Phylogeographic history of Japanese macaques

Running title: Phylogeography of Japanese macaques

Tsuyoshi Ito[1, *], Takashi Hayakawa[2, 3], Nami Suzuki–Hashido[4], Yuzuru Hamada[1], Yosuke Kurihara[1, †], Goro Hanya[1], Akihisa Kaneko[1], Takayoshi Natsume[1], Seitaro Aisu[1], Takeaki Honda[1], Syuji Yachimori[5], Tomoko Anezaki[6], Toshinori Omi[7], Shin-ichi Hayama[7], Mikiko Tanaka[1], Hikaru Wakamori[1, ‡], Hiroo Imai[1], Yoshi Kawamoto[1, §]

[1] Primate Research Institute, Kyoto University, Kanrin 41-2, Inuyama, Aichi 484-8506, Japan

[2] Faculty of Environmental Earth Science, Hokkaido University, Kita-ku N10W5, Sapporo, Hokkaido 0060-0810, Japan

[3] Japan Monkey Centre, Kanrin 26, Inuyama, Aichi 484-0081, Japan

[4] Chubu University Academy of Emerging Sciences, Matsumoto-cho 1200, Kasugai, Aichi 487-8501, Japan

[5] Shikoku Institute of Natural History, Shimobun-otsu 470-1, Susaki, Kochi 785-0023, Japan

[6] Gunma Museum of Natural History, Kamikuroiwa 1674-1, Tomioka, Gumma 370-2345, Japan

[7] Faculty of Veterinary Science, Nippon Veterinary and Life Science University, Kyonan-cho 1-7-1, Musashino, Tokyo 180-8602, Japan

Current address:

[†] Center for Education and Research in Field Sciences, Faculty of Agriculture, Shizuoka University, Kariyado 63, Fujieda, Shizuoka 426-0001, Japan

[‡] Education and Public Relations Department, Tama Zoological Park, Tokyo Zoological Park Society, Hodokubo 7-1-1, Hino-shi, Tokyo 191-0042, Japan

[§] Faculty of Veterinary Science, Nippon Veterinary and Life Science University, Kyonan-cho 1-7-1,

Musashino, Tokyo 180-8602, Japan

* Correspondence: ito.tsuyoshi.3a@kyoto-u.ac.jp

Abstract

Aim: Understanding patterns and processes of geographic genetic variation within and among closely related species is the essence of phylogeography. Japanese macaques, also called snow monkeys, have been extensively studied, particularly in the fields of sociobiology, ecology, and experimental biology; however, our knowledge of their evolutionary history is relatively limited. In this study we aimed to elucidate the geographic patterns of genetic variation in Japanese macaques and the processes that underlie them.

Location: Japan

Taxon: Japanese macaque, *Macaca fuscata*; rhesus macaque, *M. mulatta*; Taiwanese macaque, *M. cyclopis*

Methods: Double-digest restriction-site associated DNA (RAD) sequencing was used to identify genome-wide single nucleotide variants. We used FINERADSTRUCTURE, ADMIXTURE, and principal component analyses to estimate the genetic population structure. Phylogenetic relationships were then inferred based on neighbour-net, neighbour-joining, maximum likelihood, and SVDquartets algorithms. We assessed gene flow using demographic inference and ABBA-BABA tests, and estimated past distributions during the Last Glacial Maximum (LGM) using ecological niche modelling.

Results: Japanese macaques show a sister group relationship with a clade comprising Chinese rhesus, Indian rhesus, and Taiwanese macaques. Japanese macaques comprise major northeastern and southwestern clades, with a boundary located near central Japan, and gene flow between the northeastern and southwestern lineages was detected. Refugia during the LGM were estimated to be distributed in limited areas along the south coasts of the Japanese archipelago.

Main Conclusions: Phylogeographic variation of Japanese macaques is likely due mainly to northeast–southwest divergence, which resulted from withdrawal into refugia during the glacial period, and

subsequent gene flow.

Keywords: ecological niche modelling, gene flow, phylogeny, population genomics, primates, refugia

1. Introduction

A key goal of phylogeography is to understand the geographic patterns of genetic variation within and among closely related species and the processes that underlie them. Evidence accumulating from phylogeographic studies suggests that observed genetic structures strongly reflect the repeated geographic retractions and expansions of populations that accompanied the climatic oscillations of the Pleistocene epoch (Hewitt, 2000). The phylogeography of the Japanese archipelago, located on the eastern edge of Asia, has been extensively studied because it is a biodiversity hotspot (Zachos & Habel, 2011), containing many endemic species (Motokawa & Kajihara, 2017; Ohdachi, Ishibashi, Iwasa, Fukui, & Saitoh, 2015). Many temperate animals show a northeast–southwest (or east–west) divergence in genetic structure, the boundaries of which are located near the Chugoku–Chubu regions of the main island (Honshu), even though there is no apparent geophysical barrier (see Tamate, 2013 for a review of mammals; Appendix S1; Fig. 1). It is suggested that this northeast–southwest (or east–west) divergence pattern resulted from either the isolation of populations into two (or more) refugia during the glacial period or multiple migrations of continental ancestors to the Japanese archipelago (e.g., Kawamoto, 2010; Nozawa et al., 1991). However, previous genetic analyses were mostly restricted to partial sequences of mitochondrial DNA (mtDNA) or a small number of nuclear DNA (nDNA) markers (AppendixS1), and fine-scale genome-wide structures reflecting reticulate evolutionary events, such as recombination, incomplete lineage sorting, and horizontal gene transfer, are poorly understood.

Japanese macaques (*Macaca fuscata*), also known as snow monkeys, are one endemic Japanese mammal taxon that shows a northeast–southwest divergence in mtDNA haplotypes (Kawamoto et al., 2007). *Macaca fuscata* is distributed from 30°15' N to 41°30' N (Fig. 1), and forms the northernmost reach of nonhuman primates (Fooden & Aimi, 2005), making it a useful primate model for cold adaptation in humans (e.g., Buck et al., 2019; Buck, De Groote, Hamada, & Stock, 2018). *Macaca fuscata* has been extensively studied, particularly in the fields of ecology and sociobiology, since the first Japanese primatology studies were conducted in 1948 (Matsuzawa & McGrew, 2008; Nakamura, 2009; Yamagiwa, 2010). The macaques are also common subjects in biomedical, physiological, and biopsychological studies (e.g., Craig et al., 2012; Isa, Yamane, Hamai, & Inagaki, 2009; Ma et al., 2014). Ecogeographical variations in functional host genetics and

85 microbiomes have also been analysed (Hayakawa et al., 2018; Inoue-Murayama, Inoue, Watanabe,
86 Takenaka, & Murayama, 2010; Lee, Hayakawa, Kiyono, Yamabata, & Hanya, 2019; Suzuki-Hashido
87 et al., 2015). However, in contrast to extensive knowledge of their present biology, our understanding
88 of their evolutionary history is limited. Female philopatry and male-biased dispersal in macaques
89 means that their mtDNA tree reflects patterns of maternal group divergence; it does not retain traces of
90 male-mediated gene flow. Hence, mtDNA and nDNA should show different tree topologies. Partly
91 because of this, there remain some unresolved issues regarding their phylogeographic history, which
92 we discuss as below (for reviews, see Fooden & Aimi, 2005; Kawamoto, 2010).

93 The first issue is the origin of *M. fuscata*. Mitochondrial DNA trees indicate that *M. fuscata*
94 derived from the eastern (Chinese/Indochinese) population of *M. mulatta* (rhesus macaques) after the
95 east–west divergence of *M. mulatta*; that is, *M. mulatta* is a paraphyletic relative of *M. cyclopis*
96 (Taiwanese macaques) and *M. fuscata* (Chu, Lin, & Wu, 2007; Melnick, Hoelzer, Absher, & Ashley,
97 1993; Morales & Melnick, 1998; Rogers et al., 2019). In contrast, Y-chromosome trees indicate that
98 *M. fuscata* has a sister relationship with a clade comprising eastern (Chinese/Indochinese) *M. mulatta*,
99 western (Indian/Pakistan) *M. mulatta*, and *M. cyclopis*; that is, *M. mulatta* is monophyletic (Tosi,
100 Morales, & Melnick, 2000, 2002, 2003). Tosi et al. (2000, 2002, 2003) suggested that mtDNA-Y
101 discordance can be attributed to the incomplete lineage sorting of mtDNA divergence between the
102 eastern and western populations of *M. mulatta*, followed by male-mediated nuclear gene flow between
103 them, and the divergence of *M. cyclopis* and *M. fuscata* from the eastern *M. mulatta* population.
104 Although this seems plausible considering the philopatric nature of female macaques and the wide
105 distribution of *M. mulatta*, further validation using a larger number of nDNA markers is required
106 before reaching a conclusion.

107 Secondly, phylogenetic relationships among local populations of *M. fuscata* are a matter of
108 debate. As mentioned above, mtDNA trees indicate major northeastern and southwestern clades with a
109 boundary around the Chugoku–Kinki region of Honshu (Kawamoto et al., 2007). In contrast, blood
110 protein variants (encoded by autosomal DNA) indicate that populations isolated on small islands or
111 peripheral peninsulas show significant differentiation, while populations on the main islands (Honshu,
112 Shikoku, and Kyushu) are homogeneous (Tomari, 2003 cited in Kawamoto, 2010; Nozawa et al.,
113 1996, 1991). The significant differentiation of blood protein variants in island or peripheral
114 populations was originally interpreted as the consequence of multiple migrations of ancestors from the
115 continent to the Japanese archipelago (Nozawa et al., 1991); Kawamoto (2010) suggested that this
116 pattern may also be explained by population bottlenecks in these small, isolated populations.

Supporting the latter scenario, extremely low genetic diversities are observed in the autosomal and Y-chromosome microsatellite markers of the peripheral Shimokita Peninsula population (Kawamoto, Tomari, Kawai, & Kawamoto, 2008) and the mtDNA of the small Yakushima Island population (Hayaishi & Kawamoto, 2006). Alternatively, and inconsistent with the bottleneck hypothesis, genetic diversity in blood proteins is significantly high for the Shimokita population (Hayasaka, Kawamoto, Shotake, & Nozawa, 1987; Kawamoto et al., 2008). As these studies relied on a relatively small number of loci (dozens or less), the genetic population structure of the nDNA variation in *M. fuscata* remains open for investigation.

The third issue is the taxonomic status of the two recognized subspecies: *M. f. yakui* Kuroda, 1940 and *M. f. fuscata* Gray, 1870. Kuroda (1940) regarded the Yakushima population as distinct on the basis of darker pelage colouration (particularly on the back, limbs, and extremities) and smaller body size compared to *M. f. fuscata*. Although these variations in pelage colour and body size (at least in adults) overlap considerably between the two populations (Fooden & Aimi, 2005; Hamada, Watanabe, & Iwamoto, 1992, 1996), the pelage of the dorsal surfaces of the hands is distinctively darker than that of the dorsal surface of the trunk for *M. f. yakui* but not for *M. f. fuscata* (Fooden & Aimi, 2005). Cranial shape is also different: *M. f. yakui* has narrower orbits, a more protruding snout, more prominent postorbital constriction, and a broader zygomatic arch than *M. f. fuscata* (Ikeda & Watanabe, 1966; Yano, Egi, Takano, & Ogihara, 2020). However, previous molecular evidence has not supported the sister relationship of the two subspecies. In the mtDNA tree, *M. f. yakui* is placed within the southwestern clade of *M. f. fuscata* (Kawamoto et al., 2007). For the blood protein variants, *M. f. yakui* is relatively differentiated but falls within the range of variation of *M. f. fuscata* (Tomari, 2003 cited in Kawamoto, 2010; Nozawa et al., 1996, 1991). A recent study showed that simian foamy viruses found in *M. f. yakui* is distinct from most *M. f. fuscata*, *M. mulatta*, *M. cyclopis* and *M. fascicularis* (long-tailed, crab-eating, or cynomolgus macaques) (Hashimoto-Gotoh, Yoshikawa, Nakagawa, Okamoto, & Miyazawa, 2020). Thus, it remains debatable whether *M. f. yakui* is the sister taxon to *M. f. fuscata* or belongs within the subspecies.

To resolve these issues, the current study estimated the phylogenetic relationships and genetic population structures of *M. fuscata* and related species (eastern [Chinese] *M. mulatta*, western [Indian] *M. mulatta*, *M. cyclopis*, and *M. fascicularis*) based on genome-wide nDNA variants. For *M. fuscata*, we examined ten different populations, including Yakushima, the southern-most population, and Shimokita, the northernmost population (Table 1; Fig. 1). For genome-wide genotyping, we applied double-digest restriction-site associated DNA sequencing (ddRADseq) (Peterson, Weber, Kay,

149 Fisher, & Hoekstra, 2012). We also estimated the past distribution of *M. fuscata* to identify the
150 locations of potential refugia using ecological niche modelling. Based on these analyses, we inferred
151 the processes that underlie geographic patterns of genetic variation in *M. fuscata*.

152 ## 2. Materials and Methods

153 ### 2.1. Samples and ddRADseq

154 We used 113 blood- or tissue-derived DNA samples from 10 populations of *M. fuscata* (*N* =
155 102), the Chinese and Indian populations of *M. mulatta* (*N* = 6), *M. cyclopis* (*N* = 2), and *M.*
156 *fascicularis* (*N* = 3) (Fig 1; Tables 1 and S1; Appendix S2). The ddRAD library preparation followed
157 the quaddRAD procedure of Franchini et al. (2017) with minor modifications (Tables S2 and S3;
158 Appendix S3). Raw sequence reads were cleaned and demultiplexed using the STACKS 2.4 program
159 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Rochette & Catchen, 2017). The reads were
160 mapped to the RefSeq of *M. mulatta* (Mmul_10 [GCF_003339765.1]) using the BOWTIE2 program
161 (Langmead & Salzberg, 2012) and filtered to retain only high-quality reads (mapping quality of ≥ 20)
162 using the SAMTOOLS 1.9 programme (Li et al., 2009). Because some samples either lacked information
163 on sex or may have been sexed erroneously, sex was identified in terms of the ratio of numbers of
164 filtered reads mapped onto the Y-chromosome versus those mapped to the X-chromosome using
165 sam_sexing.py (https://github.com/itots/sam_sexing) (Fig. S1). Genotypes were called for autosomes,
166 X-, and Y-chromosomes using STACKS 2.5 (genotype quality of ≥ 20). The diploid single nucleotide
167 variant (SNV) calls on the male sex chromosomes were converted to haploid format (Appendix S4).
168 The SNVs were filtered using PLINK 1.90b4 (Purcell et al., 2007) with respect to minor allele count,
169 missing rate, linkage disequilibrium, and deviation from Hardy–Weinberg equilibrium (Appendix S4).
170 For the phylogenetic analyses, the SNV data (vcf format) was reformatted in nexus or phylip, wherein
171 heterogeneous sites were coded according to the IUPAC nucleotide ambiguity code using
172 vcf2phylip.py (https://github.com/itots/vcf2phylip). These SNV datasets were used unless stated
173 otherwise.

174 ### 2.2. Genetic diversity and population structure

175 Nucleotide diversity was assessed for each chromosome/population using the 'ape' package
176 (Paradis, Claude, & Strimmer, 2004), the 'pegas' package (Paradis, 2010), and a custom script written
177 in R (R Developmental Core Team, 2019) (Appendix S5). The genetic population structure in *M.*
178 *fuscata* was inferred based on the FINERADSTRUCTURE (Lawson, Hellenthal, Myers, & Falush, 2012;
179 Malinsky, Trucchi, Lawson, & Falush, 2018), ADMIXTURE (Alexander & Novembre, 2009), and

principal component analysis (PCA) (Jombart, 2008) (Appendix S6).

2.3. Phylogenetic analyses

The phylogenetic relationships between the populations of *M. fuscata* and the other three species were inferred using a neighbour-joining (NJ) algorithm (Saitou & Nei, 1987) from PAUP* 4.0a166 (https://paup.phylosolutions.com/); the maximum-likelihood (ML) method using RAXML-NG 0.90 (Kozlov, Darriba, Flouri, Morel, & Stamatakis, 2019); the singular value decomposition for quartets (SVDquartets) method (Chifman & Kubatko, 2014) using PAUP* 4.0a166; and the neighbour-net method using SPLITSTREE4 (Huson, 1998; Huson & Bryant, 2006) (Appendix S7).

2.4. Demographic inference and analysis of gene flow among populations

Site frequency spectrum-based demographic inference was analysed using autosomal SNVs and FASTSIMCOAL2 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013) (Appendix S8). For SNV filtering, only *M. fuscata* samples were extracted, singletons were not removed, and only loci with no missing data were retained after removing individuals with ≥ 10% missing genotype data; otherwise, the site frequency spectrum may be skewed. The other filtering parameters were the same as the main analysis. Three groups were assumed, namely northeastern (Shimokita, Yamagata, and Gunma), southwestern (Shiga, Takahama, Arashiyama, Wakasa, Kochi, and Kojima), and Yakushima, and eight demographic models with various settings for divergence patterns, migrations, and population size changes were tested (Fig. S2). Model comparisons were conducted based on the Akaike information criterion (AIC). The sample sizes (the number of individuals) of the northeastern, southwestern, and Yakushima groups was reduced to 10, 12, and 6 by random sampling, which was repeated five times to assess sampling bias. In the FASTSIMCOAL2 estimate, the effective population size of the southwestern group was fixed at $N_e = 16{,}650$, which was calculated as $N_e = \pi/4\mu$, where π is the mean nucleotide diversity of autosomes (Appendix S5) and μ (the mutation rate per site per generation) was assumed to be 1.0×10^{-8} (Xue et al., 2016). Population splits and admixture were inferred from the genome-wide allele frequency data calculated from the autosomal SNVs using TREEMIX 1.13 (Pickrell & Pritchard, 2012) (Appendix S8). To assess bias caused by missing data, analyses of the full dataset and the subset with no missing data, which were extracted after removing individuals with ≥ 20% of the genotype missing, were performed. Finally, to differentiate between incomplete lineage sorting and gene flow, ABBA-BABA tests (Durand, Patterson, Reich, & Slatkin, 2011) based on the autosomal SNVs dataset were performed using DSUITE 0.3r21 software (Malinsky, 2019) (Appendix S8). In these analyses, the Shiga, Takahama, Arashiyama, and Wakasa samples were treated as a single population, namely Kinki, because none could be clearly distinguished by genetic

212 population structure or phylogenetic analysis.

213 **2.5. Ecological niche modelling**

214 We performed ecological niche modelling to estimate the past distribution of *M. fuscata*
215 during the glacial period. Data for the current distribution were obtained from the second (FY1978)
216 and sixth (FY2003) distribution surveys of Japanese animals (mammals) by the Biodiversity Center of
217 Japan, Nature Conservation Bureau, Ministry of the Environment (http://gis.biodic.go.jp/webgis/). The
218 data consist of 4,111 presence records (Fig. 1). We predicted the current and past (the Last Glacial
219 Maximum [LGM]) distributions based on the maximum entropy method (Maxent; Phillips, Anderson,
220 & Schapire, 2006) using the 'ENMeval' package (Muscarella et al., 2014) in R, wherein nine
221 BIOCLIM variables (bio2, 3, 4, 8, 9, 10, 15, 17, and 18) (Hijmans, Cameron, Parra, Jones, & Jarvis,
222 2005) were the explanatory variables (Appendix S9).

223 **3. Results**

224 After filtering, the main SNV dataset consisted of 113,665 variants of 112 individuals for
225 autosomes (total genotyping rate = 0.92), 3,794 variants of 112 individuals for X-chromosomes (0.91),
226 and 133 variants of 55 individuals for Y-chromosomes (0.92) (Table S4).

227 **3.1. Genetic diversity and population structure**

228 Genetic diversity was generally much lower in populations of *M. fuscata* compared with
229 Chinese *M. mulatta*, Indian *M. mulatta*, *M. cyclopis*, and *M. fascicularis* (Tables S5 and S6). In *M.*
230 *fuscata*, genetic diversity was similar among all populations except for Yakushima (*M. f. yakui*), which
231 showed extremely low values. In all the populations, the Y-chromosomes showed much lower
232 nucleotide diversity than the autosomes, with the X-chromosomes showing intermediate diversity
233 (Table S6). Nucleotide diversity of Y-chromosomes was generally much higher in wild populations
234 (Shimokita, Yamagata, Gunma, and Kojima) than in captive populations (Takahama, Arashiyama, and
235 Wakasa).

236 The FINERADSTRUCTURE identified two major clusters in *M. fuscata*: northeastern
237 (Shimokita, Yamagata, and Gunma) and southwestern (the other populations) (Fig. 2). In the
238 northeastern cluster, the three populations showed high levels of shared ancestry (i.e., nearest
239 neighbour haplotype). In the southwestern cluster, the Yakushima population (*M. f. yakui*) appeared to
240 be differentiated, shared little co-ancestry with the other populations, and was highly homogeneous.
241 The four populations located around the Kinki region (Shiga, Takahama, Arashiyama, and Wakasa)

were closely related to each other and they shared ancestry to some extent with the northeastern populations, particularly Yamagata and Gunma. One sample (lib2-3-E) labelled as Wakasa population was classified within the cluster of Takahama; this sample could have been erroneously replaced by a sample belonging to the Takahama population during sample preparation (sex identification based on mapping rate was also discordant with the sample label; Table S1). The ADMIXTURE analysis identified seven populations in *M. fuscata* (the cross-validation error was the smallest when $K = 7$). This, together with the results of the PCA, indicating two major northeastern and southwestern clusters and the differentiation of the Yakushima population, supported the findings of the FINERADSTRUCTURE (Figs. 3 and S3).

3.2. Phylogenetic relationships

For interspecific relationships, the phylogenetic networks, NJ and ML trees, and individual and population SVDquartets trees all showed with strong bootstrap support that *M. fuscata* is monophyletic and is the sister taxon to the clade comprising Chinese *M. mulatta*, Indian *M. mulatta*, and *M. cyclopis* (Figs. 4 and S4–S8). All trees supported the monophyly of *M. mulatta*.

In *M. fuscata*, the northeastern (Shimokita, Yamagata, and Gunma) and southwestern (the other populations) clades were recovered from the NJ and ML analyses, and from individual SVDquartets trees of autosomal SNVs; i.e., the monophyly of both the northeastern and southwestern clades was strongly supported (Figs. 4, S6, and S7). The phylogenetic networks were star-like and highly reticulated in *M. fuscata*, but the major edges supported the northeast–southwest divergence (Fig. 4). For the Y-chromosome tree, the boundary between the northeastern and southwestern groups was located further east than were those of the autosomes; the northeastern clade encompassed the Takahama, Arashiyama, Wakasa, and Kochi populations that were placed within the southwestern clade of the autosomal trees (Fig. S5). In the NJ tree of the X-chromosomes and the population SVDquartets tree of the autosomes, although the monophyly of the northeastern clade (Shimokita, Yamagata, and Gunma) was strongly supported, the relationships among the other populations could not be resolved (Figs. S4 and S8).

The Yakushima population (*M. f. yakui*) was placed within the southwestern clade in all trees (Figs. 4 and S4–S7), with the exception of the population-level SVDquartets tree (Fig. S8), where the position of the Yakushima population could not be resolved (Fig. S8).

3.3. Demography and gene flow among populations

The FASTSIMCOAL2 analysis strongly supported the isolation-with-migration models, which

assumed gene flow between the northeastern and southwestern groups, and between the Yakushima and southwestern groups, rather than the isolation models (Table S7; Fig. S9). Three of five replicates supported the scenario that the northeastern and southwestern groups diverged first, followed by branching of the Yakushima lineage from the southwestern group (model 6 or 7), while another replicate supported the scenario that the Yakushima lineage diverged first (model 3), and the other replicate supported models 3 and 7 almost equally. In these isolation-with-migration models, the initial divergence time (T_{D1}) was estimated as ca. 9,300–12,000 generations ago, which corresponds to ca. 0.06–0.13 Ma [assuming a generation time of 6.5–11 years (Hernandez et al., 2007; Xue et al., 2016)]. The models were not improved by assuming recent expansions in population size or the Yakushima population bottleneck. In the TREEMIX analysis, the major northeast–southwest clades were observed (Yakushima is placed in the southwestern group) when assuming no migration events; assuming 1–3 migration events, however, the southwestern group became paraphyletic (Figs. S10 and S11). The ABBA-BABA tests detected significant gene flow between the southwestern and northeastern clade populations (Fig. 5; Tables S8–S10). The degree of inter-clade gene flow varies among southwestern populations, being generally greater in the east.

3.4. Past distribution of *M. fuscata* during the glacial period

The model with the lowest small-sample corrected AIC (AICc) value (feature class, LQHPT; regularisation parameter, 0.7) showed a moderate value for the average test area under the curve (0.69) and recovered the distribution of *M. fuscata* relatively well (Figs. 6, S12, and S13). The prediction based on this model showed that most areas of the Japanese archipelago were not suitable for *M. fuscata* during the LGM. Suitable areas were restricted to the southern coasts of the Japanese archipelago, although they were predicted to be relatively widespread in the southwestern part of the archipelago in the MPI-ESM-P model.

4. Discussion

4.1. Origin of *M. fuscata*

The trees inferred from the genome-wide nDNA SNVs are congruent with previously reported Y-chromosome trees (Tosi et al., 2000, 2002, 2003), suggesting that *M. fuscata* is the sister taxon to the clade comprising Chinese *M. mulatta*, Indian *M. mulatta*, and *M. cyclopis*. Thus, the origin of *M. fuscata* can be traced to the stem of the *M. mulatta* group. The whole nuclear genome phylogeny recently published by Osada et al. (2020) also supports this branching pattern, suggesting that the present Chinese *M. mulatta* population was formed by male-mediated gene flow from

ancestral Chinese *M. mulatta* to a ghost population generated by admixture between *M. fuscata* and *M. cyclopis* populations. A resulting nuclear swamping was likely the cause of the nDNA-mtDNA topological discordance. The *M. fuscata* divergence time has been estimated at 0.31–0.88 Ma (Marmi, Bertranpetit, Terradas, Takenaka, & Domingo-Roura, 2004); 0.38–0.42 Ma (Chu et al., 2007); 0.65–0.73 Ma (Hayasaka, Fujii, & Horai, 1996); 1.45 Ma (1.14–1.78) (Roos, Kothe, Alba, Delson, & Zinner, 2019); and 1.0 Ma (± 0.4) (Tosi et al., 2003). The latter estimate was based on a Y-chromosome marker, while the others used mtDNA. These dates are comparable to or older than the age when the land bridge was presumably formed in the Korea (Tsushima) Strait, and the oldest fossil of *M. fuscata* (Yamaguchi Prefecture, Japan) probably comes from (ca. 0.43–0.63 Ma; Aimi, 2002; Dobson & Kawamura, 1998; Fooden & Aimi, 2005). Combining these findings leads to the following hypothesis: proto-*M. fuscata* diverged from the stem of the *M. mulatta* group on the Asian continent, with some lineages contributing to the source of Chinese *M. mulatta* and others dispersing onto the Japanese archipelago via a land bridge formed in the Korea Strait during glacial periods ca. 0.43–0.63 Ma or earlier.

4.2. Phylogeographic diversification on the Japanese archipelago

Once established in the north Kyushu–west Honshu area, *M. fuscata* was probably able to disperse throughout Honshu-Shikoku-Kyushu main islands (and some surrounding small islands), which were often connected during sea level low-stands during the Middle and Late Pleistocene (Fig. S14; NOAA National Geophysical Data Center, 2009; Rohling et al., 2014). In this study, the northeastern and southwestern clades were recovered in all individual-level autosomal trees. Ecological niche modelling indicated that the ancestral populations were isolated within refugia along the southern coasts of the Japanese archipelago during the LGM (or possibly a former glacial period), resulting in a northeast–southwest divergence, although this interpretation requires caution because the data did not strictly meet a random sample design. The much lower genetic diversity of *M. fuscata* in comparison with its continental relatives suggests the presence of a strong bottleneck during dispersal within the Japanese archipelago and/or isolation in refugia. The time of the initial population divergence was estimated as ca. 0.06–0.13 Ma by demographic inference. Therefore, the current geographic diversity of *M. fuscata* was probably formed during and after the last or the second-last glacial period, which is much later than the presumed age of *M. fuscata* dispersed onto the Japanese archipelago (ca. 0.43-0.63 Ma). The boundary between the northeastern and southwestern clades is located further east (between Shiga and Gunma) by nDNA than mtDNA (Chugoku–Kinki region) (Kawamoto et al., 2007). This discrepancy in the position of the boundary can be attributed to either incomplete lineage sorting of mtDNA or male-mediated nuclear gene flow and resulting nuclear

swamping. In fact, the demographic inference and ABBA-BABA results strongly suggest that gene flow took place between the northeastern and southwestern lineages. Although there remains a possibility that the southwestern group is paraphyletic (population-level autosomal trees and X-chromosome trees did not strongly support monophyly), the geographic diversity of *M. fuscata* is probably a result of divergence and gene flow between the northeastern and southwestern lineages.

4.3. Taxonomic status of the subspecies

Macaca f. yakui is positioned within the southwestern clade of *M. f. fuscata* in all individual-level autosomal trees (i.e., *M. f. fuscata* is paraphyletic), supporting the mtDNA tree (Kawamoto et al., 2007). Star-like and reticulated phylogenetic networks suggest that the populations, including Yakushima (*M. f. yakui*) in the southwestern clade, could have diverged almost simultaneously from the ancestral pool of the southwestern clade. If this is the case, it follows that the distinctive morphological characteristics of *M. f. yakui* reflect recent rapid evolution in the Yakushima lineage isolated from the ancestral pool. The cause of this rapid evolution could be attributed to adaptation to the humid and warm environment of Yakushima Island, island effects, and/or a population bottleneck as predicted by the nearly neutral theory (Ohta, 2002). The extremely low genetic diversity and relatively long branch length of *M. f. yakui* supports the latter scenario. In Osumi Strait (sill depth ca. 100m), between Yakushima Island and Kyushu, a land-bridge was formed at the peaks of glacial periods (NOAA National Geophysical Data Center, 2009; Rohling et al., 2014). Geographical isolation (although evidence of gene flow was detected) could have contributed to the uniqueness of *M. f. yakui*. Furthermore, Hayaishi & Kawamoto (2006) suggested that *M. f. yakui* experienced a bottleneck after an environmental crash associated with the explosive eruption of the Kikai Caldera ca. 7,300 years ago. It should be noted, however, that the population-level trees assuming migration events and one replicate of demographic inference suggest that the Yakushima lineage branched prior to the northeast–southwest divergence. Thus, the results generally favour the scenario that *M. f. yakui* belongs within the southwestern clade of *M. f. fuscata* over the interpretation that *M. f. yakui* is the sister taxon to *M. f. fuscata*, although this construction should be validated further.

4.4. Limitations, future directions, and concluding remarks

Our study reconstructed the phylogeographic history of *M. fuscata*, the most northerly-distributed nonhuman primate, endemic to Japan. This is, to our knowledge, one of the first genome-scale studies relating to the phylogeography of mammals on the Japanese archipelago. In summary, we found the following: (i) *Macaca fuscata* is the sister taxon to the clade comprising Chinese *M. mulatta*, Indian *M. mulatta*, and *M. cyclopis* (i.e., *M. mulatta* is monophyletic). (ii) *Macaca fuscata*

consists of two major clades located in northeastern and southwestern Japan, with a boundary in the Chubu region; a more easterly position than that detected by mtDNA haplotypes (Kawamoto et al., 2007). This divergence may be the result of isolation within refugia during the Pleistocene, followed by gene flow between populations. (iii) *Macaca f. yakui* likely belongs within the southwestern clade (i.e., *M. f. fuscata* is paraphyletic). These findings will assist the interpretation of interpopulation phenotypic variations within *M. fuscata* and enhance our understanding of the biogeographic diversification of Japan's fauna. Our study, however, leaves room for further validation of the monophyletic status of the southwestern clade and the phylogenetic position of the Yakushima subspecies. Our estimation of intraspecific divergence time relies on some uncertain assumptions (e.g., generation time and mutation rate), which should be further validated using multiple fossil calibration points. Future studies will analyse whole genome sequences to resolve these issues, and provide a more detailed picture of the demographic history of the snow monkeys, including admixtures, population size changes, and divergence dates.

Acknowledgments

We express our gratitude to Hideo Higashi, Hidetoshi Tamate, Toshihito Takagi, Miho Hakukawa, and Mayumi Morimoto, veterinarians and caretakers at the JMC; the local governments in Shimokita (Mutsu city, Oma town, Sai village, and Kazamaura village); the NBRP "Japanese Monkeys" through the National BioResource Project of MEXT, Japan, for their help with sample preservation and/or management; and Ryosuke Kimura for his valuable comments on phylogenetic analyses. We thank the Ministry of the Environment, the Yakushima Forest Ecosystem Conservation Center, the Kagoshima Prefectural Government, and the Cooperation Research Program of the Wildlife Research Center, Kyoto University for permitting us to conduct research in Yakushima; the McDonnell Genome Institute, Washington University for making available the data for the macaque genome assembly; and the Biodiversity Center, Nature Conservation Bureau, Ministry of the Environment for permitting our use of the distribution data of Japanese macaques. We also thank Michael N Dawson, Judith Masters, and anonymous reviewers for the helpful comments that greatly improved the manuscript. Our work was funded by JSPS Grants-in-Aid for Scientific Research (grants 17K15195 and 19K16211 to T. I., 16J01208 and 18K14490 to Y. Ku., 19K06865 to Y. H., and 16K18630 and 19K16241 to T. Ha.), a Sasakawa Scientific Research Grant from the Japan Science Society, and the JSPS Core-to-Core Program, and the Keihanshin Consortium for Fostering the Next Generation of Global Leaders in Research (K-CONNEX) established by the Human Resource Development Program for Science and Technology, MEXT.

Conflicts of interests

The authors have no conflict of interests to declare.

Data availability statement

The sequence reads are available from the SRA (PRJNA646022) of NCBI. The data supporting the findings of this study are available at the Dryad public archive (Ito et al., 2021; https://doi.org/10.5061/dryad.mcvdnck0b), and the Zenodo public archive (http://doi.org/10.5281/zenodo.4659562).

References

Aimi, M. (2002). The oldest fossil macaque from Japan. *Primate Research*, *18*, 239–245 [Japanese with English abstract].

Alexander, D. H., & Novembre, J. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.vidual

Bernhardt, N., Brassac, J., Dong, X., Willing, E. M., Poskar, C. H., Kilian, B., & Blattner, F. R. (2020). Genome-wide sequence information reveals recurrent hybridization among diploid wheat wild relatives. *Plant Journal*, *102*(3), 493–506. https://doi.org/10.1111/tpj.14641

Buck, L. T., De Groote, I., Hamada, Y., Hassett, B. R., Ito, T., & Stock, J. T. (2019). Evidence of different climatic adaptation strategies in humans and non-human primates. *Scientific Reports*, *9*(1), 11025. https://doi.org/10.1038/s41598-019-47202-8

Buck, L. T., De Groote, I., Hamada, Y., & Stock, J. T. (2018). Humans preserve non-human primate pattern of climatic adaptation. *Quaternary Science Reviews*, *192*, 149–166. https://doi.org/10.1016/j.quascirev.2018.05.032

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Chala, D., Roos, C., Svenning, J. C., & Zinner, D. (2019). Species-specific effects of climate change on the distribution of suitable baboon habitats – Ecological niche modeling of current and Last Glacial Maximum conditions. *Journal of Human Evolution*, *132*, 215–226. https://doi.org/10.1016/j.jhevol.2019.05.003

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*(23), 3317–3324. https://doi.org/10.1093/bioinformatics/btu530

Chu, J. H., Lin, Y. S., & Wu, H. Y. (2007). Evolution and dispersal of three closely related macaque species, *Macaca mulatta*, *M. cyclopis*, and *M. fuscata*, in the eastern Asia. *Molecular Phylogenetics and Evolution*, *43*(2), 418–429. https://doi.org/10.1016/j.ympev.2006.11.022

Craig, A. G., Grau, G. E., Janse, C., Kazura, J. W., Milner, D., Barnwell, J. W., … Langhorne, J. (2012). The role of animal models for research on severe malaria. *PLoS Pathogens*, *8*(2). https://doi.org/10.1371/journal.ppat.1002401

Dobson, M., & Kawamura, Y. (1998). Origin of the Japanese land mammal fauna: Allocation of extant species to historically-based categories. *The Quaternary Research (Daiyonki-Kenkyu)*, *37*(5), 385–395. https://doi.org/10.4116/jaqua.37.385

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252. https://doi.org/10.1093/molbev/msr048

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905. https://doi.org/10.1371/journal.pgen.1003905

Fooden, J., & Aimi, M. (2005). Systematic review of Japanese macaques, *Macaca fuscata* (Gray, 1870). *Fieldiana Zoology*, *104*(14), 1–200.

Franchini, P., Monné Parera, D., Kautt, A. F., & Meyer, A. (2017). quaddRAD: A new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. *Molecular Ecology*, 2783–2795. https://doi.org/10.1111/mec.14077

Hamada, Y., Watanabe, T., & Iwamoto, M. (1992). Variation of body color within macaques, especially in the Japanese macaques. *Primate Research*, *8*, 1-23 [Japanese with English abstract].

Hamada, Y., Watanabe, T., & Iwamoto, M. (1996). Morphological variations among local populations of Japanese macaque (*Macaca fuscata*). In T Shotake & K. Wada (Eds.), *Variations in the Asian macaques* (pp. 97–115). Tokyo: Tokyo University Press.

Hashimoto-Gotoh, A., Yoshikawa, R., Nakagawa, S., Okamoto, M., & Miyazawa, T. (2020). Phylogenetic analyses reveal that simian foamy virus isolated from Japanese Yakushima

macaques (*Macaca fuscata yakui*) is distinct from most of Japanese Hondo macaques (*Macaca fuscata fuscata*). *Gene*, 144382. https://doi.org/10.1016/j.gene.2020.144382

Hayaishi, S., & Kawamoto, Y. (2006). Low genetic diversity and biased distribution of mitochondrial DNA haplotypes in the Japanese macaque (*Macaca fuscata yakui*) on Yakushima Island. *Primates*, *47*(2), 158–164. https://doi.org/10.1007/s10329-005-0169-1

Hayakawa, T., Sawada, A., Tanabe, A. S., Fukuda, S., Kishida, T., Kurihara, Y., … Agata, K. (2018). Improving the standards for gut microbiome analysis of fecal samples: insights from the field biology of Japanese macaques on Yakushima Island. *Primates*, *59*(5), 423–436. https://doi.org/10.1007/s10329-018-0671-x

Hayasaka, K., Fujii, K., & Horai, S. (1996). Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. *Molecular Biology and Evolution*, *13*(7), 1044–1053.

Hayasaka, K., Kawamoto, Y., Shotake, T., & Nozawa, K. (1987). Population genetical study of Japanese macaques, *Macaca fuscata*, in the Shimokita A1 troop, with special reference to genetic variability and relationships to Japanese macaques in other troops. *Primates*, *28*(4), 507–516.

Hernandez, R. D., Hubisz, M. J., Wheeler, D. A., Smith, D. G., Ferguson, B., Rogers, J., … Bustamante, C. D. (2007). Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. *Science (New York, N.Y.)*, *316*(5822), 240–243. https://doi.org/10.1126/science.1140462

Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. *Nature*, *405*(6789), 907–913. https://doi.org/10.1038/35016000

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hijmans, R. J., Phillips, S., Leathwick, J., & Edith, J. (2017). *dismo: Species Distribution Modeling (R package version 1.1-4)*. Retrieved from https://cran.r-project.org/package=dismo

Huson, D. H. (1998). SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics*, *14*(1), 68–73. https://doi.org/10.1093/bioinformatics/14.1.68

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. https://doi.org/10.1093/molbev/msj030

Ikeda, J., & Watanabe, T. (1966). Morphological studies of *Macaca fuscata*: III. Craniometry. *Primates*, *7*(2), 271–288. https://doi.org/10.1007/BF01730793

Inoue-Murayama, M., Inoue, E., Watanabe, K., Takenaka, A., & Murayama, Y. (2010). Behavior-related candidate genes in Japanese macaques. In N. Nakagawa, M. Nakamichi, & H. Sugiura (Eds.), *The Japanese macaques* (pp. 293–301). https://doi.org/10.1007/978-4-431-53886-8_13

Isa, T., Yamane, I., Hamai, M., & Inagaki, H. (2009). Japanese macaques as laboratory animals. *Experimental Animals*, *58*(5), 451–457. https://doi.org/10.1538/expanim.58.451

Ito, T., Hayakawa, T., Suzuki-Hashido, N., Hamada, Y., Kurihara, Y., Hanya, G., … Kawamoto, Y. (2021). Data from: Phylogeographic history of Japanese macaques. *Dryad, Dataset*. https://doi.org/10.5061/dryad.mcvdnck0b

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kawamoto, Y. (2010). Modes of differentiation in Japanese macaques: perspectives from population genetics. In N. Nakagawa, M. Nakamichi, & H. Sugiura (Eds.), *The Japanese Macaques* (pp. 53–76). https://doi.org/10.1007/978-4-431-53886-8_3

Kawamoto, Y., Shotake, T., Nozawa, K., Kawamoto, S., Tomari, K. ichiro, Kawai, S., … Hayaishi, S. (2007). Postglacial population expansion of Japanese macaques (*Macaca fuscata*) inferred from mitochondrial DNA phylogeography. *Primates*, *48*(1), 27–40. https://doi.org/10.1007/s10329-006-0013-2

Kawamoto, Y., Tomari, K., Kawai, S., & Kawamoto, S. (2008). Genetics of the Shimokita macaque population suggest an ancient bottleneck. *Primates.*, *49*, 32–40. https://doi.org/10.1007/s10329-007-0057-y

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btz305

Kuroda, N. (1940). *A Monograph of the Japanese Mammals Exclusive of Sirenia and Cetacea*. Tokyo: Sanseido Co. [in Japanese].

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, *8*(1), e1002453. https://doi.org/10.1371/journal.pgen.1002453

Lee, W., Hayakawa, T., Kiyono, M., Yamabata, N., & Hanya, G. (2019). Gut microbiota composition of Japanese macaques associates with extent of human encroachment. *American Journal of Primatology*, *81*(12), e23072. https://doi.org/10.1002/ajp.23072

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, *40*(4), 778–789. https://doi.org/10.1111/jbi.12058

Ma, J., Prince, A. L., Bader, D., Hu, M., Ganu, R., Baquero, K., … Aagaard, K. M. (2014). High-fat maternal diet during pregnancy persistently alters the offspring microbiome in a primate model. *Nature Communications*, *5*(3889). https://doi.org/10.1038/ncomms4889

Malinsky, M. (2019). Dsuite – fast D-statistics and related admixture evidence from VCF files. *BioRxive*. https://doi.org/http://dx.doi.org/10.1101/634477

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution*, *35*(5), 1284–1290. https://doi.org/10.1093/molbev/msy023

Marmi, J., Bertranpetit, J., Terradas, J., Takenaka, O., & Domingo-Roura, X. (2004). Radiation and phylogeography in the Japanese macaque, *Macaca fuscata*. *Molecular Phylogenetics and Evolution*, *30*(3), 676–685. https://doi.org/10.1016/S1055-7903(03)00247-1

Matsuzawa, T., & McGrew, W. C. (2008). Kinji Imanishi and 60 years of Japanese primatology. *Current Biology*, *18*(14), 587–591. https://doi.org/10.1016/j.cub.2008.05.040

Melnick, D. J., Hoelzer, G. a, Absher, R., & Ashley, M. V. (1993). mtDNA diversity in rhesus monkeys reveals overestimates of divergence time and paraphyly with neighboring species. *Molecular Biology and Evolution*, *10*(2), 282–295. https://doi.org/10.1093/oxfordjournals.molbev.a040004

Morales, J. C., & Melnick, D. J. (1998). Phylogenetic relationships of the macaques (Cercopithecidae: *Macaca*), as revealed by high resolution restriction site mapping of

mitochondrial ribosomal genes. *Journal of Human Evolution*, *34*(1989), 1–23. https://doi.org/10.1006/jhev.1997.0171

Motokawa, M., & Kajihara, H. (2017). *Species diversity of animals in Japan*. https://doi.org/10.1007/978-4-431-56432-4

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, *5*(11), 1198–1205. https://doi.org/10.1111/2041-210x.12261

Nakamura, M. (2009). Interaction studies in Japanese primatology: Their scope, uniqueness, and the future. *Primates*, *50*(2), 142–152. https://doi.org/10.1007/s10329-009-0133-6

NOAA National Geophysical Data Center. (2009). ETOPO1 1 Arc-Minute Global Relief Model. *NOAA National Centers for Environmental Information*, Accessed 7th December 2020.

Nozawa, K., Shotake, T., Minezawa, M., Kawamoto, Y., Hayasaka, K., & Kawamoto, S. (1996). Population-genetic studies of the Japanese macaque, *Macaca fuscata*. In Takayoshi Shotake & K. Wada (Eds.), *Variations in the Asian macaques* (pp. 1–36). Tokyo: Tokyo University Press.

Nozawa, K., Shotake, T., Minezawa, M., Kawamoto, Y., Hayasaka, K., Kawamoto, S., & Ito, S. ichi. (1991). Population genetics of Japanese monkeys: III. Ancestry and differentiation of local populations. *Primates*, *32*(4), 411–435. https://doi.org/10.1007/BF02381934

Ohdachi, S. D., Ishibashi, Y., Iwasa, M. A., Fukui, D., & Saitoh, T. (2015). *Wild Mammals of Japan*. Kyoto: Shokado.

Ohta, T. (2002). Near-neutrality in evolution of genes and gene regulation. *Proceedings of the National Academy of Sciences*, *99*(25), 16134–16137. https://doi.org/10.1073/pnas.252626899

Osada, N., Matsudaira, K., Hamada, Y., & Malaivijitnond, S. (2020). Testing sex-biased admixture origin of macaque species using autosomal and X-chromosomal genomic sequences. *Genome Biology and Evolution*, *evaa209*. https://doi.org/10.1093/gbe/evaa209

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

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*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modeling*, *190*, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Developmental Core Team. (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna.

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, *12*(12), 2640–2659. https://doi.org/10.1038/nprot.2017.123

Rogers, J., Raveendran, M., Harris, R. A., Mailund, T., Leppälä, K., Athanasiadis, G., … Walter, L. (2019). The comparative genomics and complex population history of *Papio* baboons. *Science Advances*, *5*(1), eaau6947. https://doi.org/10.1126/sciadv.aau6947

Rohling, E. J., Foster, G. L., Grant, K. M., Marino, G., Roberts, A. P., Tamisiea, M. E., & Williams, F. (2014). Sea-level and deep-sea-temperature variability over the past 5.3 million years. *Nature*, *508*(7497), 477–482. https://doi.org/10.1038/nature13230

Roos, C., Kothe, M., Alba, D. M., Delson, E., & Zinner, D. (2019). The radiation of macaques out of Africa: Evidence from mitogenome divergence times and the fossil record. *Journal of Human Evolution*, *133*, 114–132. https://doi.org/10.1016/j.jhevol.2019.05.017

626 Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing
627 phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.
628 https://doi.org/citeulike-article-id:93683
629 Suzuki-Hashido, N., Hayakawa, T., Matsui, A., Go, Y., Ishimaru, Y., Misaka, T., … Imai, H.
630 (2015). Rapid expansion of phenylthiocarbamide non-tasters among Japanese macaques.
631 *PLoS ONE*, *10*(7), 1–21. https://doi.org/10.1371/journal.pone.0132016
632 Tamate, H. (2013). Genetic diversity in the mammalian fauna of Japan: past, present and
633 huture. *Chikyuu Kankyou*, *18*(2), 159-167 [in Japanese].
634 Tosi, A. J., Morales, J. C., & Melnick, D. J. (2000). Comparison of Y chromosome and
635 mtDNA phylogenies leads to unique inferences of macaque evolutionary history.
636 *Molecular Phylogenetics and Evolution*, *17*(2), 133–144.
637 https://doi.org/10.1006/mpev.2000.0834
638 Tosi, A. J., Morales, J. C., & Melnick, D. J. (2002). Y-chromosome and mitochondrial
639 markers in *Macaca fascicularis* indicate introgression with Indochinese *M. mulatta* and a
640 biogeographic barrier in the Isthmus of Kra. *International Journal of Primatology*, *23*(1),
641 161–178. https://doi.org/10.1023/A:1013258109954
642 Tosi, A. J., Morales, J. C., & Melnick, D. J. (2003). Paternal, maternal, and biparental
643 molecular markers provide unique windows onto the evolutionary history of macaque
644 monkeys. *Evolution*, *57*(6), 1419–1435. https://doi.org/10.1111/j.0014-
645 3820.2003.tb00349.x
646 Xue, C., Raveendran, M., Alan Harris, R., Fawcett, G. L., Liu, X., White, S., … Rogers, J.
647 (2016). The population genomics of rhesus macaques (*Macaca mulatta*) based on whole-
648 genome sequences. *Genome Research*, *26*(12), 1–12.
649 https://doi.org/10.1101/gr.204255.116.Freely
650 Yamagiwa, J. (2010). Research history of Japanese macaques in Japan. In *The Japanese*
651 *macaques* (pp. 3–25). https://doi.org/10.1007/978-4-431-53886-8_1
652 Yano, W., Egi, N., Takano, T., & Ogihara, N. (2020). Subspecies and sexual craniofacial size
653 and shape variations in japanese macaques (*Macaca fuscata*). *Anthropological Science*,
654 *128*(2), 71–81. https://doi.org/10.1537/ase.2005052
655 Zachos, F. E., & Habel, J. C. (2011). *Biodiversity Hotspots: Distribution and Protection of*
656 *Conservation Priority Areas*. New York: Springer.
657

Biosketch

Tsuyoshi Ito is an assistant professor at the Department of Evolution and Phylogeny, Primate Research Institute, Kyoto University. He is interested in understanding phylogeny, biogeography, and the mechanisms and processes underlying morphological diversity.

Author contributions: T.I. and Y.Ka. conceived the study. T.I., Y.Ku., G.H., A.K., T.N., S.A., and T.Ho. conducted field sample collection at Yakushima Island. T.Ha., N.S.–H., Y.H., S.Y., T.A., T.O., S.H., H.I., and Y.Ka. prepared samples. T.I., T.Ha. N.S.–H., M.T., H.W., and Y.Ka. performed basic molecular experiments. T.I. prepared the ddRAD library, analysed the data, and drafted the manuscript with contributions by the other authors, and all authors approved its final version.

Editor: Judith Masters

Significant statement

Japanese macaques, also known as snow monkeys, have been extensively studied, particularly in the fields of ecology, sociobiology and experimental biology. However, our knowledge of their evolutionary history is limited. Here, we analysed genomic data of 113 samples from ten localities to reveal Japanese macaques comprise major northeastern and southwestern clades, with a boundary located in Central Japan. The northeast–southwest divergence likely resulted from withdrawal into refugia along the southern coasts of the Japanese archipelago during the glacial periods. Gene flow took place between populations during interglacial periods, causing mosaic genetic structure.

Figure legends

Figure 1. The distribution (left) and sampling sites (right) of Japanese macaques (*Macaca fuscata*) in the Japanese archipelago. The distribution map was created by the authors using the national standard area meshes (5 km × 5 km) and the csv data of the second (FY1978, sky blue) and sixth (FY2003, orange) distribution surveys of Japanese animals (mammals) by the Biodiversity Center of Japan, Nature Conservation Bureau, Ministry of the Environment (http://gis.biodic.go.jp/webgis/). Map projection is Mollweide. The regions and main islands are represented by bold lines (left). The major northeastern (NE) and southwestern (SW) clusters detected in the individual-level autosomal trees are represented by circles (right).

Figure 2. Clustered fineRADstructure coancestry matrix of Japanese macaque (*Macaca fuscata*)

687 populations based on autosomal loci. Heatmap indicates pairwise coancestry between individuals
688 with colour-range from low (yellow) to high (black). Dendrogram shows clustering based on the
689 pairwise matrix. S: Shiga, T: Takahama (lib2-3-E [denoted by asterisk] was however labelled as
690 Wakasa), W: Wakasa, A: Arashiyama.
691 Figure 3. Admixture plot of Japanese macaque (*Macaca fuscata*) populations when $K = 7$ based on the
692 autosomal SNVs (upper) and the cross-variation error (lower right). Principal component scores
693 based on the autosomal SNVs are reported lower left.
694 Figure 4. Neighbour-net phylogenetic networks (upper) and a neighbour-joining tree (lower) of Japanese
695 macaque (*Macaca fuscata*) populations and closely related species based on uncorrected *p*-
696 distances calculated from the autosomal SNVs. Node labels indicate bootstrap support values,
697 shown only on the nodes of major clades. The tree was rooted by long-tailed macaques (*M.*
698 *fascicularis*). Northeast (NE) and southwest (SW) clades are denoted.
699 Figure 5. Heatmap summarising the *D*-statistics determined by ABBA-BABA tests for Japanese
700 macaque (*Macaca fuscata*) populations, wherein the maximum *D*-statistic values for the pair of P2
701 and P3 across all possible P1 are represented (after Bernhardt et al., 2020). Trios are ordered
702 consistently with the BBAA pattern (lower diagonal) or topology: ((Shimokita, Yamagata, Gunma),
703 (Kinki, Kochi, Kojima, Yakushima)) (upper diagonal). See Tables S8–S10 for details.
704 Figure 6. The present spatial distribution of Japanese macaque (*Macaca fuscata*) populations (left) and
705 the Last Glacial Maximum (LGM) (right) in East Asia, predicted by MAXENT. Map projection is
706 Mollweide. Predicted suitability is binarised by the threshold that maximises the sum of sensitivity
707 and specificity (max SSS; Liu, White, & Newell, 2013), as calculated using the 'dismo' R package
708 (Hijmans, Phillips, Leathwick, & Edith, 2017). Colour indicates binarised predicted suitability:
709 dark (unsuitable) and light (suitable). For the LGM, the three binary maps based on the CCSM4,
710 MIROC-ESM, and MPI-ESM-P models are assembled (after Chala, Roos, Svenning, & Zinner,
711 2019); suitability is represented in four grades from dark (unsuitable in all the three maps) to light
712 (suitable in all the three maps). For un-binarised (unassemmbled) maps, see Fig. S13.

713 **Tables**

Table 1. Taxa, population, source (wild/captive), and the number of samples used in this study.

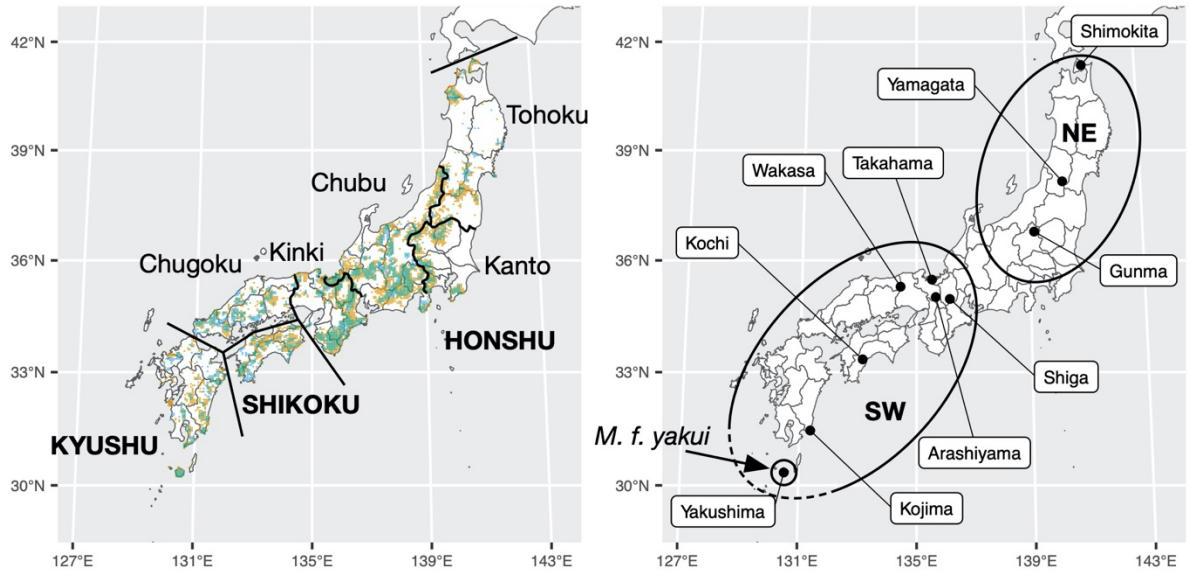
Species (subspecies)	Population	Source	*N*
Macaca fuscata fuscata	Shimokita	Wild	15
	Yamagata	Wild	5
	Gunma	Wild	14
	Shiga	Wild	10
	Takahama	PRI captive	5
	Arashiyama	PRI captive	5
	Wakasa	PRI captive	6
	Kochi	Wild	14
	Kojima	Wild	14
M. fuscata yakui	Yakushima	Wild (5)/JMC captive (7)/PRI captive (2)	14
M. mulatta	China	PRI captive	3
	India	PRI captive	3
M. cyclopis		PRI captive	2
M. fascicularis		PRI captive	3

N, the number of samples; PRI, Primate Research Institute, Kyoto University, Inuyama, Japan; JMC, Japan Monkey Centre, Inuyama, Japan. The numbers in parentheses denote the sample size for that group.

714

715

716 Figures



717

718 Fig. 1

719

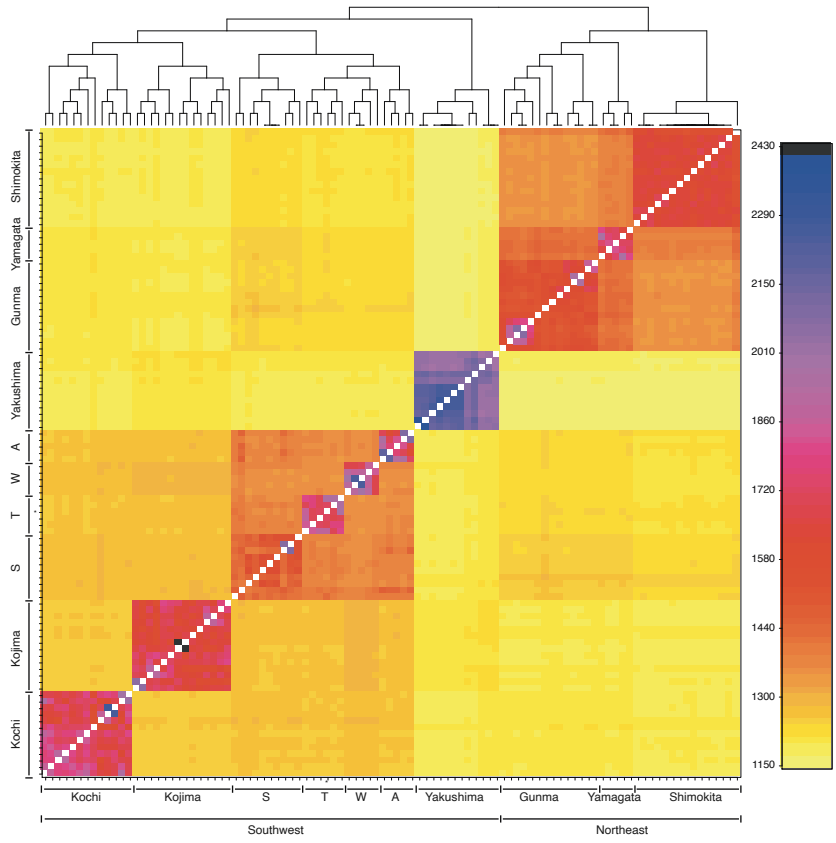

Shimokita
Yamagata
Gunma
Yakushima
A
W
T
S
Kojima
Kochi
2430
2290
2150
2010
1860
1720
1580
1440
1300
1150
Kochi
Kojima
S
T
W
A
Yakushima
Gunma
Yamagata
Shimokita
Southwest
Northeast

Fig. 2

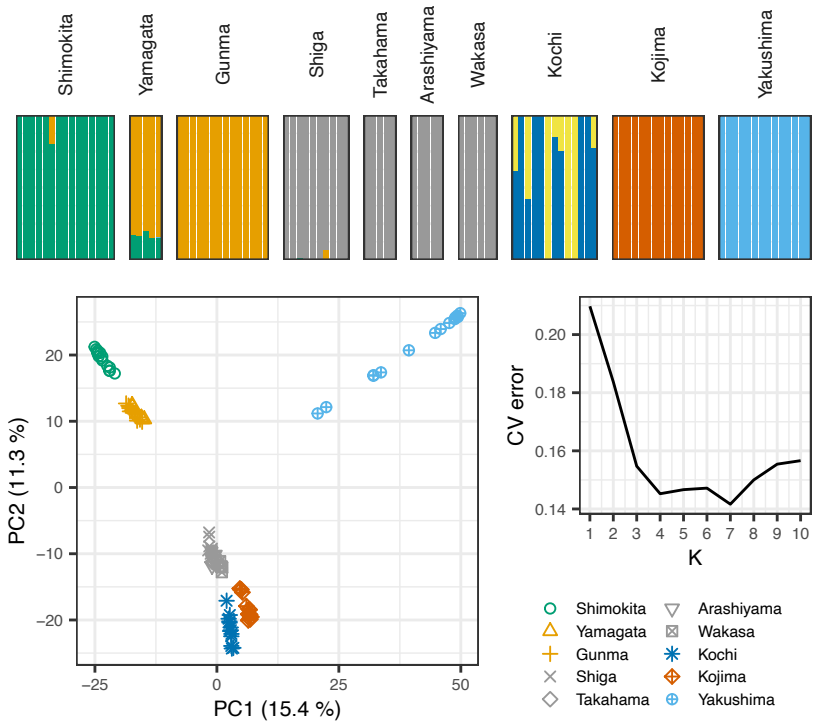
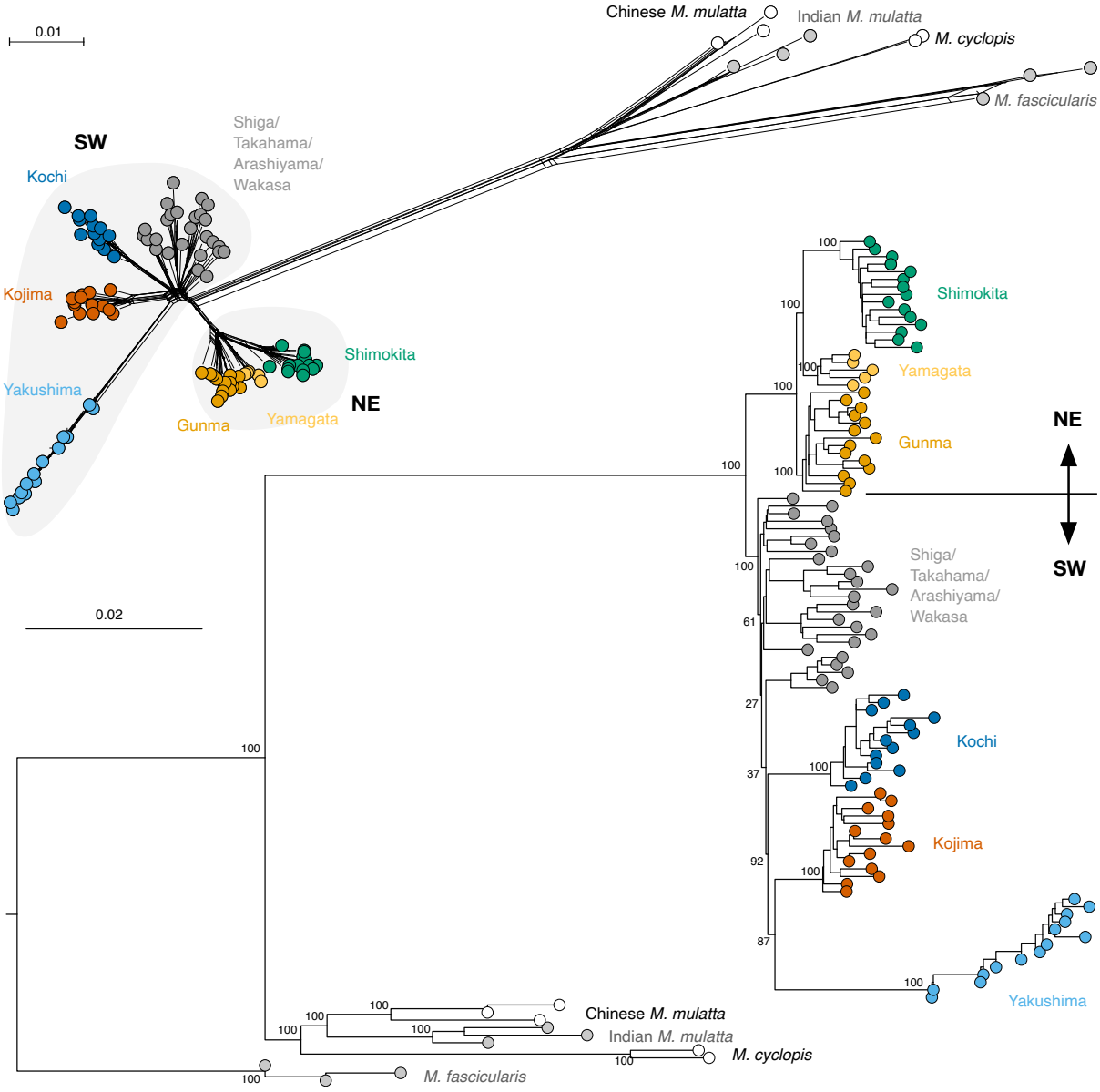

Shimokita
Yamagata
Gunma
Shiga
Takahama
Arashiyama
Wakasa
Kochi
Kojima
Yakushima
PC2 (11.3 %)
PC1 (15.4 %)
CV error
K
Shimokita
Yamagata
Gunma
Shiga
Takahama
Arashiyama
Wakasa
Kochi
Kojima
Yakushima

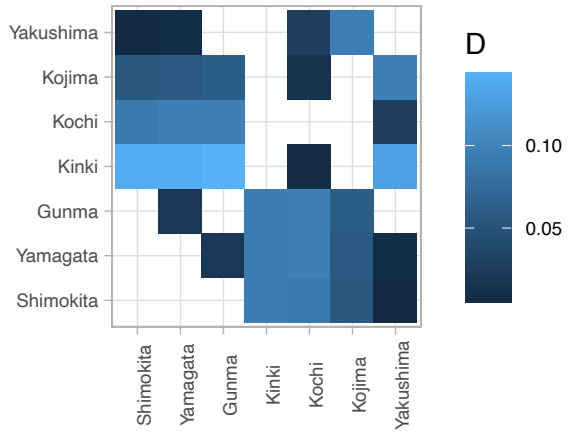
Fig. 3



728

729 Fig. 4

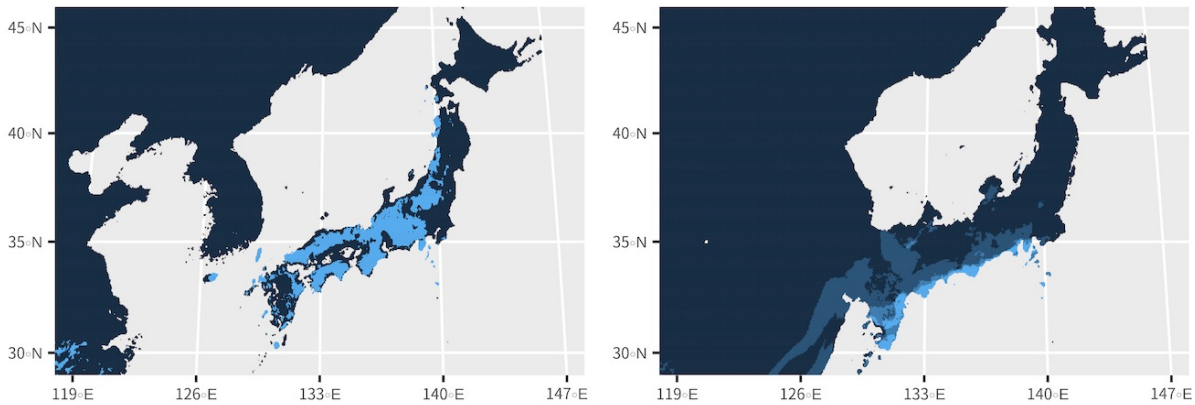
730



731

732 Fig. 5

733



734

735 Fig. 6

736