than this, at 16,922 generations ago. The migration rate ($2Nm$) was slightly larger in

382   the direction from rhesus to long-tailed macaques (1.8) than in the opposite direction (1.6).

383   When excluding the populations close to the interspecific boundary (RH-BSS, RH-WTPMH,

384   and LT-WHM) and/or LT-Sumatra, the asymmetry in the migration rate was strengthened (Table

385   S3). As predefined population size of rhesus macaques increased, divergence time became older,

386   and the population sizes of ancestry and long-tailed macaques became larger. Although halving

387   projection size skewed parameter estimates, doubling it has little influences; therefore, SFS

388   projection seems reasonable unless extremely downsizing.

389   **3.4. Geographic clines**

390   Geographic cline analysis showed that the relative tail length and the type of mtDNA

391   were drastically changed at the traditionally recognized interspecific boundary, approximately

392   17°N (Fig. 5). The Y-chromosome boundary was located at approximately 10°N (around the

393   Isthmus of Kra). The hybrid index of diagnostic markers showed a gradual change, wherein the

14

394 center of a cline was located approximately halfway between the mtDNA and Y-chromosome's

395 boundaries. Although the clinal center of each locus generally tended to shift more southwards

396 than the interspecific boundary, they were concentrated around the interspecific boundary (Figs.

397 5 and S5). This tendency was more remarkable in the X-chromosome markers than the

398 autosomal markers (kde.test: $Z = 4.89, P = 5.1e - 7$). The cline widths were smaller in the

399 X-chromosome (mean = 650.6, standard deviation [SD] = 709.9) compared with the autosomes

400 (mean = 1037.0, SD = 616.5) (kde.test: $Z = 13.45, P = 1.50e - 41$; t-test: $t_{216.86} =$

401 $6.90, P = 5.6e - 11$). Also, with a two-dimensional kernel density, the difference between the

402 X-chromosome and autosomes was significant (kde.test: $Z = 19.73, P = 6.2e - 87$). There

403 was no significant difference between SNPs near genes and the other SNPs in the position of

404 cline center (Table S4): south or north of the Isthmus of Kra ($\chi_1^2 = 0.15, P = 0.69$); within or

405 without the interspecific boundary range ($\chi_1^2 = 3.20, P = 0.07$).

4. Discussion

4.1. Population structure

408 Our analysis did not reveal the early generation of hybrids between rhesus and long-

409 tailed macaques, except for one individual of RH-WTPMH, which was likely the consequence

410 of backcross between F1 and a rhesus macaque. This interpretation was strongly suggested

411 because, except for that particular individual, coancestry was hardly shared between the two

412 species, and interspecific heterozygosity was relatively low (<0.3). The results of

413 ADMIXTURE also supported the hypothesis that current or recent interspecific admixture was

414 rare. This finding was in accordance with a previous morphological study that demonstrated

415 the rarity of contact-zone specimens in which the relative tail length was intermediate between

416 that of the two species (Fooden, 1997). In contrast, the PC1 score and the hybrid index of

417 diagnostic markers showed latitudinal cline, and the long-tailed macaque's populations north

418 of the Isthmus of Kra, namely, LT-WHM and LT-WKT, showed intermediate scores between

419 their putative parental populations, namely, RH-China and LT-Sumatra, which was in

420 accordance with the results of the previous study that analyzed 48 diagnostic SNPs (Bunlungsup

421 *et al.*, 2017b). This contrast was likely because commonly-used model-based programs, such

422 as ADMIXTURE, are designed for detecting recent or current admixture and cannot necessarily

423   detect historical admixture if the genetic admixture was pervasive and homogeneous across

424   individuals (Lawson *et al.*, 2018). These findings suggested that the hybrid zone between the

425   two species had been formed by historical admixture and that recent or current hybridization

426   was rare.

427   The pattern of population structure reveals the direction and sex bias of gene flow.

428   Long-tailed macaques exhibited larger variations in PC1 score and in the hybrid index than

429   rhesus macaques, which supported gene introgression from rhesus to long-tailed macaques

430   being more pervasive than in the opposite direction (Roos & Zinner, 2015; Bunlungsup *et al.*,

431   2017b). The present study also confirmed that the boundary of Y-chromosome ancestry was

432   located around the Isthmus of Kra, between LT-SSD and LT-KNKTK, supporting genetic

433   introgression from rhesus to long-tailed macaques as being male-induced (Tosi *et al.*, 2002;

434   Bunlungsup *et al.*, 2017a). Male rhesus macaques could be more frequently accepted by female

435   long-tailed macaques than in the opposite situation because rhesus macaques are seasonal

436   breeders and are larger in body size, while long-tailed macaques tend to be continuous breeders

437   and are relatively small in body size (Herndon, 1983; Kavanagh & Laursen, 1984; Weinbauer

438   *et al.*, 2008).

439   In long-tailed macaques, phylogenetic relationships between populations are

440   structured, and the uniqueness of the most outside lineage, the LT-WHM population, was

441   detected. On the other hand, rhesus macaques showed polytomic phylogeny and relatively

442   homogeneous genetic variations between populations. LT-WHM was largely differentiated in

443   genetic components from the other populations. Considering the negative F_{IS} (excess

444   heterozygosity) in LT-WHM, LT-WHM might have been influenced by an isolate-breaking

445   effect; i.e., gene flow from a genetically differentiated unknown population may have occurred.

446   Although we do not have any clues regarding an unknown source population, it is noteworthy

447   that LT-WHM appears to be located in or close to the area heterogeneous for lateral facial crest

448   pattern (Fig. 9 in Fooden, 1995). In this area, both the transzygomatic lateral facial crest pattern

449   typical to common long-tailed macaques (*M. fascicularis fascicularis*) and infrazygomatic

450   pattern typical to Burmese long-tailed macaques (*M. fascicularis aurea*) (Bunlungsup *et al.*,

451   2016; Matsudaira *et al.*, 2018; Gumert *et al.*, 2019) have been observed (Fooden, 1995).

16

452   Unfortunately, our present study did not examine samples from Burmese long-tailed macaques.

453   Thus, future research is expected to depict a more complex admixture history between the

454   rhesus, common, and Burmese long-tailed macaques (or their relatives).

4.2. Demographic history of hybridization

456   Our demographic analysis demonstrated that the IRM model more likely better

457   explained the observed data compared with the I, IM, and IAM models, suggesting that rhesus

458   and long-tailed macaques contacted secondarily, resulting in gene flow after long-time isolation.

459   Although the scenario that the hybridization between the two species was likely due to

460   secondary contact has already been suggested in a prior study (Stevison & Kohn, 2009), the

461   significance of the present study is that we directly tested and confirmed this hypothesis.

462   Support for the IRM model remained regardless of differences in pre-settings of the population

463   size of rhesus macaques, the projection size of SFS, and samples. Therefore, it is reasonable to

464   suggest that the gene flow between the two species occurred recently (in a historical sense) after

465   a period of complete isolation or limited gene flow.

466   The divergence between rhesus and long-tailed macaques was estimated at

467   approximately 82,000 generations ago when using all the samples and assuming rhesus

468   macaque population size at 110,000. Although the generation time of macaques is still not fully

469   understood, population genomic studies have often assumed it as six years (Osada *et al.*, 2010)

470   or 11 years (Xue *et al.*, 2016). The divergence time was approximately 0.49 MYA when

471   assuming a generation time of six years and approximately 0.90 MYA when assuming a

472   generation time of 11 years. These estimates were slightly younger than or comparable to those

473   estimated based on IM models using DNA sequences (approximately 0.9–1.3 MYA) (Osada *et*

474   *al.*, 2008; Stevison & Kohn, 2009) and were much older than those using microsatellites

475   (approximately 43 KYA) (Bonhomme *et al.*, 2009). Considering the substantial level of

476   homoplasy in microsatellites, the estimation based on microsatellites might be downwardly

477   biased. The slight discrepancy between the estimates of the present study and the previous

478   studies based on DNA sequences might be partly attributed to the differences in the model

479   (between IM and IRM) or different sampling locations.

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

17

481 interspecific boundary (although Stevison & Kohn (2009) used Indochinese long-tailed

482 macaques), the present study included samples from northern Thailand that were close to the

483 interspecific boundary. When including the samples from the Indochinese long-tailed macaques,

484 the estimate of Stevison & Kohn (2009) was approximately 0.45 MYA, close to the estimate of

485 the present study. Also, when the populations living close to the interspecific boundary were

486 removed, the older divergence time (about 120,000 generations ago) was obtained. The

487 drawback of the present study was that we fixed the effective population size of rhesus

488 macaques. When these values changed, the divergence time estimates would covary. Therefore,

489 the absolute values of the estimates of the present study should be interpreted with care because

490 they depend on several uncertain assumptions. Future research using genome-level sequences,

491 instead of only polymorphic SNPs, from samples of various localities, are expected to elucidate

492 this.

493 Asymmetric gene flow was observed, meaning that the migration rates from rhesus

494 to long-tailed macaques were larger than in the opposite direction. Stevison and Kohn (2009)

495 and Bonhomme *et al*. (2009) also detected unidirectional gene flow from rhesus to long-tailed

496 macaques, although Osada *et al*. (2010) detected symmetric gene flow. The migration rate from

497 rhesus to long-tailed macaques detected in the present study ($2Nm \approx 2$) was smaller than those

498 detected in Bonhomme *et al*. (2009) and Kanthaswamy *et al*. (2008) ($2Nm \approx 10$) and was

499 slightly more significant than those ($2Nm \approx 1$) detected in Osada *et al*. (2010) and Stevison and

500 Kohn (2009). These discrepancies might be attributed to the homoplasy in microsatellite data,

501 which were used in Bonhomme *et al*. (2009) and Kanthaswamy *et al*. (2008), and to the

502 differences in sampling locations. The present study included the two rhesus macaque

503 populations close to the interspecific boundary (RH-BSS and RH-WTPMH) and therefore,

504 likely detected weaker asymmetry than Stevison and Kohn (2009). In fact, the present study

505 showed that when the populations close to the interspecific boundary were removed, the

506 estimated migration rates and the degree of asymmetry increased. When RH-China and LT-

507 Sumatra were removed from the analysis, asymmetry was inverted, probably because the

508 remaining two populations of rhesus macaques are both close to the interspecific boundary.

509 Together, these findings suggested that gene flow from rhesus to long-tailed macaques was

510 more widespread than in the opposite direction.

18

511 **4.3. Heterogeneity of introgression**

512 Heterogeneity of introgression between genotype and phenotype and between genetic

513 loci were observed. While the types of morphological characteristics (relative tail length) and

514 mtDNA abruptly changed around the traditionally recognized interspecific boundary, the Y-

515 chromosome boundary was far south, around the Isthmus of Kra. In contrast, genomic average

516 ancestry, as inferred by the hybrid index and PC1 scores, gradually changed along latitude

517 across the interspecific boundary and the Isthmus of Kra. These findings were in accordance

518 with previous studies (Tosi *et al.*, 2002; Hamada *et al.*, 2015; Bunlungsup *et al.*, 2017b; a). The

519 significance of the present study is that we detected the heterogeneity of introgression between

520 genetic loci, giving hints to interpret discrepancies between the geographic variations of

521 morphological characteristics and average genomic ancestry. While the center and width (slope)

522 of geographic clines were considerably varied across loci, the centers for some loci were

523 concentrated around the interspecific boundary, and many of them had a small width (steep

524 slope). These loci were probably responsible for reproductive isolation, contributing to the

525 persistence of the interspecific boundary at which morphological characteristics (including

526 relative tail length and pelage color pattern) discontinuously change. Most of the other loci

527 appeared to have experienced genetic introgression of various degrees, likely due to genetic

528 drift, while a portion of the others showed considerable introgression exceptionally far south

529 beyond the Isthmus of Kra and might have experienced adaptive introgression. However, there

530 is no significant difference in the degree of introgression between SNPs near genes and the

531 other SNPs, and thus it still remains unclear whether heterogeneity of introgression is caused

532 by genetic drift or natural selection. Further research using a larger number of markers are

533 expected to elucidate this issue and clarify the properties of genes experiencing reproductive

534 isolation and adaptive introgression if any.

535 The difference between the introgression patterns of the autosomes and the X-

536 chromosome is also intriguing. The present study revealed that the cline centers were more

537 frequently concentrated around the interspecific boundary, and cline widths were smaller in the

538 loci of X-chromosomes than in autosomes. Such a difference in the introgression pattern

539 between the X-chromosomes and autosomes is commonly observed in the hybrids of mammals,

540 including mice (Tucker *et al.*, 2006) and humans (Sankararaman *et al.*, 2014). This phenomenon

19

541 probably represents the larger contribution of the X-chromosomes than autosomes to

542 reproductive isolation (the so-called large X-effect). Like many cases in mammals (including

543 the hybridization between *Homo sapiens* and *H. neanderthalensis*), the X-chromosomes might

544 have contributed more significantly to reproductive isolation than autosomes in the contact zone

545 between rhesus and long-tailed macaques. Alternatively, sex bias in migration also contributes

546 to the discrepancy between the degrees of introgression of the X-chromosomes and autosomes.

547 **4.4. Conclusion**

548 　　The present study analyzed genome-wide SNPs to elucidate the population structure,

549 demographic history, and geographic clines of morphological characteristics and allele

550 frequencies in the rhesus and long-tailed macaques in the Indochina Peninsula. The genetic

551 structure of the Indochinese long-tailed macaque-populations could not be solely explained by

552 the admixture between Chinese–Indochinese rhesus and Indonesian common long-tailed

553 macaques and might have been influenced by an unknown third lineage. The hybridization

554 between the two species probably occurred by secondary contact after a period of isolation.

555 Although many genes are largely introgressed from rhesus to long-tailed macaques, some genes

556 are likely responsible for reproductive isolation and might have contributed to the maintenance

557 of an interspecific boundary along with species-specific morphological characteristics. This is

558 likely the mechanism underlying the inconsistency that genetic components (on average)

559 gradually changed along latitude while morphological characteristics discontinuously changed

560 at the interspecific boundary. These findings are expected to help in the understanding of

561 hybridization and its consequences as well as speciation in primates, including humans.

562 **References**

563 Abbott, R.J., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., *et al.* 2013. Hybridization
564 　　　and speciation. *J. Evol. Biol.* **26**: 229–246.

565 Abbott, R.J., Barton, N.H. & Good, J.M. 2016. Genomics of hybridization and its evolutionary
566 　　　consequences. *Mol. Ecol.* 2325–2332.

567 Ackermann, R.R. & Bishop, J.M. 2010. Morphological and molecular evidence reveals recent
568 　　　hybridization between gorilla taxa. *Evolution (N. Y).* **64**: 271–290.

569 Akaike, H. 1998. Information theory and an extension of the maximum likelihood principle. In:
570 　　　*Selected Papers of Hirotugu Akaike* (E. Parzen, K. Tanabe, & G. Kitagawa, eds), pp. 199–213.
571 　　　Springer, New York.

572 Alexander, D.H. & Novembre, J. 2009. Fast model-based estimation of ancestry in unrelated
573 　　　individuals. *Genome Res.* **19**: 1655–1664.

574 Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. & Hohenlohe, P.A. 2016. Harnessing the power
575   of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* **17**: 81–92.
576 Arnold, M.L. & Kunte, K. 2017. Adaptive genetic exchange: A tangled history of admixture and
577   evolutionary innovation. *Trends Ecol. Evol.* **32**: 601–611.
578 Arnold, M.L. & Meyer, A. 2006. Natural hybridization in primates: one evolutionary mechanism.
579   *Zoology* **109**: 261–276.
580 Barr, A., Premasuthan, A., Satkoski, J., Smith, D.G., George, D. & Kanthaswamy, S. 2011. A rapid
581   quantitative real-time PCR-based DNA quantification assay coupled with species-assignment
582   capabilities for two hybridizing macaca species. *Folia Primatol.* **82**: 71–80.
583 Barton, N.H. 2001. The role of hybridization in evolution. *Mol. Ecol.* **10**: 551–568.
584 Bell, M.A. & Travis, M.P. 2005. Hybridization, transgressive segregation, genetic covariation, and
585   adaptive radiation. *Trends Ecol. Evol.* **20**: 358–361.
586 Blancher, A., Bonhomme, M., Crouau-Roy, B., Terao, K., Kitano, T., Saitou, N., *et al.* 2008.
587   Mitochondrial DNA sequence phylogeny of 4 populations of the widely distributed cynomolgus
588   macaque (*Macaca fascicularis fascicularis*). *J. Hered.* **99**: 254–264.
589 Bonhomme, M., Cuartero, S., Blancher, A. & Crouau-Roy, B. 2009. Assessing natural introgression in
590   2 biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque
591   (*Macaca fascicularis*). *J. Hered.* **100**: 158–169.
592 Bunlungsup, S., Imai, H., Hamada, Y., Gumert, M.D., San, A.M. & Malaivijitnond, S. 2016.
593   Morphological characteristics and genetic diversity of Burmese long-tailed macaques (*Macaca*
594   *fascicularis aurea*). *Am. J. Primatol.* **78**: 441–455.
595 Bunlungsup, S., Imai, H., Hamada, Y., Matsudaira, K. & Malaivijitnond, S. 2017a. Mitochondrial
596   DNA and two Y-chromosome genes of common long-tailed macaques (*Macaca fascicularis*
597   *fascicularis*) throughout Thailand and vicinity. *Am. J. Primatol.* **79**: 1–13.
598 Bunlungsup, S., Kanthaswamy, S., Oldt, R.F., Smith, D.G., Houghton, P., Hamada, Y., *et al.* 2017b.
599   Genetic analysis of samples from wild populations opens new perspectives on hybridization
600   between long-tailed (*Macaca fascicularis*) and rhesus macaques (*Macaca mulatta*). *Am. J.*
601   *Primatol.* **79**: e22726.
602 Cortés-Ortiz, L., Duda, T.F., Canales-Espinosa, D., García-Orduña, F., Rodríguez-Luna, E. &
603   Bermingham, E. 2007. Hybridization in large-bodied new world primates. *Genetics* **176**: 2421–
604   2425.
605 Cortés-Ortiz, L., Roos, C. & Zinner, D. 2019. Introduction to special Iisue on primate hybridization
606   and hybrid zones. *Int. J. Primatol.* **40**: 1–8.
607 Derryberry, E.P., Derryberry, G.E., Maley, J.M. & Brumfield, R.T. 2014. Hzar: Hybrid zone analysis
608   using an R software package. *Mol. Ecol. Resour.* **14**: 652–663.
609 Duong, T., Goud, B. & Schauer, K. 2012. Closed-form density-based framework for automatic
610   detection of cellular morphology changes. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 8382–7. National
611   Academy of Sciences.
612 Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C. & Foll, M. 2013. Robust demographic
613   inference from genomic and SNP data. *PLoS Genet.* **9**: e1003905.
614 Fitzpatrick, B.M. 2013. Alternative forms for genomic clines. *Ecol. Evol.* **3**: 1951–1966.
615 Fooden, J. 2006. Comparative review of *fascicularis*-group species of macaques (Primates: *Macaca*).
616   *Fieldiana Zool.* **107**: 1–44.
617 Fooden, J. 1964. Rhesus and crab-eating macaques: Intergradation in Thailand. *Science* **143**: 363–364.
618 Fooden, J. 1995. Systematic review of southeast Asian longtail macaques, *Macaca fascicularis*
619   (Raffles, 1821). *Fieldiana Zool.* **81**: 1–206.
620 Fooden, J. 2000. Systematic review of the rhesus macaques, *Macaca mulatta* (Zimmermann, 1780).
621   *Fieldiana Zool.* **96**: 1–180.
622 Fooden, J. 1997. Tail length variation in *Macaca fascicularis* and *M. mulatta*. *Primates* **38**: 221–231.
623 Fooden, J. & Albrecht, G.H. 1999. Tail-length evolution in *fascicularis*-group macaques
624   (Cercopithecidae: *Macaca*). *Int. J. Primatol.* **20**: 431–440.
625 Genner, M.J. & Turner, G.F. 2012. Ancient hybridization and phenotypic novelty within lake
626   Malawi's cichlid fish radiation. *Mol. Biol. Evol.* **29**: 195–206.
627 Gumert, M.D., Tan, A.W.Y., Luncz, L. V., Chua, C.T., Kulik, L., Switzer, A.D., *et al.* 2019.
628   Prevalence of tool behaviour is associated with pelage phenotype in intraspecific hybrid long-

629   tailed macaques (*Macaca fascicularis aurea* × *M. f. fascicularis*). *Behaviour*, doi:
630   10.1163/1568539X-00003557.
631   Hamada, Y., Malaivijitnond, S., Kingsada, P. & Bounnam, P. 2007. The distribution and present status
632   of primates in the northern region of Lao PDR. *Nat. Hist. J. Chulalongkorn Univ.* **7**: 161–191.
633   Hamada, Y., San, A.M. & Malaivijitnond, S. 2015. Assessment of the hybridization between rhesus
634   (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*) based on morphological characters.
635   *Am. J. Phys. Anthropol.* **159**: 189–198.
636   Hamada, Y., Suryobroto, B., Goto, S. & Malaivijitnond, S. 2008. Morphological and body color
637   variation in Thai *Macaca fascicularis fascicularis* north and south of the Isthmus of Kra. *Int. J.*
638   *Primatol.* **29**: 1271–1294.
639   Hamada, Y., Watanabe, T., Chatani, K., Hayakawa, S. & Iwamoto, M. 2005. Morphometrical
640   comparison between Indian- and Chinese-derived rhesus macaques (*Macaca mulatta*).
641   *Anthropol. Sci.* **113**: 183–188.
642   Hayasaka, K., Fujii, K. & Horai, S. 1996. Molecular phylogeny of macaques: implications of
643   nucleotide sequences from an 896-base pair region of mitochondrial DNA. *Mol. Biol. Evol.* **13**:
644   1044–1053.
645   Hernandez, R.D., Hubisz, M.J., Wheeler, D.A., Smith, D.G., Ferguson, B., Rogers, J., *et al.* 2007.
646   Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus
647   macaques. *Science* **316**: 240–243.
648   Herndon, J.G. 1983. Seasonal breeding in rhesus monkeys: Influence of the behavioral environment.
649   *Am. J. Primatol.* **5**: 197–204.
650   Huson, D.H. 1998. SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics* **14**: 68–73.
651   Narnia.
652   Huson, D.H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol.*
653   *Biol. Evol.* **23**: 254–267.
654   Jadejaroen, J., Kawamoto, Y., Hamada, Y. & Malaivijitnond, S. 2016. An SNP marker at the STAT6
655   locus can identify the hybrids between rhesus (*Macaca mulatta*) and long-tailed macaques (*M.*
656   *fascicularis*) in Thailand: a rapid and simple screening method and its application. *Primates* **57**:
657   93–102.
658   Kanthaswamy, S., Satkoski, J., George, D., Kou, A., Erickson, B.J.-A.A. & Smith, D.G. 2008.
659   Hybridization and stratification of nuclear genetic variation in *Macaca mulatta* and *M.*
660   *fascicularis*. *Int. J. Primatol.* **29**: 1295–1311.
661   Kanthaswamy, S., Satkoski, J., Kou, A., Malladi, V. & Glenn Smith, D. 2010. Detecting signatures of
662   inter-regional and inter-specific hybridization among the Chinese rhesus macaque specific
663   pathogen-free (SPF) population using single nucleotide polymorphic (SNP) markers. *J. Med.*
664   *Primatol.* **39**: 252–265.
665   Kavanagh, M. & Laursen, E. 1984. Breeding seasonality among long-tailed macaques, *Macaca*
666   *fascicularis*, in Peninsular Malaysia. *Int. J. Primatol.* **5**: 17–29.
667   Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**:
668   357–359.
669   Lawson, D.J., van Dorp, L. & Falush, D. 2018. A tutorial on how not to over-interpret STRUCTURE
670   and ADMIXTURE bar plots. *Nat. Commun.* **9**: 3258.
671   Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and
672   population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.
673   Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* 2009. The sequence
674   alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
675   Liedigk, R., Kolleck, J., Böker, K.O., Meijaard, E., Md-Zain, B.M., Abdul-Latiff, M.A.B., *et al.* 2015.
676   Mitogenomic phylogeny of the common long-tailed macaque (*Macaca fascicularis fascicularis*).
677   *BMC Genomics* **16**: 222.
678   Malaivijitnond, S. & Hamada, Y. 2008. Current situation and status of long-tailed macaques (*Macaca*
679   *fascicularis*) in Thailand. *Nat. Hist. J. Chulalongkorn Univ.* **8**: 185–204.
680   Malaivijitnond, S., Sae-Low, W. & Hamada, Y. 2008. The human-ABO blood groups of free-ranging
681   long-tailed macaques (*Macaca fascicularis*) and parapatric rhesus macaques (*M. mulatta*) in
682   Thailand. *J. Med. Primatol.* **37**: 31–37.

683   Malinsky, M., Trucchi, E., Lawson, D.J. & Falush, D. 2018. RADpainter and fineRADstructure:
684   Population Inference from RADseq Data. *Mol. Biol. Evol.* **35**: 1284–1290.

685   Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**: 229–237.

686   Malukiewicz, J., Boere, V., Fuzessy, L.F., Grativol, A.D., De Oliveira E Silva, I., Pereira, L.C.M., *et*
687   *al.* 2015. Natural and anthropogenic hybridization in two species of eastern Brazilian marmosets
688   (*Callithrix jacchus* and *C. penicillata*). *PLoS One* **10**: 1–22.

689   Maruki, T. & Lynch, M. 2017. Genotype calling from population-genomic sequencing data.
690   *G3&#58; Genes|Genomes|Genetics* **7**: 1393–1404.

691   Matsudaira, K., Hamada, Y., Bunlungsup, S., Ishida, T., San, A.M. & Malaivijitnond, S. 2018. Whole
692   mitochondrial genomic and Y-chromosomal phylogenies of Burmese long-tailed macaque
693   (*Macaca fascicularis aurea*) suggest ancient hybridization between *fascicularis* and *sinica*
694   species groups. *J. Hered.* **109**: 360–371.

695   Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. 2017. Ancient
696   hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**: 14363.

697   Melnick, D.J., Kidd, K.K., Melnick1, D.J. & Kidd, K.K. 1985. Genetic and evolutionary relationships
698   among Asian macaques. *Int. J. Primatol.* **6**: 123–160.

699   Oldt, R.F., Kanthaswamy, S., Montes, M., Schumann, L., Grijalva, J., Bunlungsup, S., *et al.* 2019.
700   Population genetics of the ABO locus within the rhesus (*Macaca mulatta*) and cynomolgus (*M.*
701   *fascicularis*) macaque hybrid zone. *Int. J. Immunogenet.* **46**: 38–48.

702   Ortiz, E.M. 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic
703   analysis. DOI:10.5281/zenodo.2540861.

704   Osada, N., Hashimoto, K., Kameoka, Y., Hirata, M., Tanuma, R., Uno, Y., *et al.* 2008. Large-scale
705   analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M.*
706   *fascicularis* and *M. mulatta*. *BMC Genomics* **9**: 90.

707   Osada, N., Uno, Y., Mineta, K., Kameoka, Y., Takahashi, I. & Terao, K. 2010. Ancient genome-wide
708   admixture extends beyond the current hybrid zone between *Macaca fascicularis* and *M. mulatta*.
709   *Mol. Ecol.* **19**: 2884–2895.

710   Osuna, C.E., Lim, S.-Y., Deleage, C., Griffin, B.D., Stein, D., Schroeder, L.T., *et al.* 2016. Zika viral
711   dynamics and shedding in rhesus and cynomolgus macaques. *Nat. Med.* **22**: 1448–1455.

712   Parnell, N.F., Hulsey, C.D. & Streelman, J.T. 2008. Hybridization produces novelty when the
713   mapping of form to function is many to one. *BMC Evol. Biol.* **8**: 122.

714   Parsons, K.J., Son, Y.H. & Albertson, R.C. 2011. Hybridization promotes evolvability in African
715   cichlids: connections between transgressive segregation and phenotypic integration. *Evol. Biol.*
716   **38**: 306–315.

717   Pereira, R.J., Barreto, F.S. & Burton, R.S. 2014. Ecological novelty by hybridization: experimental
718   evidence for increased thermal tolerance by transgressive segregation in *Tigriopus californicus*.
719   *Evolution (N. Y).* **68**: 204–215.

720   Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. 2012. Double digest RADseq:
721   An inexpensive method for de novo SNP discovery and genotyping in model and non-model
722   species. *PLoS One* **7**: e37135.

723   Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., *et al.* 2014. The complete
724   genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**: 43–49.

725   R Developmental Core Team. 2019. R: A language and environment for statistical computing. Vienna.

726   Rochette, N.C. & Catchen, J.M. 2017. Deriving genotypes from RAD-seq short-read data using
727   Stacks. *Nat. Protoc.* **12**: 2640–2659.

728   Roos, C. & Zinner, D. 2015. Diversity and evolutionary history of macaques with special focus on
729   *Macaca mulatta* and *Macaca fascicularis*. *Nonhum. Primate Nonclinical Drug Dev. Saf. Assess.*
730   3–16.

731   Roos, C., Zinner, D., Kubatko, L.S., Schwarz, C., Yang, M., Meyer, D., *et al.* 2011. Nuclear versus
732   mitochondrial DNA: evidence for hybridization in colobine monkeys. *BMC Evol. Biol.* **11**: 77.

733   Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., *et al.* 2014. The
734   genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**: 354–357.

735   Satkoski Trask, J.A., Garnica, W.T., Smith, D.G., Houghton, P., Lerche, N. & Kanthaswamy, S. 2013.
736   Single-nucleotide polymorphisms reveal patterns of allele sharing across the species boundary

737   between rhesus (*Macaca mulatta*) and cynomolgus (*M. fascicularis*) macaques. *Am. J. Primatol.*
738   **75**: 135–144.

739   Seehausen, O. 2004. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**: 198–207.

740   Sibal, L.R. & Samson, K.J. 2001. Nonhuman primates: a critical role in current disease research. *ILAR*
741   *J.* **42**: 74–84.

742   Simonti, C.N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D.S., Chisholm, R.L., *et al.* 2016. The
743   phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**: 737–
744   741.

745   Smith, D.G., Ng, J., George, D., Trask, J.S., Houghton, P., Singh, B., *et al.* 2014. A genetic
746   comparison of two alleged subspecies of Philippine cynomolgus macaques. *Am. J. Phys.*
747   *Anthropol.* **155**: 136–148.

748   Soltis, P.S. 2013. Hybridization, speciation and novelty. *J. Evol. Biol.* **26**: 291–293.

749   Stevison, L.S. & Kohn, M.H. 2009. Divergence population genetic analysis of hybridization between
750   rhesus and cynomolgus macaques. *Mol. Ecol.* **18**: 2457–75.

751   Street, S.L., Kyes, R.C., Grant, R. & Ferguson, B. 2007. Single nucleotide polymorphisms (SNPs) are
752   highly conserved in rhesus (*Macaca mulatta*) and cynomolgus (*Macaca fascicularis*) macaques.
753   *BMC Genomics* **8**: 480.

754   Svardal, H., Jasinska, A.J., Apetrei, C., Coppola, G., Huang, Y., Schmitt, C.A., *et al.* 2017. Ancient
755   hybridization and strong adaptation to viruses across African vervet monkey populations. *Nat.*
756   *Genet.* **49**: 1705–1713.

757   Svensson, O., Smith, A., García-alonso, J. & Oosterhout, C. Van. 2016. Hybridization generates a
758   hopeful monster : a hermaphroditic selfing cichlid. *R. Soc. Open Sci.* **3**: 150684.

759   Taylor, S.A. & Larson, E.L. 2019. Insights from genomes into the evolutionary importance and
760   prevalence of hybridization in nature. *Nat. Ecol. Evol.* **3**: 170–177.

761   Tosi, A.J., Morales, J.C. & Melnick, D.J. 2002. Y-chromosome and mitochondrial markers in *Macaca*
762   *fascicularis* indicate introgression with Indochinese *M. mulatta* and a biogeographic barrier in
763   the Isthmus of Kra. *Int. J. Primatol.* **23**: 161–178.

764   Tucker, P.K., Sage, R.D., Warner, J., Wilson, A.C. & Eicher, E.M. 2006. Abrupt cline for sex
765   chromosomes in a hybrid zone between two species of mice. *Evolution (N. Y).* **46**: 1146.

766   Weinbauer, G.F., Niehoff, M., Niehaus, M., Srivastav, S., Fuchs, A., Van Esch, E., *et al.* 2008.
767   Physiology and endocrinology of the ovarian cycle in macaques. *Toxicol. Pathol.* **36**: 7S-23S.

768   Xue, C., Raveendran, M., Alan Harris, R., Fawcett, G.L., Liu, X., White, S., *et al.* 2016. The
769   population genomics of rhesus macaques (*Macaca mulatta*) based on whole-genome sequences.
770   *Genome Res.* **26**: 1–12.

771   Yan, G., Zhang, G., Fang, X., Zhang, Y.Y., Li, C., Ling, F., *et al.* 2011. Genome sequencing and
772   comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus
773   macaques. *Nat. Biotechnol.* **29**: 1019–1023.

774   Yao, L., Li, H., Martin, R.D., Moreau, C.S. & Malhi, R.S. 2017. Tracing the phylogeographic history
775   of Southeast Asian long-tailed macaques through mitogenomes of museum specimens. *Mol.*
776   *Phylogenet. Evol.* **116**: 227–238.

777   Zinner, D., Arnold, M.L. & Roos, C. 2011. The strange blood: Natural hybridization in primates. *Evol.*
778   *Anthropol.* **20**: 96–103.

779   Zinner, D., Groeneveld, L.F., Keller, C. & Roos, C. 2009. Mitochondrial phylogeography of baboons
780   (*Papio* spp.): indication for introgressive hybridization? *BMC Evol. Biol.* **9**: 83.

781

782   **Data accessibility**

783   The obtained sequencing reads were deposited in the NCBI Sequence Read Archive

784   (PRJNA578019), and the data sets and code used in this study are available from the Dryad

785   public archive (https://doi.org/10.5061/dryad.1ns1r n8rf).

24

786 **Declaration of interests**

787　　　　We declare that we have no competing interests.

788 **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 Prasrinakharin Chumphon	7	9.9	99.0
	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.

789

790

Table 2. Evaluation of demographic models.

Model	MaxEstLhood	Number of parameters (K)	AIC	ΔAIC
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.

791

792

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

793

794

795   **Figures**
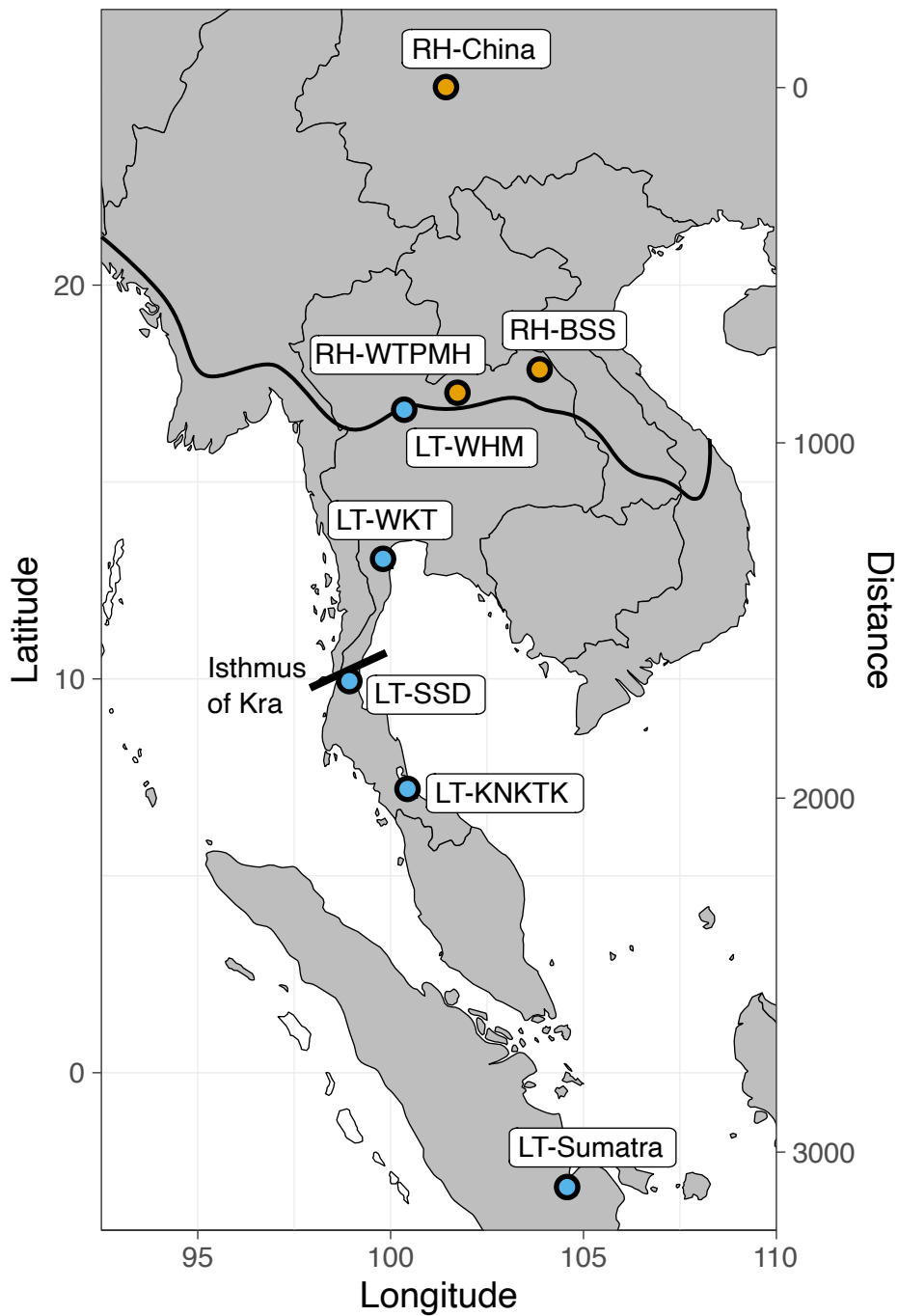
796

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
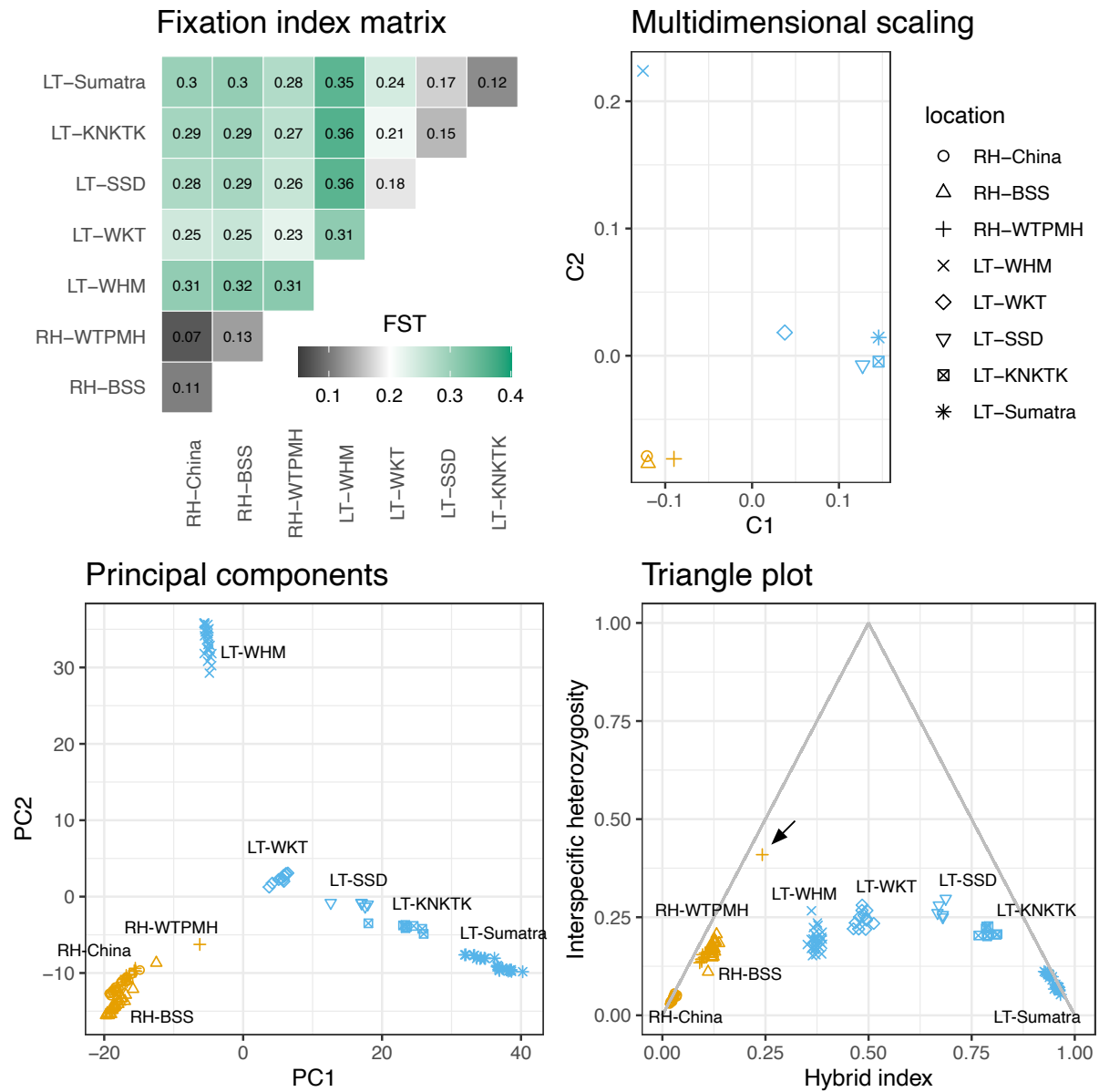
Figure 2. Pairwise F_{ST} between populations (upper left), its multidimensional scaling scores (upper right), PC scores (lower left), and the triangle plot of the hybrid index and interspecific heterozygosity (lower right). For scatter plots, species are coded by color: rhesus (orange) and long-tailed 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.
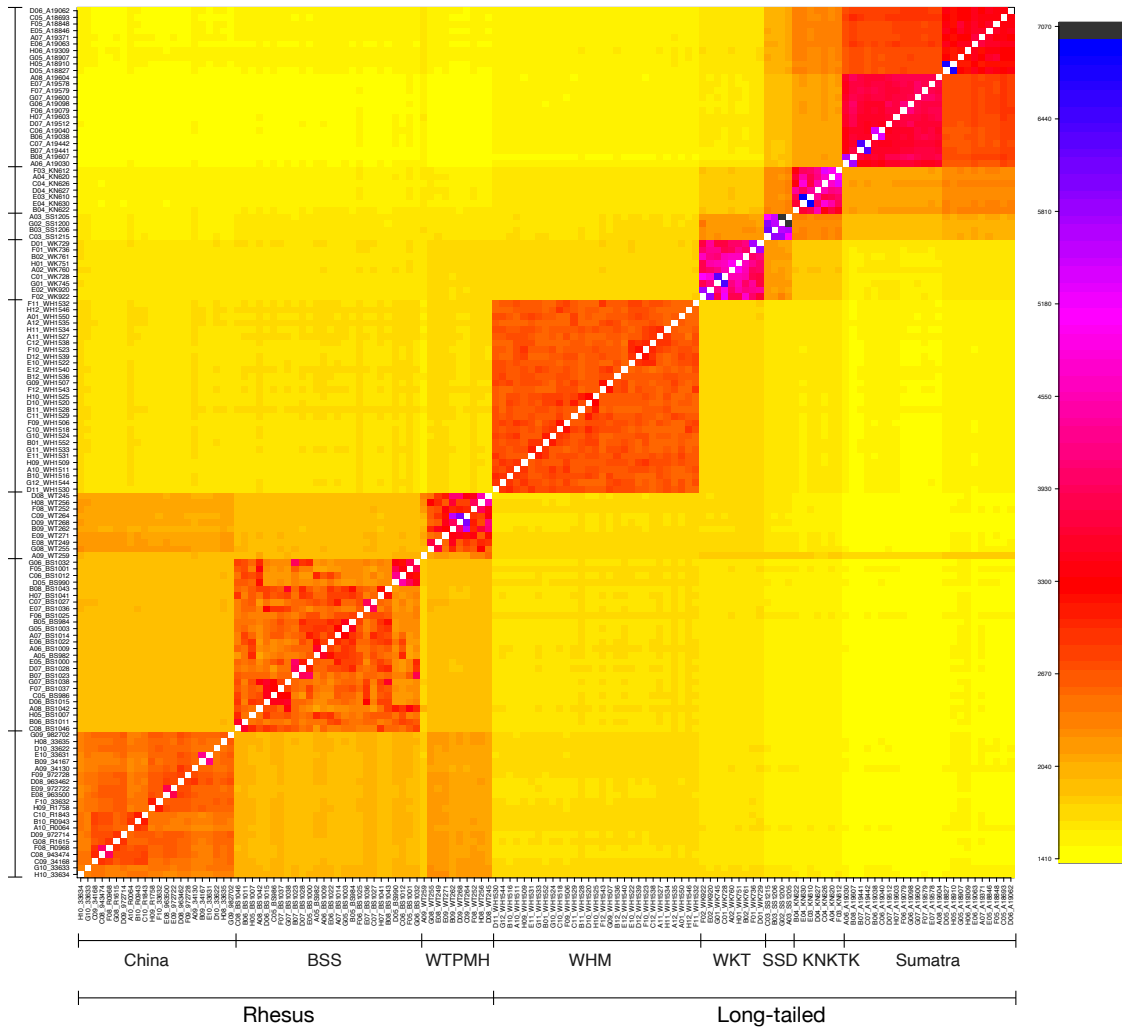
Figure 3. fineRADstructure coancestry matrix. The heatmap of fineRADstructure depicts variation in pairwise coancestry between individuals according to the scale shown on the right.
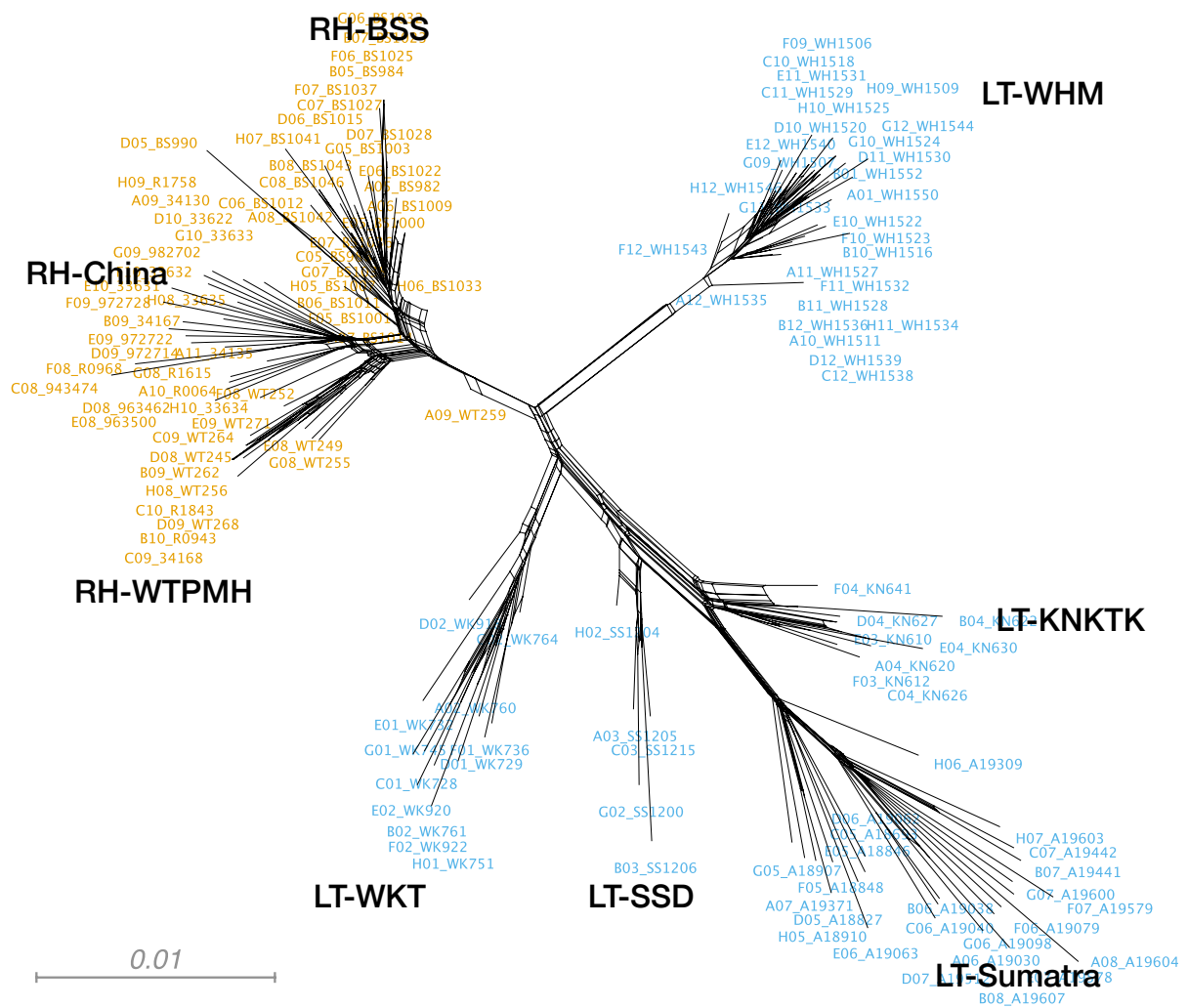
Figure 4. Phylogenetic networks based on the neighbor-net algorithm. The name of the operational taxonomic unit (sample ID) is color-coded by species: rhesus (orange) and long-tailed macaques (sky blue).
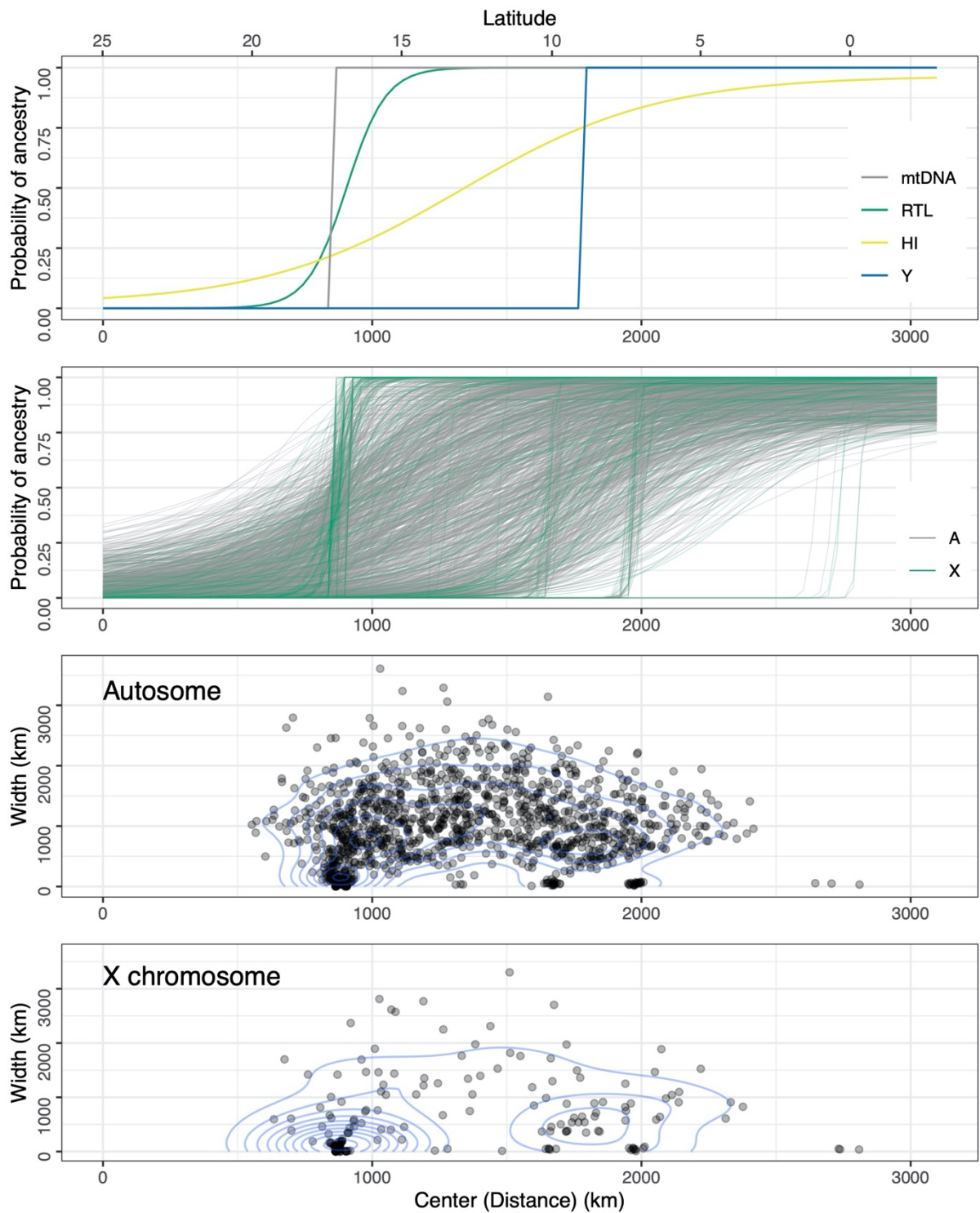
Figure 5. Geographic clines. The top panel denotes the geographic clines for the hybrid index
of diagnostic markers (HI; gray), mitochondrial DNA (mtDNA; green), relative tail
length (RTL; yellow), and Y-chromosome (Y; blue). The second panel denotes the
geographic clines for the allele frequency of 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
Cline center is south of the Isthmus of Kra (10° N)		
Yes	54	334
No	134	901
Cline center is around the interspecific boundary (100 km north–south range)		
Yes	46	230
No	142	1005

832

Autosome (K = 5)
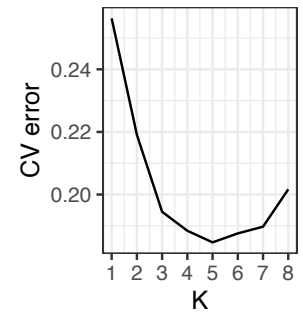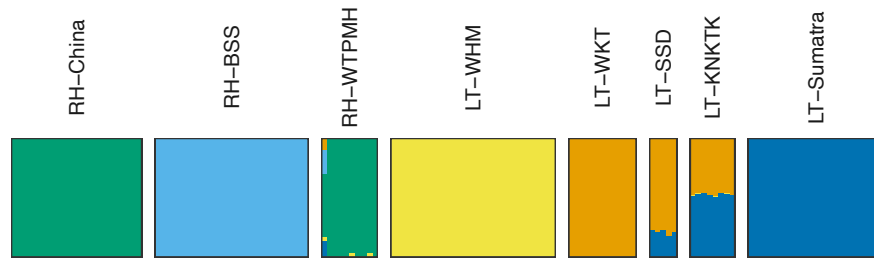
X chromosome (K = 8)
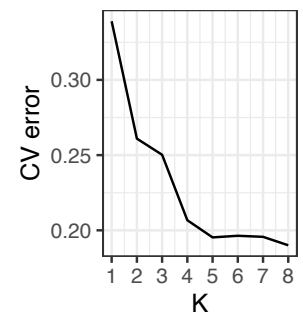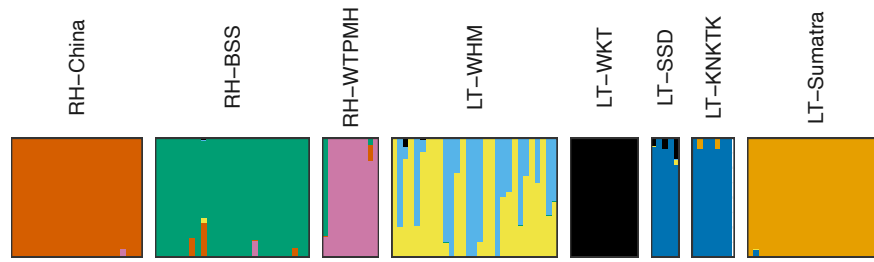
Figure S1 ADMIXTURE barplot of auosome (a) and X-chromosome (b). The cross-validation error is the smallest when $K = 5$ for autosome and $K = 8$ for X-chromosome.
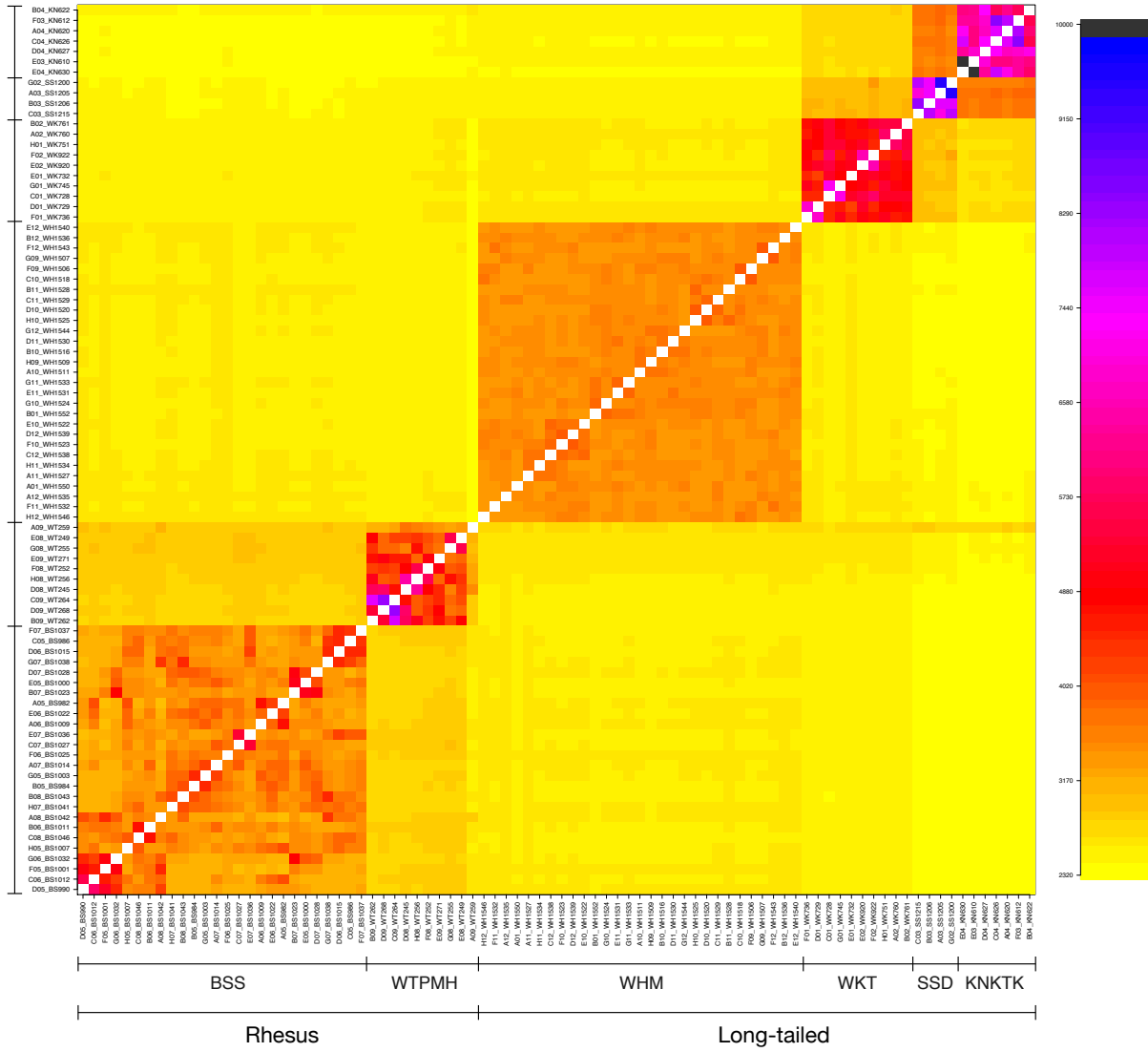
Figure S2 fineRADstructure coancestry matrix for wild-derived samples only. The heatmap of fineRADstructure depicts variation in pairwise coancestry between individuals according to the scale shown on the right.
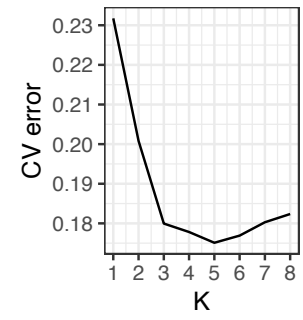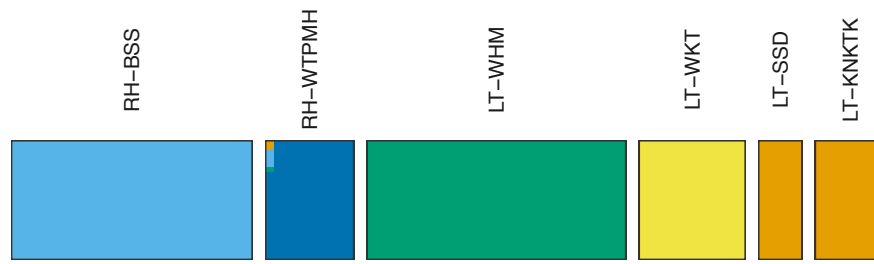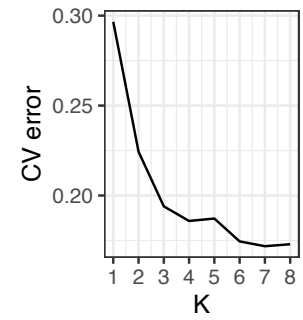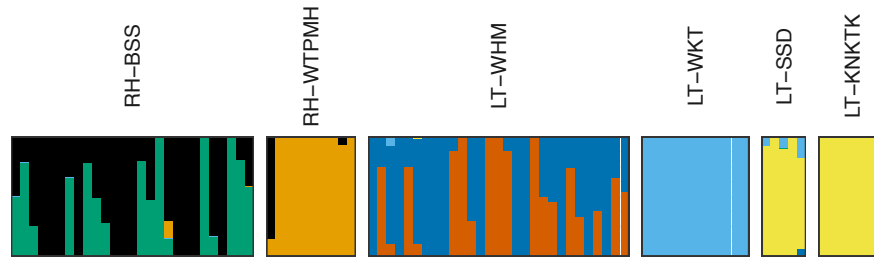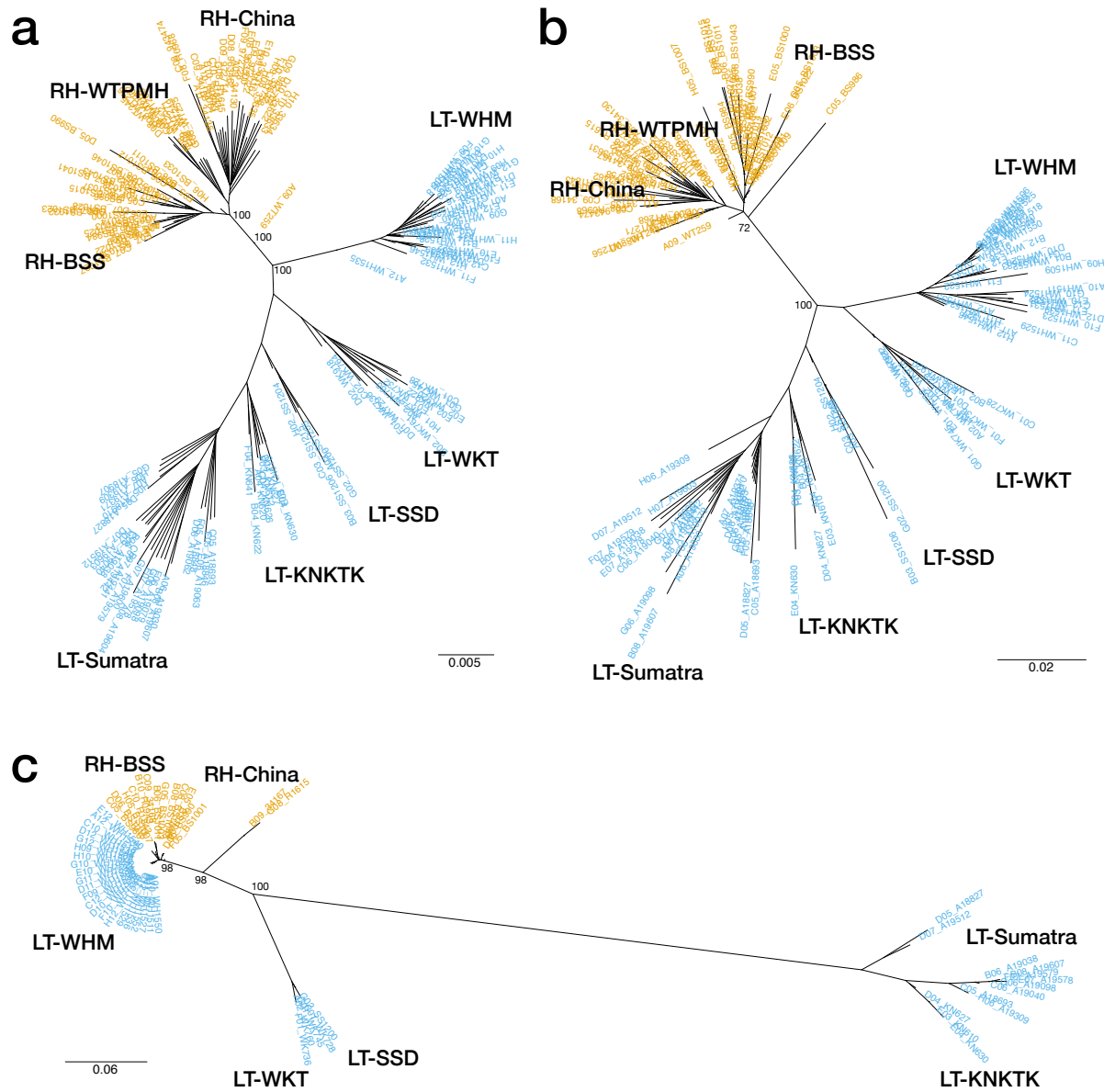
Figure S3 ADMIXTURE barplot of auosome (a) and X-chromosome (b) for wild-derived samples only. The cross-validation error is the smallest when $K = 5$ for autosome and $K = 7$ for X-chromosome.
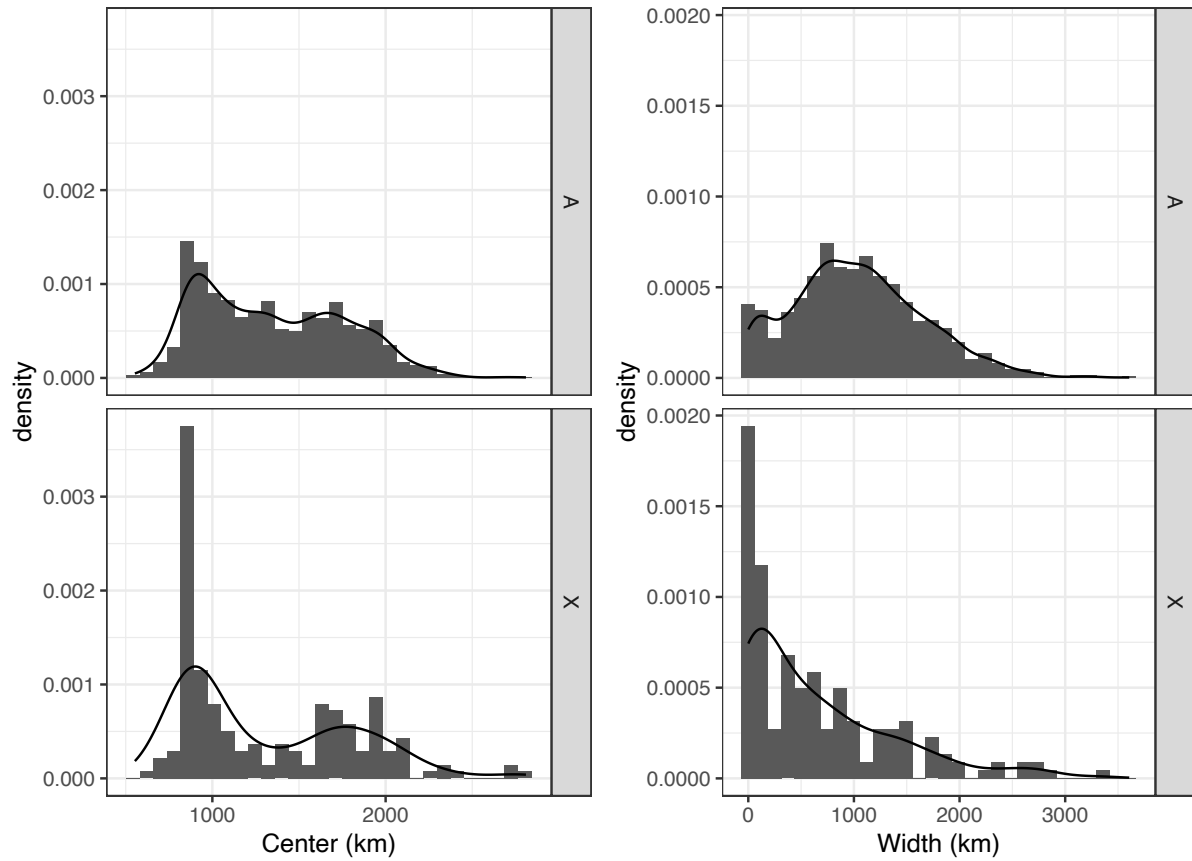
Figure S4 Neighbor-joining tree of autosome (a), X-chromosome (b), and Y-chromosome (c).
Bootstrap support values are shown on the nodes of major clades.

851

852 **Figure S5** Density histograms of cline centers (left) and widths (right) overlayed by their kernel density

853 profiles of autosomes (A, upper) and X-chromosomes (X, lower), respectively.

854

855

856