Inference of Selective Force on House Mouse Genomes during Secondary Contact in East Asia

Kazumichi Fujiwara[1,2], Shunpei Kubo[2], Toshinori Endo[2], Toyoyuki Takada[3], Toshihiko Shiroishi[4], Hitoshi Suzuki[5], Naoki Osada[2,*]

[1]Mouse Genomics Resource Laboratory, National Institute of Genetics, Mishima, Japan

[2]Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan

[3]Integrated BioResource Information Division, RIKEN BioResource Research Center, Tsukuba, Japan

[4]RIKEN BioResource Research Center, Tsukuba, Japan

[5]Graduate School of Environmental Science, Hokkaido University, Sapporo, Japan

*Author for correspondence: Naoki Osada, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan; phone: +81-11-706-7332; email: nosada@ist.hokudai.ac.jp

Running title: East Asian wild house mouse genome analysis

Abstract

The house mouse (*Mus musculus*), commensal to humans, has spread globally via human activities, leading to secondary contact between genetically divergent subspecies. This pattern of genetic admixture can provide insights into the selective forces at play in this well-studied model organism. Our analysis of 163 house mouse genomes, with a particular focus on East Asia, revealed substantial admixture between the subspecies *castaneus* and *musculus*, particularly in Japan and southern China. We revealed, despite the admixture, all Y chromosomes in the East Asian samples belonged to the *musculus*-type haplogroup, potentially explained by genomic conflict under sex ratio distortion due to varying copy numbers of ampliconic genes on sex chromosomes. We also investigated the influence of selection on the post-hybridization of the subspecies *castaneus* and *musculus* in Japan. Even though the genetic background of most Japanese samples closely resembles the subspecies *musculus*, certain genomic regions overrepresented the *castaneus*-like genetic components, particularly in immune-related genes. Furthermore, a large genomic block containing a vomeronasal/olfactory receptor gene cluster predominantly harbored *castaneus*-type haplotypes in the Japanese samples, highlighting the crucial role of olfaction-based recognition in shaping hybrid genomes.

Introduction

The house mouse (*Mus musculus*) is a pivotal laboratory animal, which is commensal to humans and has spread worldwide due to human activities such as colonization, agricultural expansion, and trade (GÜndüz et al. 2001; Geraldes et al. 2008; Searle et al. 2009a; Searle et al. 2009b; Gabriel et al. 2011; Macholán et al. 2012; Jones et al. 2013; Li et al. 2020). The house mouse exhibits high morphological and genetic diversity, with three main subspecies, recognized: *M. musculus castaneus*, *M. m. musculus*, and *M. m. domesticus* (Boursot et al. 1996; Salcedo et al. 2007; Phifer-Rixey et al. 2020; Fujiwara et al. 2022a). These subspecies diverged around 180–500 kya and underwent secondary contact after dispersing from their native habitat around South Asia (Boursot et al. 1993; Din et al. 1996) with asymmetric hybrid incompatibility; the subspecies *domesticus* shows relatively strong hybrid incompatibility with the other two subspecies (White et al. 2011; White et al. 2012). Despite partial reproductive isolation between subspecies, genetic studies have revealed a non-negligible amount of gene flow, and occasional identification of hybrid genotypes (Orth et al. 1998; Ďureje et al. 2012; Fujiwara et al. 2022a). Studying the genetic differentiation and hybridization patterns in *M. musculus* thus provides a unique opportunity to gain insights into human history and the genetic architecture of hybridization among genetically distinct subspecies of a well-studied model organism.

In East Asia, two major subspecies have been widely observed; the subspecies *castaneus* is distributed in southern China and Taiwan, and the subspecies *musculus* distributed in northern China, the Korean Peninsula, and the Russian Far East (Din et al. 1996; Jing et al. 2014; Fujiwara et al. 2022a). Another subspecies, *M. m. molossinus*, has been recognized in the Japanese archipelago, where researchers have shown through genetic analysis that it is derived from a hybrid between the subspecies *castaneus* and *musculus* (Yonekawa et al. 1980; Yonekawa et al. 1982; Yonekawa et al. 1988; Sakai et al. 2005; Takada et al. 2013), while most of wild Japanese house mice have major genetic background of the subspecies *musculus* (Fujiwara et al. 2022a).

59 Interestingly, mitochondrial haplogroups representing the subspecies *castaneus* are mainly
60 distributed in northern Japan, while mitochondrial haplogroups representing the subspecies
61 *musculus* are observed in southern Japan (Terashima et al. 2006; Kuwayama et al. 2017; Li et al.
62 2020), leading the hypothesis that the subspecies *castaneus* was introduced from southern China
63 to the Japanese archipelago and *musculus* was introduced from the Korean Peninsula with wet-
64 rice field cultivation system (Terashima et al. 2006; Kuwayama et al. 2017). This hypothesis
65 would be similar to the well-known model of the origin of modern Japanese, called the dual
66 structure model by Hanihara (1991). The dual structure model assumes that the genetic makeup
67 of modern Japanese has been composed of both indigenous Jomon hunter-gatherers and newly
68 migrated Yayoi rice farmers (Hanihara 1991). While recent genetic analyses of modern and
69 ancient human genome studies have revealed that the Jomon originated from one of the basal East
70 Asian populations (McColl et al. 2018; Kanzawa-Kiriyama et al. 2019; Gakuhari et al. 2020;
71 Osada and Kawai 2021), it is largely unknown when and from where the subspecies *castaneus*
72 arrived in the Japanese archipelago.

73 Although the genetic diversity of maternally-inherited mitochondrial haplotypes in wild house
74 mice has been well studied, natural variation in the paternally-inherited Y chromosome was
75 largely unknown except for a few examples (Morgan and Pardo-Manuel de Villena 2017). Recent
76 experimental studies have discovered X- and Y-linked loci that are crucial for hybrid
77 incompatibility, where copy number variation (CNV) plays a significant role in determining
78 hybrid incompatibility and sex-ratio distortion (SD). Previous studies have demonstrated that Y-
79 linked *Sly* suppresses post-meiotic expression of X- and Y-linked genes in sperms (Ellis et al.
80 2011). Copy number increase in *Sly* results in male-biased progenies, while copy number increase
81 in X-linked *Slx1* are known to counteract it (Scavetta and Tautz 2010; Cocquet et al.). Morgan et
82 al. (2017) showed that the subspecies *musculus* tended to have higher *Sly*/*Slx* copy numbers than
83 other subspecies, but their analyzed samples are biased toward wild-derived inbred lines [31].

84 Therefore, the analysis for CNVs of *Sly*/*Slx* and their natural distribution range in wild samples
85 would provide how the sex-ratio distortion alleles affect the genetic differentiation of mouse
86 subspecies.

87 The selective forces on the formation of hybrid genomes are also an important issue in
88 evolutionary biology research. With the advancement of genome sequencing technology, many
89 genome-scale studies have been conducted to examine the effects of natural selection under
90 interspecific or intersubspecific gene flow. In particular, adaptive gene introgression, in which
91 beneficial alleles cross the boundary of populations and spread rapidly to another population, has
92 attracted the interest of many researchers (Song et al. 2011; Pardo-Diaz et al. 2012; Huerta-
93 Sánchez et al. 2014; Enard and Petrov 2018). The distinctive genetic characteristic of wild
94 Japanese house mice would provide a great opportunity to investigate the effects of selection
95 during hybridization. Having the major genetic background of the subspecies *musculus*, the
96 preference for the genomic region of the *castaneus*-ancestry in the Japanese population leaves a
97 similar signature to that of adaptive introgression and is a robust signal of natural selection during
98 hybrid genome formation.

99 In this study, we analyze 163 high-coverage whole genome sequences of *M. musculus*,
100 including newly sequenced samples primarily from Japan, to quantify the pattern of genetic
101 admixture among the samples. By contrasting the population differentiation patterns of autosomal
102 and sex-linked genomic regions, we infer the evolutionary history of the two East Asian
103 subspecies populations. We also analyze Japanese samples to identify specific regions of the
104 genome where genetic components from the subspecies *castaneus* are more prevalent, even
105 though the majority of the Japanese samples have a genetic background from the subspecies
106 *musculus*. Focusing on the evolutionary history of the *M. musculus* in East Asia, this study
107 provides new insights into how the commensal animal has shaped its genetic traits through
108 secondary contact mediated by human activities.

Result

Genome-wide pattern of differentiation in East Asia

In this study, we sequenced the genome of 37 wild house mice mainly collected from Japan, with an average coverage of 26.3 (Supplementary Table 1). Those data were merged with a previously published dataset that included *M. spretus*, yielding the genotypes 170 samples. After filtering, 133,886,237 autosomal biallelic SNVs were used for the initial population study.

Principal component analysis (PCA) was performed on all *M. musculus* samples (Supplementary Figure 1). The results were in high agreement with those previously shown by Fujiwara et al. (2022). Three major genetic components corresponding to subspecies *castaneus*, *domesticus*, and *musculus*, were observed. Most of the East Asian samples aligned on the *castaneus-musculus* cline, which means that the genetic characteristics of the East Asian samples can be modeled by different degrees of admixture between the subspecies *castaneus* and *musculus*. Genetic clustering was performed using ADMIXTURE software assuming three ancestral populations ($K = 3$). The results yielded three genetic components corresponding to the three main subspecies, consistent with the PCA results (Supplementary Figure 2).

To quantify the degree of genetic admixture in each sample, the f_4-ratio value developed by Patterson et al. (2012) was calculated for each sample (see Methods), assuming that the East Asian samples were derived from admixture of two unknown lineages that shared ancestry with the Korean *musculus* and Indian *castaneus* populations. The estimated percentage of *musculus* ancestry, expressed as α, varied from 0.109 to 1 among the East Asian samples, as shown in Fig. 1. In the Japanese archipelago, all samples had α values greater than 0.564, and the mean value of α was 0.878 (Fig. 1). Interestingly, the samples from the Sea of Japan side have lower α values than those from the Pacific Ocean side. When the sample from Okinawa (a subtropical island) was excluded and the other samples were divided into two classes according to the Central

Watershed of Japan, the Pacific-side samples had significantly higher α than the Sea of Japan-side samples ($P = 0.0002$, MWU test). In China, α values were close to 1 in northern and western China, but much lower in southern China (Fig. 1). In the southern Chinese samples, the averaged α was 0.295.

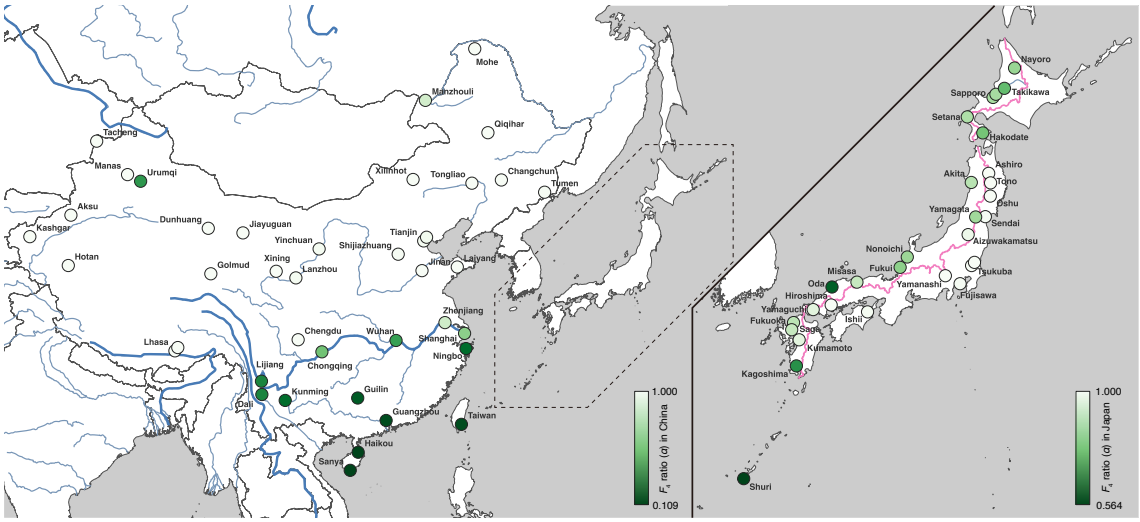


Fig. 1: The estimated ratio of *musculus/castaneus* ancestry in East Asia.

The colors of circles indicate *musculus*/*castaneus* ancestry of samples at each collection site (left pane: East Asian samples except Japanese archipelago, right pane: Japanese archipelago samples). The *musculus* ancestry and *castaneus* ancestry of each sample are represented by a range of colors from white to green, with the stronger green color indicating higher *castaneus* ancestry. The range of ancestry ratio is indicated by the color bar in each panel. The pink line in Panel B represents the Central Watershed in the Japanese archipelago.

Genetic differentiation at mitochondrial and sex-linked loci

We compared genetic differentiation patterns at autosomal, X-chromosomal, Y-chromosomal, and mitochondrial loci. We observed generally less mixing of subspecies genomes on the X chromosome; *i.e.*, there was less *castaneus*-ancestry in the Japanese population and less

150 *musculus*-ancestry in the southern Chinese population. The estimated α values on the X
151 chromosome ranged from 0.317 to 1 (average α: 0.860) for Japanese samples and from 0.034 to
152 0.952 (average α: 0.207) for southern Chinese samples, showing that the X chromosome has a
153 low level of introgression into the predominant genome. The difference in α between autosomes
154 and X chromosomes were significant in both Japanese (P = 0.0008) and southern Chinese (P =
155 0.0134) populations (Wilcoxon signed rank test).

156 Genealogies of the Y-chromosomal and mitochondrial loci inferred by the maximum likelihood
157 (ML) method are shown in Supplementary Figure 3 and 4. The Y-chromosomal locus formed
158 three distinct haplogroups corresponding to the three major subspecies, whereas the mitochondrial
159 locus showed five major haplogroups. In addition to the three mitochondrial haplogroups
160 representing the subspecies *castaneus*, *domesticus,* and *musculus*, samples from Madagascar and
161 Nepal formed distinct mitochondrial haplogroups, as previously shown (Fujiwara et al. 2022a;
162 Fujiwara et al. 2022b).

163 The geographic distribution of the subspecies *castaneus* and *musculus* haplotypes was
164 markedly different between mitochondrial and Y-chromosomal loci (Fig. 2). The distribution of
165 mitochondrial haplotypes was consistent with previous findings (Fig. 2a). In the Japanese
166 archipelago, *castaneus*- and *musculus*-type mitochondrial haplotypes are distributed in the
167 northern and southern regions of mainland Japan, respectively. In China, the distribution of
168 mitochondrial haplotypes was clearly divided around the Yangtze River basin (or divided at
169 Qinling Mountains), with *musculus*-type haplotypes in northern China and *castaneus*-type
170 haplotypes in southern China. However, all Y-chromosomal haplotypes in East Asia were
171 assigned exclusively to *musculus*-type haplotypes (Fig. 2b). The Y chromosome of the reference
172 genome (C57BL/6J) also had a *musculus*-type haplotype, and it was closely related to the
173 haplotype of Japanese samples from the Tohoku area (Supplementary Figure 4), supporting the
174 idea that the Y chromosome of classical inbred strains originated from Japanese house mice

(Japanese fancy mouse) (Bishop et al. 1985; Nagamine et al. 1992; Tucker et al. 1992; Takada et al. 2013).

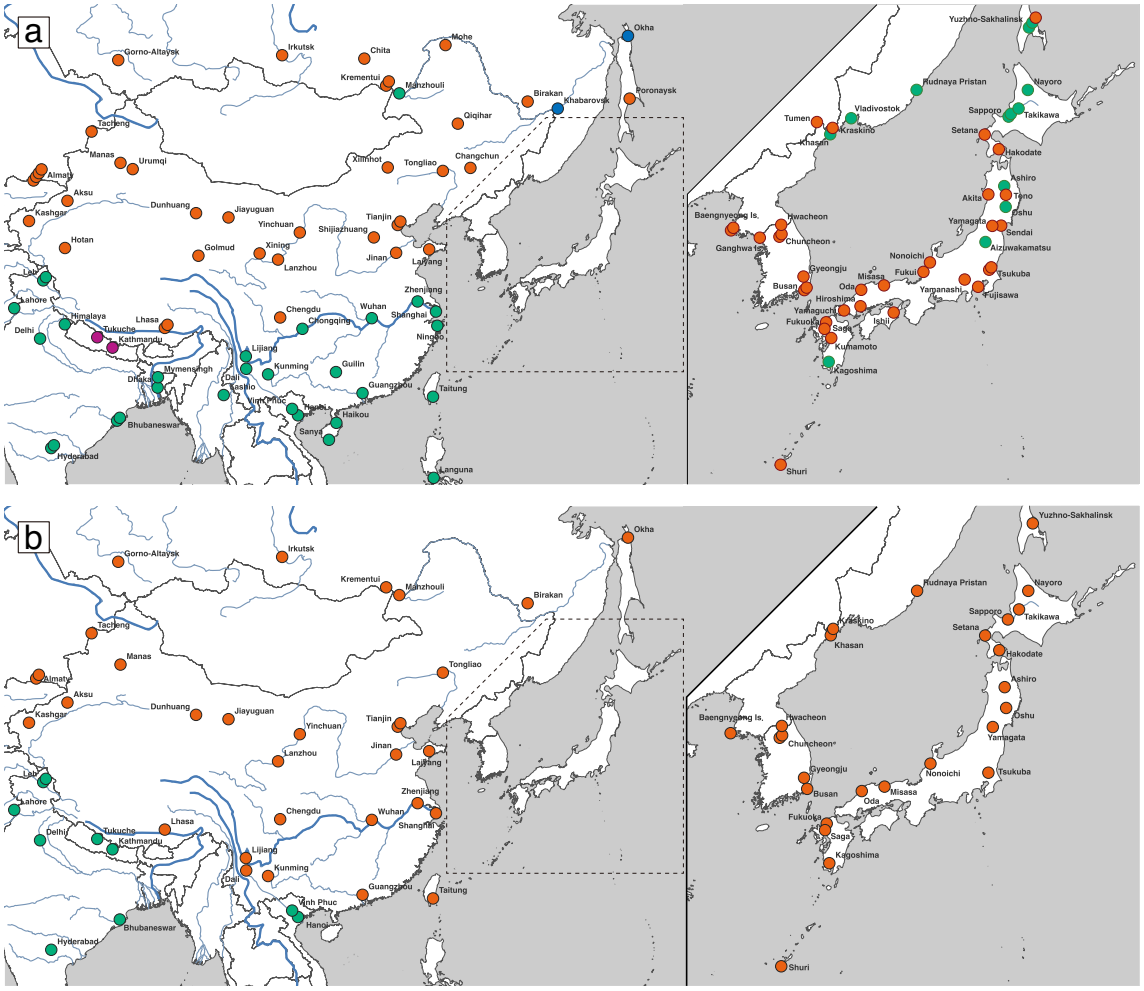
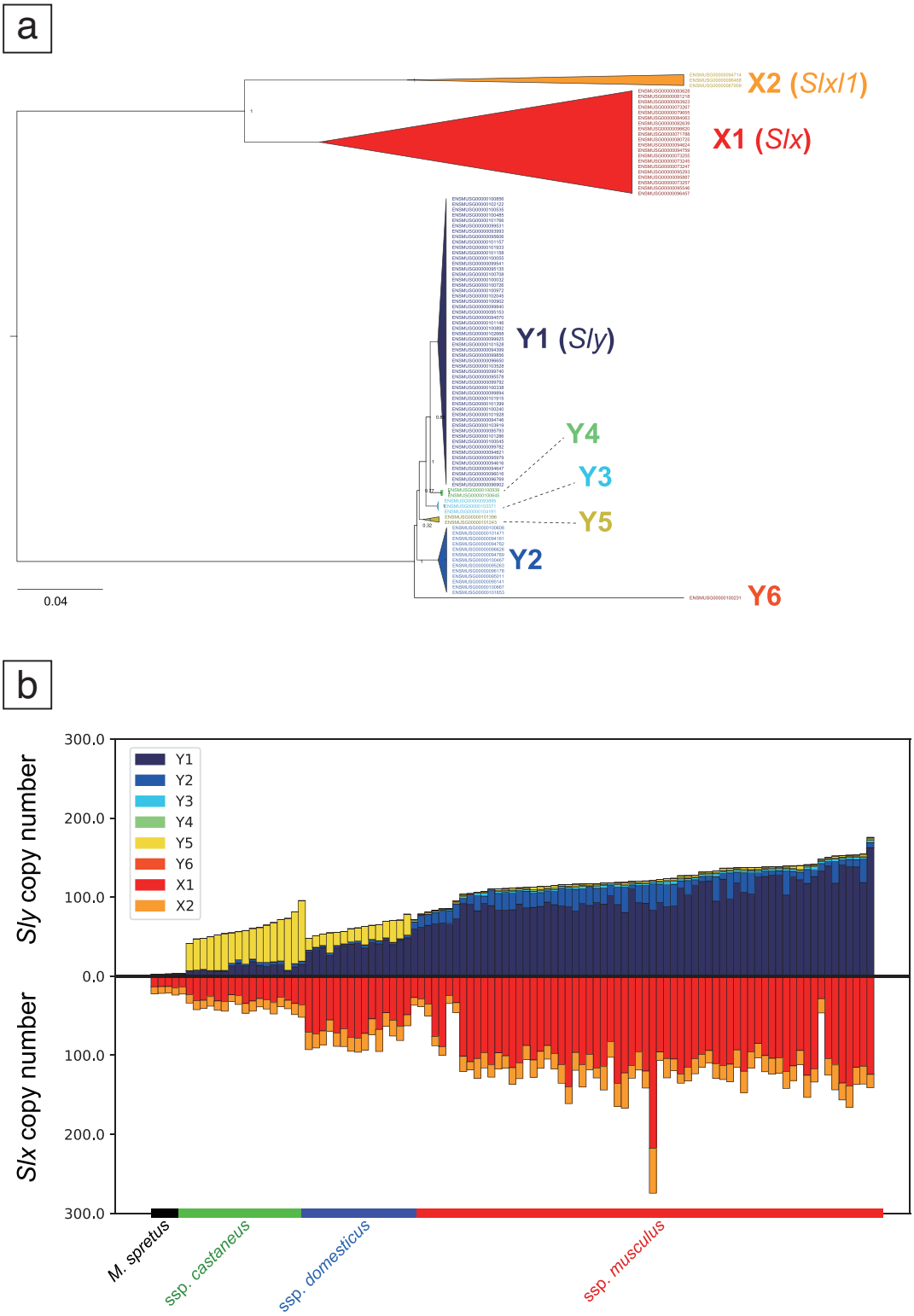


Fig. 2: The geographic distribution of the subspecies *castaneus* and *musculus* haplotypes of mitochondrial and Y-chromosomal loci.

The upper panel (a) and lower panel (b) shows the mitochondrial and Y-chromosomal haplotypes, respectively. The colors of circles indicate subspecies haplotypes based on genotyping: *castaneus*-type (green), *domesticus*-type (blue), *musculus*-type (red), and geographically confined group of Nepalese-type (purple).

185 We hypothesized that the dominance of the *musculus*-type Y chromosome in East Asia is due
186 to a distorted sex ratio caused by a dosage imbalance in the *Sly*/*Slx* genes; *Sly* and *Slx* are
187 gametologs under X-Y sexual conflict and are highly duplicated genes (>100 copies). Previous
188 studies have shown that *musculus*-type Y chromosomes encompass more *Sly* copies and the
189 subspecies *musculus* have more copies of *Slx* on the X chromosome (Morgan and Pardo-Manuel
190 de Villena 2017). Estimating the copy numbers of *Sly* and *Slx*, however, is challenging due to the
191 presence of highly similar but distinct classes of repetitive sequences in the genome. To tackle
192 this, we first aligned the paralogous sequences of the *Sly* and *Slx* loci present in the reference
193 genome and used these alignments to reconstruct a phylogenetic tree. The reconstructed tree
194 showed two clusters for *Slx*, which correspond to *Slx* (X1) and *Slxl1* (X2), and six clusters for *Sly*
195 (Y1–Y6), with one cluster (Y1) being the most abundant in the reference genome and including
196 the canonical *Sly* (Fig. 3a). We then estimated the copy numbers for each cluster based on the
197 short-read depth of cluster-specific SNP alleles (see Methods). We observed that different
198 subspecies-level Y haplotypes harbored different patterns of *Sly* clusters (Fig. 3b). The cluster Y5
199 was the major cluster in the *castaneus*-type haplotypes, while the cluster Y1 was the major cluster
200 in the *domesticus*-type, and *musculus*-type Y haplotypes, respectively. Particularly, *castaneus*-
201 and *domesticus*-type Y haplotypes had low *Sly* copy numbers, while *musculus*-type Y haplotypes
202 had high copy numbers among the three haplogroups. In contrast, the copy number ratios of *Slx*
203 and *Slxl1* were mostly consistent among samples. As shown in Fig. 3b and the previous study, *Sly*
204 and *Slx* copy numbers were highly correlated ($P < 10^{-22}$, Spearman's rank correlation test)
205 (Morgan and Pardo-Manuel de Villena 2017).



206

207 Fig. 3: Copy number evolution of *Sly* and *Slx*

208 The phylogenetic tree of *Sly*/*Slx* genes on the mouse reference genome, GRCm38 (a). Bootstrap
209 values are described at each node. Estimated copy numbers of *Sly* and *Slx* for each sample (b).
210 Different colors represent the copy numbers of different cluster. The samples were aligned
211 according to the Y chromosome haplotypes first and sorted ascendingly by the total copy
212 numbers of *Sly*.

213

214 In order to investigate selective forces shaping the sweep-like pattern of Y chromosome
215 haplotypes in southern China, we performed Wright-Fisher computer simulations mimicking the
216 introgression from the subspecies *musculus* population to subspecies *castaneus* populations in
217 southern China. The model assumes that a small fraction of the subspecies *musculus* individuals
218 with longer haplotypes of SD locus on the sex chromosomes (corresponding to *Sly* and *Slx*)
219 continuously migrate to the subspecies *castaneus* population. An F1 hybrid between a *musculus*-
220 type male and a *castaneus*-type female tends to have more male offspring than female offspring
221 owing to the segregation distortion of X/Y chromosomes in a male germline. We also modeled
222 loci responsible for partial hybrid incompatibility on the X chromosome (XHI loci). Detailed
223 methods and results are documented in the Supplementary Note 1, Supplementary Figure 5, and
224 Supplementary Table 2. Briefly, without assuming the mutations on copy numbers at the SD locus,
225 the simulations showed that the introgressed Y chromosomes with longer SD haplotypes
226 (corresponding to the *musculus*-type *Sly*/*Slx* haplotypes) quickly spread in the recipient
227 population. The X chromosomes with longer SD haplotypes also eventually fix in the recipient
228 population but the fixation of longer haplotypes was in general faster on the Y chromosomes than
229 on the X chromosomes, because we assumed that SD occurs on male germline. The difference of
230 fixation time between the Y and X chromosomes becomes longer when the effect of XHI is strong
231 and the recombination between XHI loci and SD locus on the X chromosome is rare. When the
232 effect of XHI is sufficiently strong, the introgression of X chromosomes with longer SD

233 haplotypes are completely suppressed.

234 *Selection during the admixture in the Japanese population*

235 To identify genomic regions affected by natural/sexual selection during intersubspecific
236 admixture, we focused on samples from the Japanese archipelago and searched for genomic
237 regions highly enriched in *castaneus*-ancestry haplotypes. Following the results presented in Fig.
238 1, we estimated the values of α, representing the proportion of *musculus*-like genomic components,
239 for each non-overlapping 20-kb-long window of the autosomal genome.

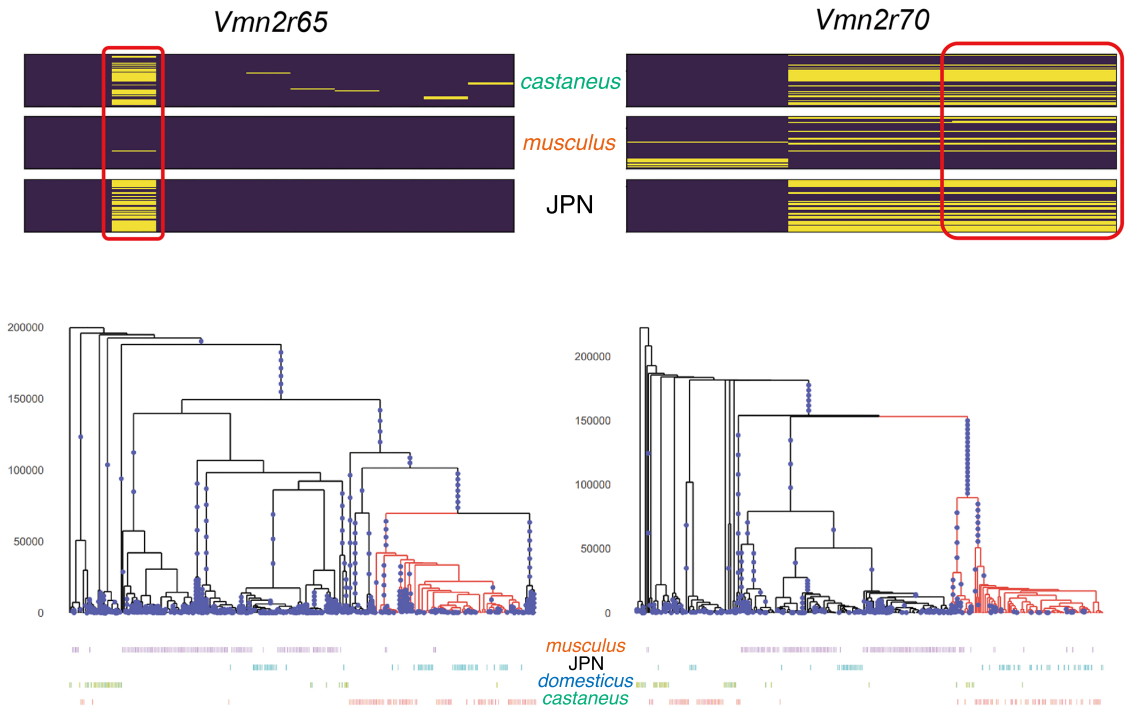
240 To discern the neutral distribution of *castaneus*-enriched genomic blocks within the Japanese
241 house mouse population, we employed the Fastsimcoal2 software. This aided in inferring
242 demographic parameters from our sample set, which were used in estimating α (See
243 Supplementary Note 2, Supplementary Figure 6, and Supplementary Table 3 for details). Using
244 these parameters, we executed 100,000 coalescent simulations to ascertain a distribution of α
245 under neutrality. From our results, we identified 4194 windows. These windows, which fell below
246 the 5% threshold of the neutral distribution, encompassed the protein-coding regions of 842
247 distinct genes. A comprehensive list of these genes can be found in Supplementary Table 4.

248 We performed a functional enrichment analysis on the 842 candidate genes using MetaScape
249 software (Zhou et al. 2019). A total of 20 gene categories were found to be enriched in the list (p
250 values < 0.01), with the complete list of significantly overrepresented functional categories
251 presented in Supplementary Figure 7. The enriched term list contained a variety of gene categories,
252 but notably, immune-related gene categories such as immunoglobulin production (GO:002377),
253 antibody-dependent cellular cytotoxicity (GO:0001788), and Herpes simplex virus 1 infection
254 (mmu05168) were overrepresented. Indeed, our screening accurately identified the retroelement-
255 like *Fv1* (Friend virus susceptibility protein 1) gene, an antiviral factor (Best et al. 1996) as an
256 *castaneus*-ancestry enriched gene in Japan. Previous studies have demonstrated that the $Fv1^b$

257 allele, which lacks the 1.3-kb-length deletion commonly found in the subspecies *domesticus* and
258 *musculus*, is observed with high frequency in the subspecies *castaneus* and the Japanese samples
259 (Boso et al. 2021). We confirmed that the deletion started from chr4:147,870,288 (GRCm38) and
260 had 1.2 kb in length, resulting in the truncation of C-terminal region of *Fv1*. The allele frequencies
261 of deletion were 0.92, 0.64, and 0.16 in the subspecies *domesticus*, *musculus*, and *castaneus*,
262 respectively. The allele frequency in the Japanese samples was 0.02, showing that almost no
263 Japanese house mice harbor the deletion.

264 We also observed that 23 vomeronasal receptor genes (*Vmn1r* and *Vmn2r*), including one
265 pseudogene, were present in the gene list. In the enrichment analysis, a category, detection of
266 chemical stimulus involved in sensory perception (GO0050907) showed the strongest bias. These
267 *castaneus*-enriched vomeronasal receptor genes were scattered across different
268 olfactory/vomeronasal genomic clusters on the mouse genome, with most genes located in a large
269 cluster on chromosome 7: region chr7:84,853,553-87,037,968, containing 17 olfactory and 15
270 vomeronasal receptors. Of the 23 vomeronasal receptor genes identified in our genomic scan, 12
271 were located in the cluster. To confirm that the estimated low α in the region was not due to errors
272 in statistical inference, we computed F_{ST} between Japanese and subspecies *musculus* samples
273 ($F_{ST\text{-}MUS/JPN}$) and F_{ST} between Japanese samples and subspecies *castaneus* samples ($F_{ST\text{-}CAS/JPN}$).
274 The subspecies *castaneus* and *musculus* samples were assigned from the result of the
275 ADMIXTURE analysis. We plotted window-averaged $F_{ST\text{-}CAS/JPN}$ and $F_{ST\text{-}MUS/JPN}$ across the region
276 including chr7:84,853,553-87,037,968 (Fig. 4). As expected, outside of this region, $F_{ST\text{-}MUS/JPN}$
277 values were consistently smaller than $F_{ST\text{-}CAS/JPN}$, supporting the notion that Japanese samples
278 were generally *musculus*-like. However, the pattern was completely reversed within the
279 olfactory/vomeronasal clusters, indicating an unusual pattern of genetic admixture in this region.
280 We also estimated genealogies around the target nonsynonymous SNVs of *Vmn2r* genes using
281 RELATE software (Speidel et al. 2019). The genealogies encompassing the SNVs showed that

282 the *castaneus*-type Japanese alleles coalesced with the other *castaneus*-type alleles after the split
283 of the subspecies *castaneus* and *musculus*, which was estimated to occur around 200 kya in our
284 data. The examples of nonsynonymous SNV segregation patterns and estimated genealogies are
285 shown in Fig. 5.



286

287 Fig. 4: Haplotype structure of nonsynonymous SNVs and genealogies of wild house mouse
288 *Vmn2r65* and *Vmn2r70*.

289 The upper panels show the segregation pattern of nonsynonymous SNVs. The sites in the red
290 rectangles represent the focal sites that showed *castaneus*-ancestry enrichment in the Japanese
291 (JPN) samples. The lower panels show the inferred genealogies around the focal sites. The
292 branches with the derived mutations at the focal sites are colored in red. The labels of samples
293 are shown at the bottom of the tree.

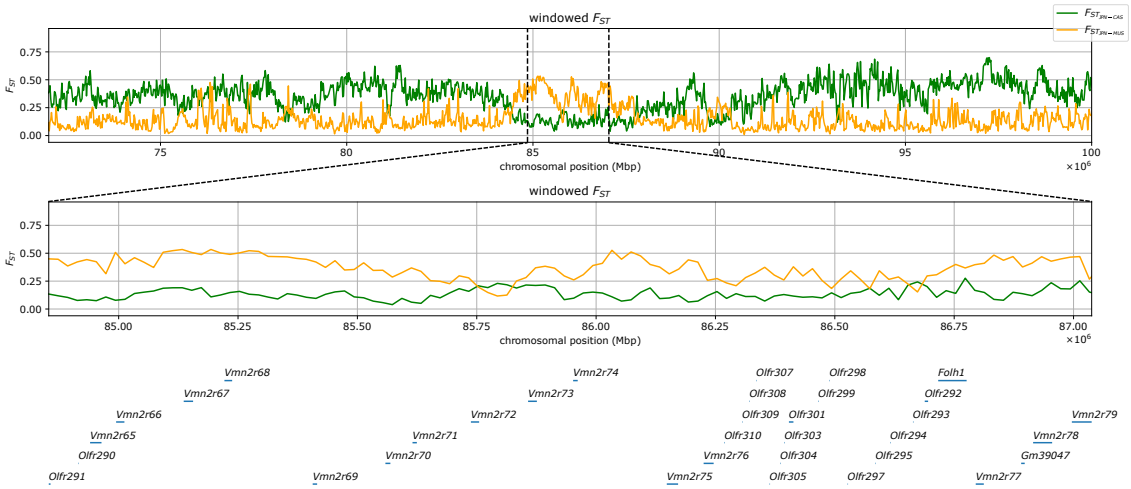


Fig. 5: Distribution of F_{ST} in 20kb-length-windows on chromosome 7.

The yellow and green lines indicate F_{ST} between the Japanese and subspecies *musculus* samples and F_{ST} between the Japanese and subspecies *castaneus* samples, respectively. The lower panel shows the pattern in the olfactory/vomeronasal genes in this cluster (chr7:84,853,553-87,037,968, GRCm38).

300 **Discussion**

301 *Genetic differentiation and inferred history of M. musculus in East Asia*

302 The widespread coexistence of the subspecies *castaneus* and *musculus* in East Asia suggests
303 that they formed a broad hybrid zone following their initial migration from their homeland.
304 Previous studies have inferred migration scenarios primarily based on phylogeographical patterns
305 of mitochondrial loci (Jing et al. 2014; Li et al. 2020). We present a combined view from
306 autosomal, mitochondrial, X-chromosome, and Y-chromosome loci results in Supplementary
307 Note 3.

308 *Spread of musculus-type Y chromosomes in East Asia*

309 In this study we uncovered natural variation in Y chromosome of house mouse across Eurasia,
310 identifying distinct clusters corresponding to major subspecies. Surprisingly, all Y chromosomes
311 in East Asian samples were of the *musculus*-type, which could be explained by various
312 mechanisms such as natural/sexual selection, male-biased migration, and genetic drift. Genomic
313 conflict under the *Sly*/*Slx* system is likely a key factor, where increased *Sly* copy number promotes
314 *musculus*-type Y chromosome spread in admixed populations like Japan and southern China. In
315 Europe, hybridization reduces the selective advantage of *musculus*-type Y chromosomes due to
316 male hybrid sterility, while *musculus*-type Y chromosomes have an advantage over *castaneus*-
317 type Y chromosomes in interbreeding with subspecies *castaneus*. Our cluster-level analysis of *Sly*
318 copy number variations revealed that *Sly* alleles, characterized by distinct amino acid sequences,
319 expanded disparately across various Y chromosome lineages. These distinct sequences might
320 confer unique functional attributes to each *Sly* allele. The question of whether these different
321 amino acid sequences result in functional differentiation among *Sly* alleles remains a captivating
322 question in future studies.

323 *Post-admixture natural selection in the Japanese archipelago*

324 The Japanese samples showed high *castaneus*-ancestry in certain genomic regions despite their
325 predominantly *musculus*-like genetic background. As demonstrated in previous studies (Lawal et
326 al. 2021), genes involved in host defense mechanisms were significantly enriched in the outlier
327 loci. In the case of humans, several studies have shown evidence of adaptive introgression from
328 archaic to modern humans at immune-related loci (Racimo et al. 2015; Enard and Petrov 2018).
329 For example, we identified *Irgm1* and *Irgm2*, which are GTPases involved in the interferon
330 signaling pathway, in our candidate list. In particular, *Irgm2* has been shown to be responsible for
331 defense against Toxoplasma, and numerous *castaneus*-like tightly-linked nonsynonymous SNVs
332 were common in the Japanese samples (Supplementary Figure 8 and 9). In another example, the
333 *castaneus*-ancestry enriched regions contained six cathepsin genes on chromosome 13. Two of
334 them, *Ctsj* and *Ctsr*, are exclusive to adult placentas, and may be responsible for the process of
335 viral infection, such as SARS-CoV-2 (Zhao et al. 2021). Although the roles of *Ctsj* and *Ctsr* in
336 viral infection has to be demonstrated, maternal–fetal viral transmission might be a critical factor
337 during the process of hybrid formation.

338 In addition to those genes, vomeronasal receptors exhibited a strongly biased pattern of
339 segregation. These receptors are expressed in the vomeronasal organ and form a large receptor
340 family involved in vomeronasal chemosensation (Wynn et al. 2012). They recognize a wide
341 variety of chemical cues, such as pheromones from different sexes and predator odors (Dulac and
342 Torello 2003). In some vomeronasal receptor genes, most Japanese samples harbor *castaneus*-
343 type nonsynonymous alleles. Although certain vomeronasal receptor genes are highly
344 differentiated between house mouse subspecies and could serve as markers for discrimination
345 (Wynn et al. 2012), our genealogical analysis revealed that the target genes frequently crossed
346 subspecies boundaries (Fig. 4). It is thus possible that some vomeronasal alleles become adaptive
347 in other subspecies and quickly spread following introgression. Interestingly, frequent

348 introgression of olfactory and other chemosensory receptor genes, but not particularly
349 vomeronasal receptor genes, have been observed in the analysis between the subspecies
350 *domesticus* and *musculus* (Staubach et al. 2012; Janousek et al. 2015).

351 The identified vomeronasal receptors were primarily clustered in a region on mouse
352 chromosome 7, chr7:84,853,553-87,037,968. This entire ≈2Mb region displayed a strongly biased
353 pattern of *castaneus*-ancestry (Fig. 5). Interestingly, previous extensive studies revealed that
354 *Vmn2r* genes in this cluster respond to cues from conspecific mating partners (Isogai et al. 2011).
355 Detailed studies, such as gene replacement experiments, will elucidate how different vomeronasal
356 receptor alleles contribute to the formation of natural house mouse hybrids in the future.

357 Altogether, we have elucidated the genomic landscape of wild house mice in East Asia,
358 demonstrating a deep connection between human and mouse genetic histories. In East Asian
359 house mouse populations, which experienced secondary contact after the subspecies split, we
360 have inferred the effects of natural/sexual selection and genomic conflict from the observed
361 pattern of genetic structure at various genomic loci. These findings will pave the way for future
362 studies on genetic admixture between different house mouse subspecies and provide insights into
363 the genetic architecture of hybrid genomes.

364 **Methods**

365 *Genomic samples*

366 We have sequenced the whole genomes of 37 *M. musculus* samples, with mitochondrial DNA
367 sequences published in previous studies (Moriwaki et al. 1986; Suzuki et al. 2004; Terashima et
368 al. 2006; Nunome et al. 2010; Suzuki et al. 2013; Kodama et al. 2015; Kuwayama et al. 2017).
369 We combined this data with the global sample dataset used in our prior research (Fujiwara et al.
370 2022a). A detailed list of these samples can be found in Supplementary Table 1. In total, our
371 analysis included 163 *M. musculus* and 7 *M. spretus* samples.

372 *Mapping and genotype calling*

373 We sequenced the genomes of the 37 samples using the DNBSEQ platform (100- or 150-bp-
374 length paired-end). These newly sequenced whole genome samples were the samples used
375 primarily for mitochondrial analysis in previous studies. The methodology for quality checks,
376 mapping reads, and genotype calling follows the same approach as in our previous study
377 (Fujiwara et al. 2022a). We used the GRCm38 (mm10) assembly as the autosomal reference, and
378 the RefSeq sequences NC_005089 and NC_205952 as the mitochondrial references for *M.*
379 *musculus* and *M. spretus*, respectively. To determine the sexes of the 37 samples, we counted the
380 number of mapped reads to sex chromosomes as described in our previous studies (Fujiwara et
381 al. 2022a). The determined sexes for these samples can be found in Supplementary Table 1.

382 *Autosome re-sequencing analysis*

383 We conducted a principal component analysis on 163 *M. musculus* samples to investigate
384 population structure. This analysis was performed using smartpca from the Eigensoft software
385 with default parameters, except that outliers were not excluded (Patterson et al. 2006).

386 To further analyze population structure and admixture, we employed ADMIXTURE with the
387 number of clusters $K = 3$ for the 163 *M. musculus* samples. Before admixture inference, we pruned

388 SNPs under linkage disequilibrium (LD) using PLINK with the option "--indep-pairwise 50 5
389 0.5". In downstream analyses, the majority of ancestral proportions determined by ADMIXTURE
390 analysis define the subspecies of each sample.

391 To quantify the fraction of genomes contributed by the subspecies *musculus* (α) in the East
392 Asian samples, we estimated f_4-ratio statistics (Patterson et al. 2012). The estimation of f_4 statistics
393 was carried out using the "patterson_d" function in the Scikit-allel Python package. Assuming the
394 source of *musculus* genomic components was more closely related to the Korean *musculus*
395 samples than to Kazakh *musculus* samples, we estimated the α value for sample X as equation
396 (1):

397 $$\hat{\alpha} = \frac{f_4(KAZ,SPR;X,IND)}{f_4(KAZ,SPR;KOR,IND)} \qquad (1)$$

398 , where KAZ, SPR, KOR, and IND represent the samples from Kazakhstan (5 samples), *M.*
399 *spretus* (7 samples), Korea (9 samples), and India (2 samples from Leh), respectively. The model
400 is also presented in Supplementary Figure 6. In principle, α is equal to or smaller than 1, but it
401 may slightly exceed 1 because of statistical fluctuation. When α >1, we therefore set the α to 1.
402 Additionally, we estimated the values of α for each 20kb-length window for all Japanese samples.
403 To remove unreliably estimated values, we excluded windows with fewer than 100 SNVs. Both
404 the denominator and numerator of the right side of equation 1 are expected to be positive,
405 considering that the Japanese and Korean samples are genetically closer to the subspecies
406 *musculus* (KAZ samples) than to the subspecies *castaneus* (IND) samples. However, a very small
407 fraction of windows showed negative denominators and/or numerators due to statistical errors,
408 introgression, or ancestral polymorphisms. Such windows were filtered out from the following
409 analysis. F_{ST} was estimated using the "hudson_fst" function in Scikit-allel.

410 *Mitochondrial genome assembly*

411 Complete mitochondrial sequences of all 37 samples were *de novo* assembled using

412 GetOrganelle software (Jin et al. 2020) with maximum extension rounds set to 10 and *k*-mer
413 parameters set to "21, 45, 65, 85, 105, 127". The qualities of assembled mitochondrial genomes
414 were visually checked using Bandage software (Wick et al. 2015). All sample sequences were
415 aligned using Clustal Omega with the "--auto" option. All D-Loop regions and gapped sites were
416 removed from the alignments. A maximum-likelihood phylogenetic tree was constructed using
417 IQ-TREE2 software (Nguyen et al. 2015) with the bootstrapping of 1000 replications. Before
418 constructing a phylogenetic tree, we performed ModelFinder software (Kalyaanamoorthy et al.
419 2017) implemented in IQ-TREE2 to determine the best substitution models. According to
420 ModelFinder, "TPM2+F+I+G4" was chosen as the best substitution model using the Bayesian
421 Information Criterion.

422 *Y chromosome re-sequencing*

423 We re-sequenced the Y chromosomes using raw read data obtained by whole genome
424 sequencing. The list of male samples that were used for the Y genotyping is shown in
425 Supplementary Table 1. The SNVs and Indels were called by GATK4 (McKenna et al. 2010)
426 HaplotypeCaller with "-ERC GVCF" option to obtain genomic variant calling files (gVCF). All
427 sample gVCF were then genotyped by GATK4 GenotypeGVCFs with "-all-sites" option. We
428 used GenMap software to calculate mappability scores with "-K 30" and "-E 2" options, and
429 filtered out the positions with mappability score <1.

430 For the phylogenetic analysis, we restricted our analysis to the short arm of this Y chromosome
431 (the first 3.4 Mb length), because the vast majority of the long arm of Y is composed of amplicon
432 sequences. We also excluded SNVs that were labeled as "LowQual" by the GATK4
433 HaplotypeCaller and those with read depths ≤3. The obtained VCF file of the Y chromosome was
434 converted to Fasta, and Nexus files. The maximum likelihood phylogenetic tree of the Y
435 chromosome was constructed by IQ-TREE software (Nguyen et al. 2015) with the bootstrapping
436 of 1000 replications. According to ModelFinder, "TIM2+F+R2" was chosen as the best

437 substitution model using Bayesian Information Criterion.

438 *Copy number estimation of Sly/Slx*

439 We downloaded all the paralogous sequences of *Sly* (ENSMUSG00000101155) and *Slx*
440 (ENSMUSG00000095063) on the sex chromosomes, including introns, from Ensembl Release
441 102. The sequences were then aligned using MAFFT (Katoh et al. 2002; Katoh and Standley
442 2013). Following the alignment, we reconstructed a phylogenetic tree using MEGA X software
443 (Kumar et al. 2018), employing a maximum likelihood method with the GTR (General Time
444 Reversible) model. From this alignment, we identified tag-SNP sites with cluster-specific alleles.

445 We started by mapping the short reads from samples to either *Sly* or *Slx* genomic sequences,
446 incorporating 1000-bp-length flanking sequences for both ends. The copy numbers for each
447 cluster were then estimated by counting the depth of cluster-specific alleles at the tag-SNP sites,
448 using the pileup format files generated from samtools mpileup command. To filter out results
449 from repetitive sequences that might cloud our analysis, we focused solely on sites with a
450 mappability score of 1, calculated using autosomal sequences and either *Sly* or *Slx* sequences. To
451 ensure the accuracy and relevance of our results, a filter was applied to exclude sites that exhibited
452 a depth smaller than 1/3 or larger than three times the median. Finally, we estimated the copy
453 number for each sample from the mean depth, which was then divided by half the average depth
454 of the whole genome for the sample.

455 *Deletion detection from sequence data*

456 DELLY (v.1.1.6) was used for genotyping retroelement-like *Fv1* (Friend virus susceptibility
457 protein 1) gene deletion. Since location of *Fv1* gene positioned between chr4:147,868,979–
458 147,870,358 in chromosome 4, we restrict regions for detecting deletions between
459 chr4:145,000,000–150,000,000 to speed up calculations.

460 *Construction of genome-wide genealogies using RELATE*

461 We estimated the genome-wide genealogies using RELATE v.1.1.9 (Speidel et al. 2019). The
462 input files were converted from variant call format (.vcf) to haplotype format (.haps) using the
463 RelateFileFormats command with the "--mode ConvertFromVcf" option implemented in the
464 RELATE package. Non-biallelic SNVs were removed from the analysis, and the *M. spretus*
465 genome was used to assign derived mutations in *M. musculus*. A germline mutation rate of $5.7 \times$
466 10^{-9} per bp per generation (Milholland et al. 2017) and a generation time of 1 year, and an initial
467 effective population size parameter of 120,000 were used for estimation. We then used the
468 EstimatePopulationSize.sh script in the RELATE package to re-estimate the effective population
469 size changes over time. We set the threshold to remove uninformative trees at 0.5.

470 **Data access**

471 Supplementary material is available online. All raw short-read sequencing data generated in
472 this study have been submitted to the DNA Data Bank of Japan (DDBJ) DDBJ Sequence Read
473 Archive (DRA) database under accession number PRJDB16017. All complete mitochondrial
474 sequence generated in this study have been also submitted to the DDBJ with accession number
475 LC772928-LC772964. The data sets, parameterization files, and scripts required to reproduce the
476 analysis are deposited in the Dryad digital repository (https://datadryad.org/) under doi:
477 [https://datadryad.org/stash/share/AxCAUN6Te8zqxpLB9C87BbtTmjJbBaBHBPdNcGZsqss].

478 **Competing interests**

479 The authors declare that they have no competing interests

480 **Acknowledgments**

481 We would like to express our sincere gratitude to the late Dr. Kazuo Moriwaki for collecting
482 valuable wild house mouse samples from around the world. We thank the members of the
483 Yaponesian Genome project funded by the MEXT KAKENHI, for their valuable comments and

feedback to our research. This work was supported by MEXT KAKENHI (grant 18H05511 and 23H04846 to N. O. and grant 18H05508 to H. S.)

486 **References**

487

488 Best S, Tissier PL, Towers G, Stoye JP. 1996. Positional cloning of the mouse
489 retrovirus restriction gene Fvl. *Nature* **382**: 826-829.

490 Bishop CE, Boursot P, Baron B, Bonhomme F, Hatat D. 1985. Most classical Mus
491 musculus domesticus laboratory mouse strains carry a Mus musculus musculus Y
492 chromosome. *Nature* **315**: 70-72.

493 Boso G, Lam O, Bamunusinghe D, Oler AJ, Wollenberg K, Liu Q, Shaffer E, Kozak
494 CA. 2021. Patterns of Coevolutionary Adaptations across Time and Space in Mouse
495 Gammaretroviruses and Three Restrictive Host Factors. *Viruses* **13**: 1864.

496 Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F. 1993. The Evolution of
497 House Mice. *Annu Rev Ecol Syst* **24**: 119-152.

498 Boursot P, Din W, Anand R, Darviche D, Dod B, Von Deimling F, Talwar GP,
499 Bonhomme F. 1996. Origin and radiation of the house mouse: mitochondrial DNA
500 phylogeny. *J Evol Biol* **9**: 391-415.

501 Cocquet J, Ellis PJI, Mahadevaiah SK, Affara NA, Vaiman D, Burgoyne PS. 2012. A
502 Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the
503 Mouse. *PLoS Genet* **8**: e1002900.

504 Din W, Anand R, Boursot P, Darviche D, Dod B, Jouvin-Marche E, Orth A, Talwar
505 GP, Cazenave P-A, Bonhomme F. 1996. Origin and radiation of the house mouse: clues
506 from nuclear genes. *J Evol Biol* **9**: 519-539.

507 Dulac C, Torello AT. 2003. Molecular detection of pheromone signals in mammals:
508 from genes to behaviour. *Nature Reviews Neuroscience* **4**: 551-562.

Ďureje Ľ, Macholán M, Baird SJE, Piálek J. 2012. The mouse hybrid zone in Central Europe: from morphology to molecules. *Folia Zoologica* **61**: 308-318, 311.

Ellis PJI, Bacon J, Affara NA. 2011. Association of Sly with sex-linked gene amplification during mouse evolution: a side effect of genomic conflict in spermatids? *Hum Mol Genet* **20**: 3010-3021.

Enard D, Petrov DA. 2018. Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. *Cell* **175**: 360-371.e313.

Fujiwara K, Kawai Y, Takada T, Shiroishi T, Saitou N, Suzuki H, Osada N. 2022a. Insights into Mus musculus Population Structure across Eurasia Revealed by Whole-Genome Analysis. *Genome Biol Evol* **14**.

Fujiwara K, Ranorosoa MC, Ohdachi SD, Arai S, Sakuma Y, Suzuki H, Osada N. 2022b. Whole-genome sequencing analysis of wild house mice (Mus musculus) captured in Madagascar. *Genes Genet Syst* **97**: 193-207.

Gabriel SI, Stevens MI, Mathias MdL, Searle JB. 2011. Of Mice and 'Convicts': Origin of the Australian House Mouse, Mus musculus. *PLOS ONE* **6**: e28622.

Gakuhari T, Nakagome S, Rasmussen S, Allentoft ME, Sato T, Korneliussen T, Chuinneagáin BN, Matsumae H, Koganebuchi K, Schmidt R et al. 2020. Ancient Jomon genome sequence analysis sheds light on migration patterns of early East Asian populations. *Communications Biology* **3**: 437.

Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu H-T, Bulatova N, Ziv Y, Nachman MW. 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Mol Ecol* **17**: 5349-5363.

GÜndüz İ, Auffray J-C, Britton-Davidian J, Catalan J, Ganem G, Ramalhinho MG,

532 Mathias ML, Searle JB. 2001. Molecular studies on the colonization of the Madeiran
533 archipelago by house mice. *Mol Ecol* **10**: 2023-2029.

534 Hanihara K. 1991. Dual Structure Model for the Population History of the Japanese.
535 *Japan Review*: 1-33.

536 Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X,
537 He M, Somel M et al. 2014. Altitude adaptation in Tibetans caused by introgression of
538 Denisovan-like DNA. *Nature* **512**: 194-197.

539 Isogai Y, Si S, Pont-Lezica L, Tan T, Kapoor V, Murthy VN, Dulac C. 2011. Molecular
540 organization of vomeronasal chemoreception. *Nature* **478**: 241-245.

541 Janousek V, Munclinger P, Wang L, Teeter KC, Tucker PK. 2015. Functional
542 organization of the genome may shape the species boundary in the house mouse. *Mol*
543 *Biol Evol* **32**: 1208-1220.

544 Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020.
545 GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle
546 genomes. *Genome Biology* **21**: 241.

547 Jing M, Yu H-T, Bi X, Lai Y-C, Jiang W, Huang L. 2014. Phylogeography of Chinese
548 house mice (Mus musculus musculus/castaneus): distribution, routes of colonization and
549 geographic regions of hybridization. *Mol Ecol* **23**: 4387-4405.

550 Jones EP, Eager HM, Gabriel SI, Jóhannesdóttir F, Searle JB. 2013. Genetic tracking
551 of mice and other bioproxies to infer human history. *Trends Genet* **29**: 298-308.

552 Kanzawa-Kiriyama H, Jinam TA, Kawai Y, Sato T, Hosomichi K, Tajima A, Adachi
553 N, Matsumura H, Kryukov K, Saitou N et al. 2019. Late Jomon male and female genome
554 sequences from the Funadomari site in Hokkaido, Japan. *Anthropological Science* **127**:

555 83-108.

556 Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid
557 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059-
558 3066.

559 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version
560 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.

561 Kodama S, Nunome M, Moriwaki K, Suzuki H. 2015. Ancient onset of geographical
562 divergence, interpopulation genetic exchange, and natural selection on theMc1rcoat-
563 colour gene in the house mouse (Mus musculus). *Biol J Linn Soc* **114**: 778-794.

564 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular
565 Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**: 1547-
566 1549.

567 Kuwayama T, Nunome M, Kinoshita G, Abe K, Suzuki H. 2017. Heterogeneous
568 genetic make-up of Japanese house mice (Mus musculus) created by multiple
569 independent introductions and spatio-temporally diverse hybridization processes. *Biol J*
570 *Linn Soc* **122**: 661-674.

571 Lawal RA, Arora UP, Dumont BL. 2021. Selection shapes the landscape of functional
572 variation in wild house mice. *BMC Biol* **19**: 239.

573 Li Y, Fujiwara K, Osada N, Kawai Y, Takada T, Kryukov AP, Abe K, Yonekawa H,
574 Shiroishi T, Moriwaki K et al. 2020. House mouse Mus musculus dispersal in East
575 Eurasia inferred from 98 newly determined complete mitochondrial genome sequences.
576 *Heredity* doi:10.1038/s41437-020-00364-y.

577 Macholán M, Baird SJE, Munclinger P, Piálek J. 2012. *Evolution of the House Mouse*.

578 Cambridge University Press, Cambridge.

579 McColl H, Racimo F, Vinner L, Demeter F, Gakuhari T, Moreno-Mayar JV, van Driem
580 G, Gram Wilken U, Seguin-Orlando A, de la Fuente Castro C et al. 2018. The prehistoric
581 peopling of Southeast Asia. *Science* **361**: 88-92.

582 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella
583 K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: A
584 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome*
585 *Res* **20**: 1297-1303.

586 Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. 2017. Differences between
587 germline and somatic mutation rates in humans and mice. *Nat Commun* **8**: 15183.

588 Morgan AP, Pardo-Manuel de Villena F. 2017. Sequence and Structural Diversity of
589 Mouse Y Chromosomes. *Mol Biol Evol* **34**: 3186-3204.

590 Moriwaki K, Miyashita N, Suzuki H, Kurihara Y, Yonekawa H. 1986. Genetic
591 Features of Major Geographical Isolates of Mus musculus. doi:10.1007/978-3-642-
592 71304-0_6 (ed. M Potter, et al.), pp. 55-61. Springer.

593 Nagamine CM, Nishioka Y, Moriwaki K, Boursot P, Bonhomme F, Lau YFC. 1992.
594 The musculus-type Y Chromosome of the laboratory mouse is of Asian origin. *Mamm*
595 *Genome* **3**: 84-91.

596 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and
597 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol*
598 *Biol Evol* **32**: 268-274.

599 Nunome M, Ishimori C, Aplin KP, Tsuchiya K, Yonekawa H, Moriwaki K, Suzuki H.
600 2010. Detection of recombinant haplotypes in wild mice (Mus musculus) provides new

601 insights into the origin of Japanese mice. *Mol Ecol* **19**: 2474-2489.

602 Orth A, Adama T, Din W, Bonhomme F. 1998. [Natural hybridization between two
603 subspecies of the house mouse, Mus musculus domesticus and Mus musculus castaneus,
604 near Lake Casitas, California]. *Genome* **41**: 104-110.

605 Osada N, Kawai Y. 2021. Exploring models of human migration to the Japanese
606 archipelago using genome-wide genetic data. *Anthropol Sci* **129**: 45-58.

607 Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M,
608 McMillan WO, Jiggins CD. 2012. Adaptive Introgression across Species Boundaries in
609 Heliconius Butterflies. *PLoS Genet* **8**: e1002752.

610 Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T,
611 Webster T, Reich D. 2012. Ancient Admixture in Human History. *Genetics* **192**: 1065-
612 1093.

613 Patterson N, Price AL, Reich D. 2006. Population Structure and Eigenanalysis. *PLoS*
614 *Genet* **2**: e190.

615 Phifer-Rixey M, Harr B, Hey J. 2020. Further resolution of the house mouse (Mus
616 musculus) phylogeny by integration over isolation-with-migration histories. *BMC Evol*
617 *Biol* **20**: 120.

618 Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. 2015. Evidence for archaic
619 adaptive introgression in humans. *Nature Reviews Genetics* **16**: 359-371.

620 Sakai T, Kikkawa Y, Miura I, Inoue T, Moriwaki K, Shiroishi T, Satta Y, Takahata N,
621 Yonekawa H. 2005. Origins of mouse inbred strains deduced from whole-genome
622 scanning by polymorphic microsatellite loci. *Mamm Genome* **16**: 11-19.

623 Salcedo T, Geraldes A, Nachman MW. 2007. Nucleotide Variation in Wild and Inbred

Mice. *Genetics* **177**: 2277-2291.

Scavetta RJ, Tautz D. 2010. Copy Number Changes of CNV Regions in Intersubspecific Crosses of the House Mouse. *Mol Biol Evol* **27**: 1845-1856.

Searle JB, Jamieson PM, Gündüz İ, Stevens MI, Jones EP, Gemmill CEC, King CM. 2009a. The diverse origins of New Zealand house mice. *Proceedings of the Royal Society B: Biological Sciences* **276**: 209-217.

Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry RJ, Giménez MD et al. 2009b. Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society B: Biological Sciences* **276**: 201-207.

Song Y, Endepols S, Klemann N, Richter D, Matuschka F-R, Shih C-H, Nachman Michael W, Kohn Michael H. 2011. Adaptive Introgression of Anticoagulant Rodent Poison Resistance by Hybridization between Old World Mice. *Curr Biol* **21**: 1296-1301.

Speidel L, Forest M, Shi S, Myers SR. 2019. A method for genome-wide genealogy estimation for thousands of samples. *Nat Genet* **51**: 1321-1329.

Staubach F, Lorenc A, Messer PW, Tang K, Petrov DA, Tautz D. 2012. Genome Patterns of Selection and Introgression of Haplotypes in Natural Populations of the House Mouse (Mus musculus). *PLoS Genet* **8**: e1002891.

Suzuki H, Nunome M, Kinoshita G, Aplin KP, Vogel P, Kryukov AP, Jin M-L, Han S-H, Maryanto I, Tsuchiya K et al. 2013. Evolutionary and dispersal history of Eurasian house mice Mus musculus clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity* **111**: 375-390.

Suzuki H, Shimada T, Terashima M, Tsuchiya K, Aplin K. 2004. Temporal, spatial,

647 and ecological modes of evolution of Eurasian Mus based on mitochondrial and nuclear
648 gene sequences. *Mol Phylogen Evol* **33**: 626-646.

649 Takada T, Ebata T, Noguchi H, Keane TM, Adams DJ, Narita T, Shin-I T, Fujisawa
650 H, Toyoda A, Abe K et al. 2013. The ancestor of extant Japanese fancy mice contributed
651 to the mosaic genomes of classical inbred strains. *Genome Res* **23**: 1329-1338.

652 Terashima M, Furusawa S, Hanzawa N, Tsuchiya K, Suyanto A, Moriwaki K,
653 Yonekawa H, Suzuki H. 2006. Phylogeographic origin of Hokkaido house mice (Mus
654 musculus) as indicated by genetic markers with maternal, paternal and biparental
655 inheritance. *Heredity* **96**: 128-138.

656 Tucker PK, Lee BK, Lundrigan BL, Eicher EM. 1992. Geographic origin of the Y
657 chromosomes in "old" inbred strains of mice. *Mamm Genome* **3**: 254-261.

658 White MA, Steffy B, Wiltshire T, Payseur BA. 2011. Genetic Dissection of a Key
659 Reproductive Barrier Between Nascent Species of House Mice. *Genetics* **189**: 289-304.

660 White MA, Stubbings M, Dumont BL, Payseur BA. 2012. Genetics and Evolution of
661 Hybrid Male Sterility in House Mice. *Genetics* **191**: 917-934.

662 Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of
663 de novo genome assemblies. *Bioinformatics* **31**: 3350-3352.

664 Wynn EH, Sánchez-Andrade G, Carss KJ, Logan DW. 2012. Genomic variation in the
665 vomeronasal receptor gene repertoires of inbred mice. *BMC Genomics* **13**: 415.

666 Yonekawa H, Moriwaki K, Gotoh O, Miyashita N, Matsushima Y, Shi LM, Cho WS,
667 Zhen XL, Tagashira Y. 1988. Hybrid origin of Japanese mice "Mus musculus
668 molossinus": evidence from restriction analysis of mitochondrial DNA. *Mol Biol Evol* **5**:
669 63-78.

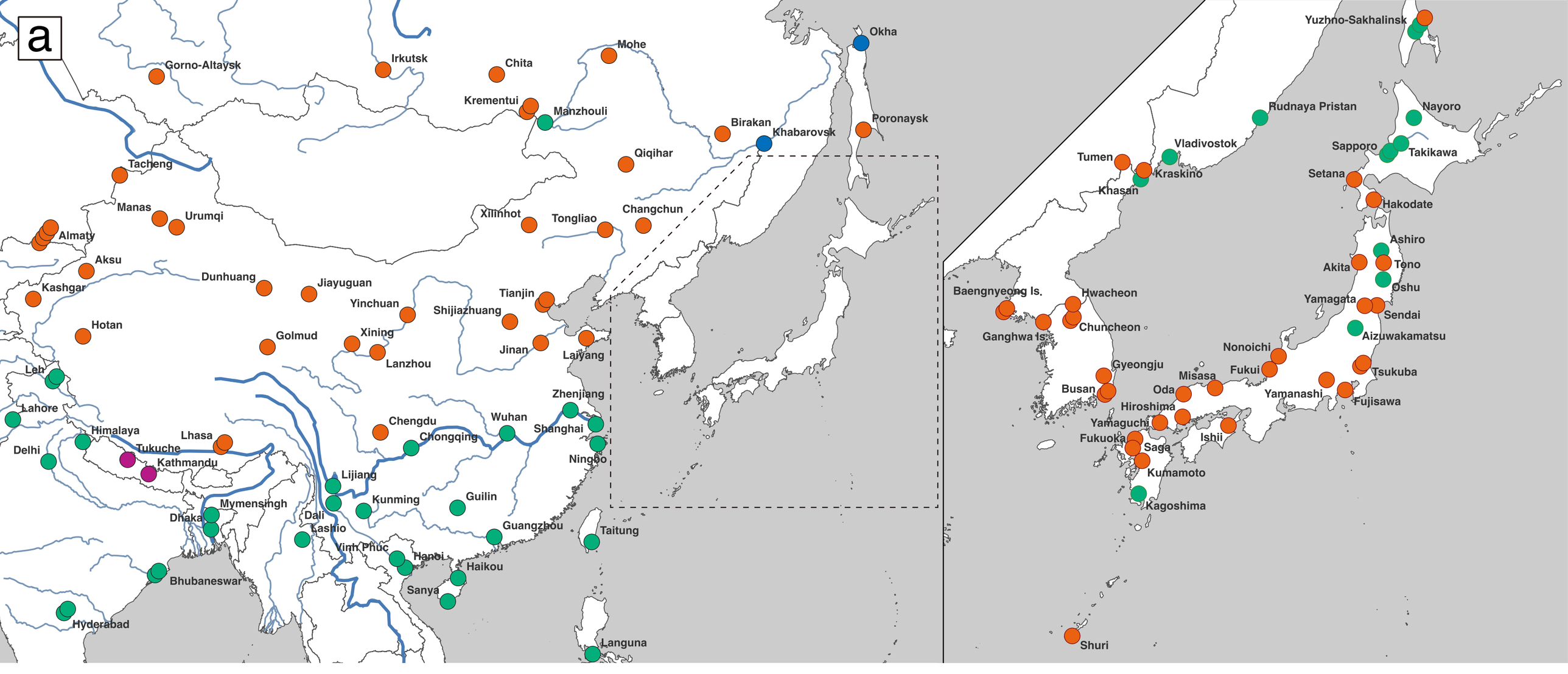
670 Yonekawa H, Moriwaki K, Gotoh O, Miyashita N, Migita S, Bonhomme F, Hjorth JP,
671 Petras ML, Tagashira Y. 1982. Origins of laboratory mice deduced from restriction
672 patterns of mitochondrial DNA. *Differentiation* **22**: 222-226.

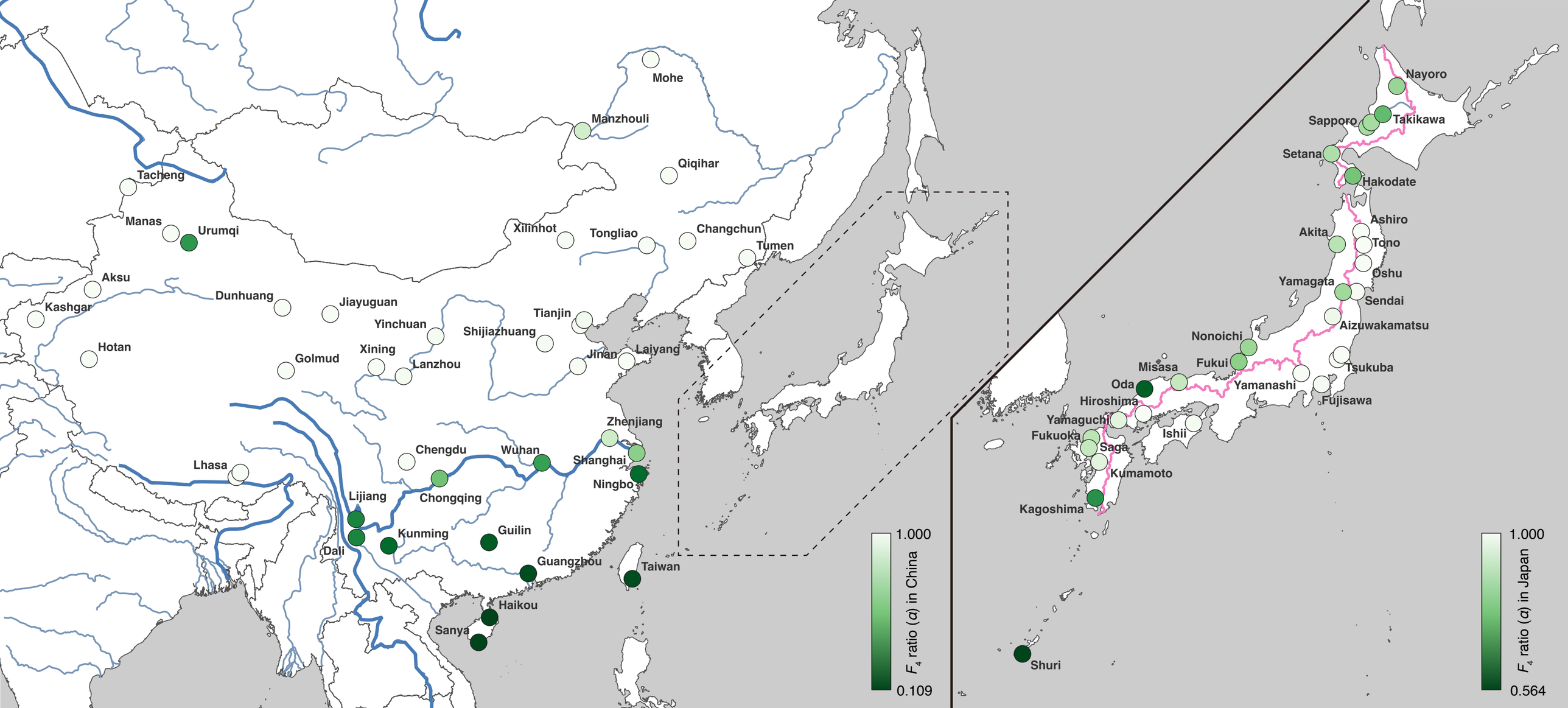
673 Yonekawa H, Moriwaki K, Gotoh O, Watanabe J, Hayashi J-I, Miyashita N, Petras
674 ML, Tagashira Y. 1980. Relationship Between Laboratory Mice and the Subspecies *Mus*
675 *musculus domesticus* Based on Restriction Endonuclease Cleavage Patterns of
676 Mitochondrial DNA. *The Japanese journal of genetics* **55**: 289-296.

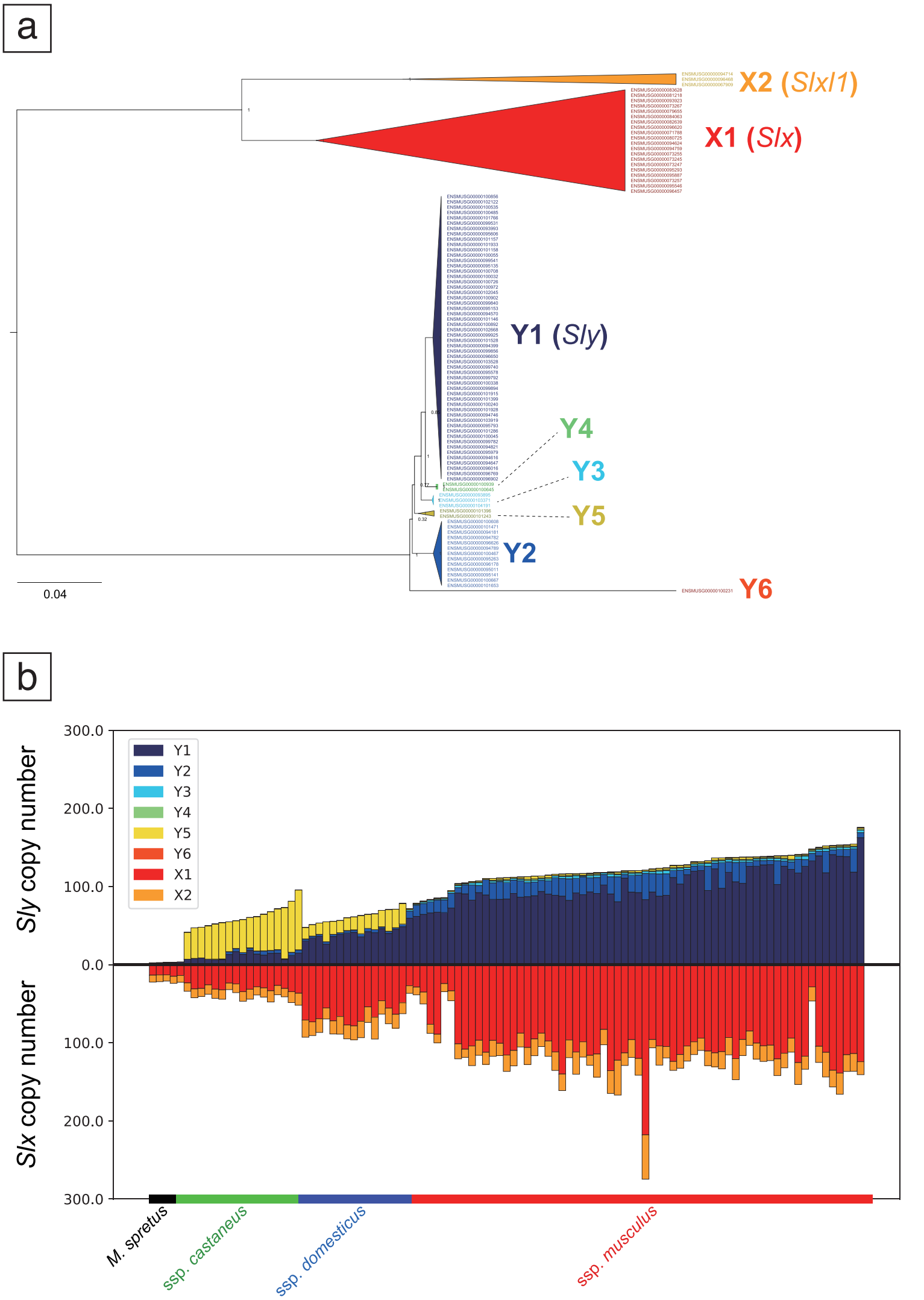
677 Zhao M-M, Yang W-L, Yang F-Y, Zhang L, Huang W-J, Hou W, Fan C-F, Jin R-H,
678 Feng Y-M, Wang Y-C et al. 2021. Cathepsin L plays a key role in SARS-CoV-2 infection
679 in humans and humanized mice and is a promising target for new drug development.
680 *Signal Transduction and Targeted Therapy* **6**: 134.

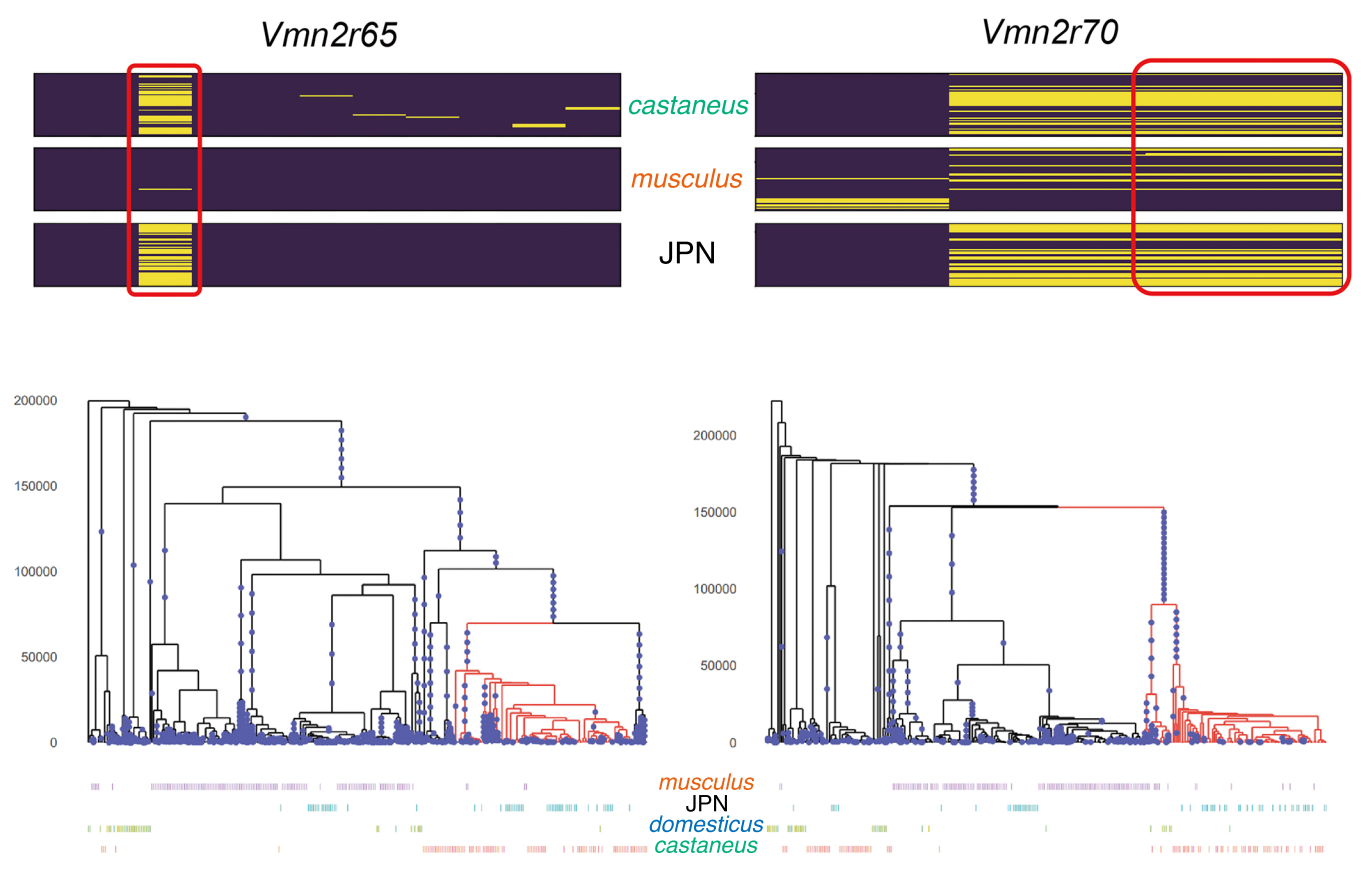
681 Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C,
682 Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of
683 systems-level datasets. *Nat Commun* **10**: 1523.

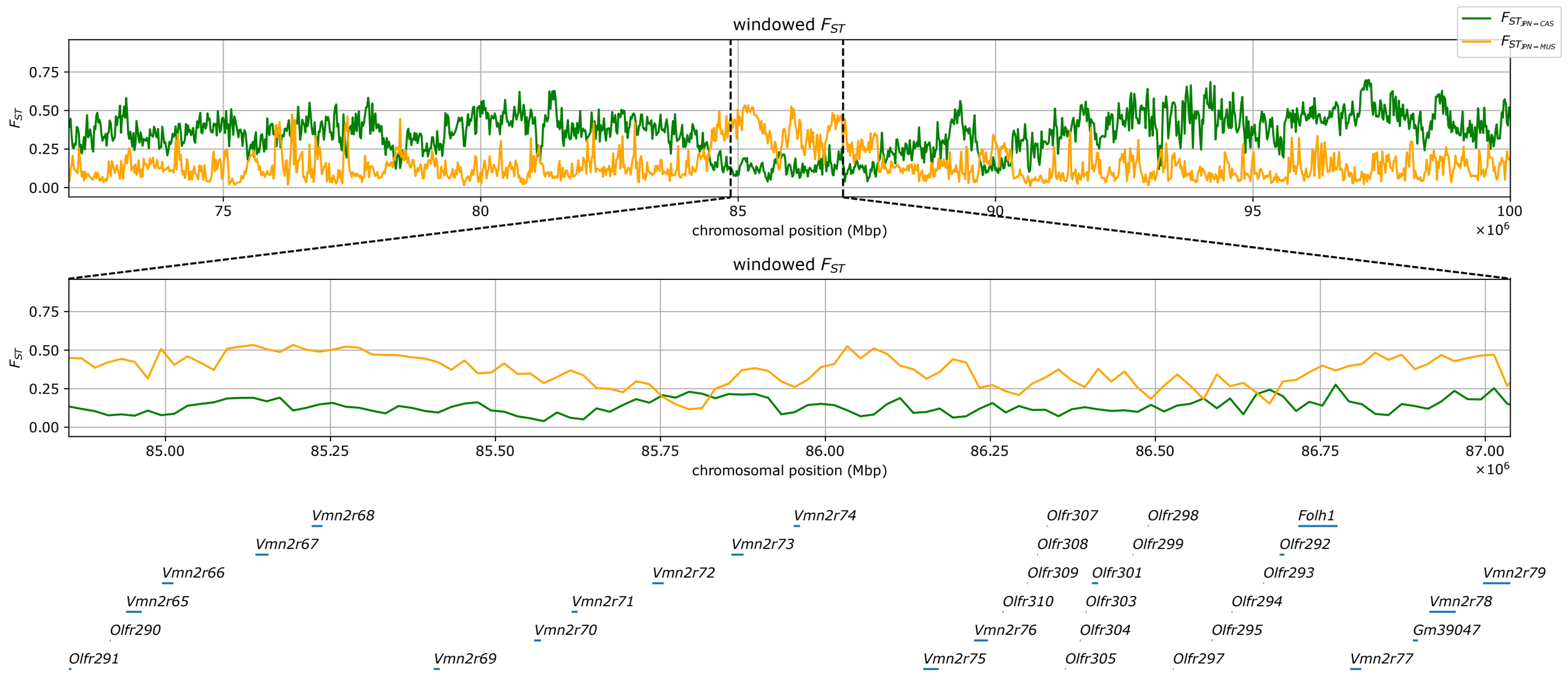
684


a
Gorno-Altaysk
Irkutsk
Chita
Mohe
Okha
Krementui
Manzhouli
Birakan
Khabarovsk
Poronaysk
Qiqihar
Tacheng
Manas
Urumqi
Xilinhot
Tongliao
Changchun
Almaty
Aksu
Kashgar
Dunhuang
Jiayuguan
Yinchuan
Shijiazhuang
Tianjin
Hotan
Golmud
Xining
Jinan
Lanzhou
Laiyang
Leh
Zhenjiang
Lahore
Chengdu
Wuhan
Shanghai
Himalaya
Lhasa
Chongqing
Delhi
Tukuche
Kathmandu
Ningbo
Lijiang
Kunming
Guilin
Mymensingh
Dhaka
Dali
Lashio
Guangzhou
Taitung
Vinh Phuc
Hanoi
Haikou
Bhubaneswar
Sanya
Hyderabad
Languna
Yuzhno-Sakhalinsk
Rudnaya Pristan
Nayoro
Vladivostok
Sapporo
Takikawa
Tumen
Kraskino
Setana
Khasan
Hakodate
Ashiro
Akita
Tono
Oshu
Baengnyeong Is.
Hwacheon
Yamagata
Sendai
Ganghwa Is.
Chuncheon
Aizuwakamatsu
Nonoichi
Gyeongju
Fukui
Tsukuba
Misasa
Busan
Oda
Yamanashi
Fujisawa
Hiroshima
Yamaguchi
Ishii
Fukuoka
Saga
Kumamoto
Kagoshima
Shuri
b
Gorno-Altaysk
Irkutsk
Okha
Krementui
Manzhouli
Birakan
Tacheng
Manas
Tongliao
Almaty
Aksu
Kashgar
Dunhuang
Jiayuguan
Yinchuan
Tianjin
Jinan
Lanzhou
Laiyang
Leh
Zhenjiang
Lahore
Chengdu
Wuhan
Shanghai
Lhasa
Delhi
Tukuche
Kathmandu
Lijiang
Kunming
Dali
Guangzhou
Taitung
Vinh Phuc
Hanoi
Bhubaneswar
Hyderabad
Yuzhno-Sakhalinsk
Rudnaya Pristan
Nayoro
Sapporo
Takikawa
Kraskino
Setana
Khasan
Hakodate
Ashiro
Oshu
Baengnyeong Is.
Hwacheon
Yamagata
Chuncheon
Gyeongju
Nonoichi
Tsukuba
Busan
Misasa
Oda
Fukuoka
Saga
Kagoshima
Shuri


Mohe
Manzhouli
Qiqihar
Tacheng
Manas
Urumqi
Xilinhot
Tongliao
Changchun
Tumen
Aksu
Kashgar
Dunhuang
Jiayuguan
Yinchuan
Shijiazhuang
Tianjin
Hotan
Golmud
Xining
Lanzhou
Jinan
Laiyang
Zhenjiang
Chengdu
Wuhan
Shanghai
Lhasa
Ningbo
Lijiang
Chongqing
Kunming
Guilin
Dali
Guangzhou
Taiwan
Haikou
Sanya
1.000
F_4 ratio (α) in China
0.109
Nayoro
Sapporo
Takikawa
Setana
Hakodate
Ashiro
Akita
Tono
Oshu
Yamagata
Sendai
Aizuwakamatsu
Nonoichi
Fukui
Misasa
Tsukuba
Oda
Yamanashi
Hiroshima
Fujisawa
Yamaguchi
Ishii
Fukuoka
Saga
Kumamoto
Kagoshima
Shuri
1.000
F_4 ratio (α) in Japan
0.564


a
X2 (Slxl1)
X1 (Slx)
Y1 (Sly)
Y4
Y3
Y5
Y2
Y6
0.04
b
Sly copy number
Slx copy number
300.0
200.0
100.0
0.0
Y1
Y2
Y3
Y4
Y5
Y6
X1
X2
M. spretus
ssp. castaneus
ssp. domesticus
ssp. musculus


Vmn2r65
Vmn2r70
castaneus
musculus
JPN
200000
150000
100000
50000
0
musculus
JPN
domesticus
castaneus


windowed F_{ST}
$F_{ST_{JPN-CAS}}$
$F_{ST_{JPN-MUS}}$
F_{ST}
0.75
0.50
0.25
0.00
75
80
85
90
95
100
chromosomal position (Mbp)
$\times 10^6$
windowed F_{ST}
F_{ST}
0.75
0.50
0.25
0.00
85.00
85.25
85.50
85.75
86.00
86.25
86.50
86.75
87.00
chromosomal position (Mbp)
$\times 10^6$
Vmn2r68
Vmn2r67
Vmn2r66
Vmn2r65
Olfr290
Olfr291
Vmn2r69
Vmn2r70
Vmn2r71
Vmn2r72
Vmn2r73
Vmn2r74
Vmn2r75
Vmn2r76
Olfr310
Olfr309
Olfr308
Olfr307
Olfr305
Olfr304
Olfr303
Olfr301
Olfr297
Olfr299
Olfr298
Olfr295
Olfr294
Olfr293
Olfr292
Folh1
Vmn2r77
Gm39047
Vmn2r78
Vmn2r79