**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

## 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 [1-8]. The house mouse exhibits high morphological and genetic diversity, with three main subspecies, recognized: *M. musculus castaneus*, *M. m. musculus*, and *M. m. domesticus* [9-12]. These subspecies diverged around 180–500 kya and underwent secondary contact after dispersing from their native habitat around South Asia [13,14] with asymmetric hybrid incompatibility; the subspecies *domesticus* shows relatively strong hybrid incompatibility with the other two subspecies [15,16]. Despite partial reproductive isolation between subspecies, genetic studies have revealed a non-negligible amount of gene flow, and occasional identification of hybrid genotypes [12,17,18]. 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 [12,14,19]. 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* [20-24], while most of wild Japanese house mice have major genetic background of the subspecies *musculus* [12]. Interestingly, mitochondrial haplogroups representing the subspecies *castaneus* are mainly distributed in northern Japan, while mitochondrial haplogroups representing the subspecies *musculus* are observed in southern Japan [7,25,26], leading the hypothesis that the subspecies *castaneus* was introduced from southern China to the Japanese archipelago and *musculus* was introduced from the Korean Peninsula with wet-rice field cultivation system [25,26]. This hypothesis would be similar to the well-known model of the origin of modern Japanese, called the dual structure model by Hanihara (1991). The dual structure model assumes that the genetic makeup of modern Japanese has been composed of both indigenous Jomon hunter-gatherers and newly migrated Yayoi rice farmers [27]. While recent genetic analyses of modern and ancient human genome studies have revealed that the Jomon originated from one of the basal East Asian populations [28-31], it is largely unknown when and from where the subspecies *castaneus* arrived in the Japanese archipelago.

Although the genetic diversity of maternally-inherited mitochondrial haplotypes in wild house mice has been well studied, natural variation in the paternally-inherited Y chromosome was largely unknown except for a few examples [32]. Recent experimental studies have discovered X- and Y-linked loci that are crucial for hybrid incompatibility, where copy number variation (CNV) plays a significant

role in determining hybrid incompatibility and sex-ratio distortion (SD). Previous studies have demonstrated that Y-linked *Sly* suppresses post-meiotic expression of X- and Y-linked genes in sperms [33]. Copy number increase in *Sly* results in male-biased progenies, while copy number increase in X-linked *Slx1* are known to counteract it [34,35]. Morgan et al. (2017) showed that the subspecies *musculus* tended to have higher *Sly*/*Slx* copy numbers than other subspecies, but their analyzed samples are biased toward wild-derived inbred lines [31]. Therefore, the analysis for CNVs of *Sly*/*Slx* and their natural distribution range in wild samples would provide how the sex-ratio distortion alleles affect the genetic differentiation of mouse subspecies.

The selective forces on the formation of hybrid genomes are also an important issue in evolutionary biology research. With the advancement of genome sequencing technology, many genome-scale studies have been conducted to examine the effects of natural selection under interspecific or intersubspecific gene flow. In particular, adaptive gene introgression, in which beneficial alleles cross the boundary of populations and spread rapidly to another population, has attracted the interest of many researchers [36-39]. The distinctive genetic characteristic of wild Japanese house mice would provide a great opportunity to investigate the effects of selection during hybridization. Having the major genetic background of the subspecies *musculus*, the preference for the genomic region of the *castaneus*-ancestry in the Japanese population leaves a similar signature to that of adaptive introgression and is a robust signal of natural selection during hybrid genome formation.

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

## Results

*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 Fig. 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 Fig. 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.

*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 *musculus*-ancestry in the southern Chinese population. The estimated α values on the X chromosome ranged from 0.317 to 1 (average α: 0.860) for Japanese samples and from 0.034 to 0.952 (average α: 0.207) for southern

Chinese samples, showing that the X chromosome has a low level of introgression into the predominant genome. The difference in α between autosomes and X chromosomes were significant in both Japanese ($P$ = 0.0008) and southern Chinese ($P$ = 0.0134) populations (Wilcoxon signed rank test).

Genealogies of the Y-chromosomal and mitochondrial loci inferred by the maximum likelihood (ML) method are shown in Supplementary Fig. 3 and 4. The Y-chromosomal locus formed three distinct haplogroups corresponding to the three major subspecies, whereas the mitochondrial locus showed five major haplogroups. In addition to the three mitochondrial haplogroups representing the subspecies *castaneus*, *domesticus,* and *musculus*, samples from Madagascar and Nepal formed distinct mitochondrial haplogroups, as previously shown [12,40].

The geographic distribution of the subspecies *castaneus* and *musculus* haplotypes was markedly different between mitochondrial and Y-chromosomal loci (Fig. 2). The distribution of mitochondrial haplotypes was consistent with previous findings (Fig. 2a). In the Japanese archipelago, *castaneus*- and *musculus*-type mitochondrial haplotypes are distributed in the northern and southern regions of mainland Japan, respectively. In China, the distribution of mitochondrial haplotypes was clearly divided around the Yangtze River basin (or divided at Qinling Mountains), with *musculus*-type haplotypes in northern China and *castaneus*-type haplotypes in southern China. However, all Y-chromosomal haplotypes in East Asia were assigned exclusively to *musculus*-type haplotypes (Fig. 2b). The Y chromosome of the reference genome (C57BL/6J) also had a *musculus*-type haplotype, and it was closely related to the haplotype of Japanese samples from the Tohoku area (Supplementary Fig. 4), supporting the idea that the Y chromosome of classical inbred strains originated from Japanese house mice (Japanese fancy mouse) [21,41-43].

We hypothesized that the dominance of the *musculus*-type Y chromosome in East Asia is due to a distorted sex ratio caused by a dosage imbalance in the *Sly*/*Slx* genes; *Sly* and *Slx* are gametologs under X-Y sexual conflict and are highly duplicated genes (>100 copies). Previous studies have shown that *musculus*-type Y chromosomes encompass more *Sly* copies and the subspecies *musculus* have more copies of *Slx* on the X chromosome [32]. Estimating the copy numbers of *Sly* and *Slx*, however, is challenging due to the presence of highly similar but distinct classes of repetitive sequences in the genome. To tackle this, we first aligned the paralogous sequences of the *Sly* and *Slx* loci present in the reference genome and used these alignments to reconstruct a phylogenetic tree. The reconstructed tree showed two clusters for *Slx*, which correspond to *Slx* (X1) and *Slxl1* (X2), and six clusters for *Sly* (Y1–Y6), with one cluster (Y1) being the most abundant in the reference genome and including the canonical *Sly* (Fig. 3a). We then estimated the copy numbers for each cluster based on the short-read depth of cluster-specific SNP alleles (see methods). We observed that different subspecies-level Y

haplotypes harbored different patterns of *Sly* clusters (Fig. 3b). The cluster Y5 was the major cluster in the *castaneus*-type haplotypes, while the cluster Y1 was the major cluster in the *domesticus*-type, and *musculus*-type Y haplotypes, respectively. Particularly, *castaneus*- and *domesticus*-type Y haplotypes had low *Sly* copy numbers, while *musculus*-type Y haplotypes had high copy numbers among the three haplogroups. In contrast, the copy number ratios of *Slx* and *Slxl1* were mostly consistent among samples. As shown in Fig. 3b and the previous study, *Sly* and *Slx* copy numbers were highly correlated ($P < 10^{-22}$, Spearman's rank correlation test) [32].

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

*Selection during the admixture in the Japanese population*

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

To discern the neutral distribution of *castaneus*-enriched genomic blocks within the Japanese house mouse population, we employed the Fastsimcoal2 software. This aided in inferring demographic

parameters from our sample set, which were used in estimating $\alpha$ (See Supplementary Note 2, Supplementary Fig. 6, and Supplementary Table 3 for details). Using these parameters, we executed 100,000 coalescent simulations to ascertain a distribution of $\alpha$ under neutrality. From our results, we identified 4194 windows. These windows, which fell below the 5% threshold of the neutral distribution, encompassed the protein-coding regions of 842 distinct genes. A comprehensive list of these genes can be found in Supplementary Table 4.

We performed a functional enrichment analysis on the 842 candidate genes using MetaScape software [44]. A total of 20 gene categories were found to be enriched in the list ($p$ values < 0.01), with the complete list of significantly overrepresented functional categories presented in Supplementary Fig. 7. The enriched term list contained a variety of gene categories, but notably, immune-related gene categories such as immunoglobulin production (GO:002377), antibody-dependent cellular cytotoxicity (GO:0001788), and Herpes simplex virus 1 infection (mmu05168) were overrepresented. Indeed, our screening accurately identified the retroelement-like *Fv1* (Friend virus susceptibility protein 1) gene, an antiviral factor [45] as an *castaneus*-ancestry enriched gene in Japan. Previous studies have demonstrated that the $Fv1^b$ allele, which lacks the 1.3-kb-length deletion commonly found in the subspecies *domesticus* and *musculus*, is observed with high frequency in the subspecies *castaneus* and the Japanese samples [46]. We confirmed that the deletion started from chr4:147,870,288 (GRCm38) and had 1.2 kb in length, resulting in the truncation of C-terminal region of *Fv1*. The allele frequencies of deletion were 0.92, 0.64, and 0.16 in the subspecies *domesticus*, *musculus*, and *castaneus*, respectively. The allele frequency in the Japanese samples was 0.02, showing that almost no Japanese house mice harbor the deletion.

We also observed that 23 vomeronasal receptor genes (*Vmn1r* and *Vmn2r*), including one pseudogene, were present in the gene list. In the enrichment analysis, a category, detection of chemical stimulus involved in sensory perception (GO0050907) showed the strongest bias. These *castaneus*-enriched vomeronasal receptor genes were scattered across different olfactory/vomeronasal genomic clusters on the mouse genome, with most genes located in a large cluster on chromosome 7: region chr7:84,853,553-87,037,968, containing 17 olfactory and 15 vomeronasal receptors. Of the 23 vomeronasal receptor genes identified in our genomic scan, 12 were located in the cluster. To confirm that the estimated low $\alpha$ in the region was not due to errors in statistical inference, we computed $F_{ST}$ between Japanese and subspecies *musculus* samples ($F_{ST\text{-}MUS/JPN}$) and $F_{ST}$ between Japanese samples and subspecies *castaneus* samples ($F_{ST\text{-}CAS/JPN}$). The subspecies *castaneus* and *musculus* samples were assigned from the result of the ADMIXTURE analysis. We plotted window-averaged $F_{ST\text{-}CAS/JPN}$ and $F_{ST\text{-}MUS/JPN}$ across the region including chr7:84,853,553-87,037,968 (Fig. 4). As expected, outside of this region, $F_{ST\text{-}MUS/JPN}$ values were consistently smaller than $F_{ST\text{-}CAS/JPN}$, supporting the notion that

Japanese samples were generally *musculus*-like. However, the pattern was completely reversed within the olfactory/vomeronasal clusters, indicating an unusual pattern of genetic admixture in this region. We also estimated genealogies around the target nonsynonymous SNVs of *Vmn2r* genes using RELATE software [47]. The genealogies encompassing the SNVs showed that the *castaneus*-type Japanese alleles coalesced with the other *castaneus*-type alleles after the split of the subspecies *castaneus* and *musculus*, which was estimated to occur around 200 kya in our data. The examples of nonsynonymous SNV segregation patterns and estimated genealogies are shown in Fig. 5.

## Discussions

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

The widespread coexistence of the subspecies *castaneus* and *musculus* in East Asia suggests that they formed a broad hybrid zone following their initial migration from their homeland. Previous studies have inferred migration scenarios primarily based on phylogeographical patterns of mitochondrial loci [7,19]. We present a combined view from autosomal, mitochondrial, X-chromosome, and Y-chromosome loci results in Supplementary Note 3.

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

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

*Post-admixture natural selection in the Japanese archipelago*

The Japanese samples showed high *castaneus*-ancestry in certain genomic regions despite their predominantly *musculus*-like genetic background. As demonstrated in previous studies [48], genes involved in host defense mechanisms were significantly enriched in the outlier loci. In the case of humans, several studies have shown evidence of adaptive introgression from archaic to modern humans at immune-related loci [39,49]. For example, we identified *Irgm1* and *Irgm2*, which are GTPases involved in the interferon signaling pathway, in our candidate list. In particular, *Irgm2* has been shown to be responsible for defense against Toxoplasma, and numerous *castaneus*-like tightly-linked nonsynonymous SNVs were common in the Japanese samples (Supplementary Fig. 8 and 9). In another example, the *castaneus*-ancestry enriched regions contained six cathepsin genes on chromosome 13. Two of them, *Ctsj* and *Ctsr*, are exclusive to adult placentas, and may be responsible for the process of viral infection, such as SARS-CoV-2 [50]. Although the roles of *Ctsj* and *Ctsr* in viral

infection has to be demonstrated, maternal–fetal viral transmission might be a critical factor during the process of hybrid formation.

In addition to those genes, vomeronasal receptors exhibited a strongly biased pattern of segregation. These receptors are expressed in the vomeronasal organ and form a large receptor family involved in vomeronasal chemosensation [51]. They recognize a wide variety of chemical cues, such as pheromones from different sexes and predator odors [52]. In some vomeronasal receptor genes, most Japanese samples harbor *castaneus*-type nonsynonymous alleles. Although certain vomeronasal receptor genes are highly differentiated between house mouse subspecies and could serve as markers for discrimination [51], our genealogical analysis revealed that the target genes frequently crossed subspecies boundaries (Fig. 4). It is thus possible that some vomeronasal alleles become adaptive in other subspecies and quickly spread following introgression.

The identified vomeronasal receptors were primarily clustered in a region on mouse chromosome 7, chr7:84,853,553-87,037,968. This entire ≈2Mb region displayed a strongly biased pattern of *castaneus*-ancestry (Fig. 5). Interestingly, previous extensive studies revealed that *Vmn2r* genes in this cluster respond to cues from conspecific mating partners [53]. Detailed studies, such as gene replacement experiments, will elucidate how different vomeronasal receptor alleles contribute to the formation of natural house mouse hybrids in the future.

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

## Material and Methods

*Genomic samples*

We have sequenced the whole genomes of 37 *M. musculus* samples, with mitochondrial DNA sequences published in previous studies [25,26,54-58]. We combined this data with the global sample dataset used in our prior research [12]. A detailed list of these samples can be found in Supplementary Table 1. In total, our analysis included 163 *M. musculus* and 7 *M. spretus* samples.

*Mapping and genotype calling*

We sequenced the genomes of the 37 samples using the DNBSEQ platform (100- or 150-bp-length paired-end). These newly sequenced whole genome samples were the samples used primarily for mitochondrial analysis in previous studies. The methodology for quality checks, mapping reads, and genotype calling follows the same approach as in our previous study [12]. We used the GRCm38 (mm10) assembly as the autosomal reference, and the RefSeq sequences NC_005089 and NC_205952 as the mitochondrial references for *M. musculus* and *M. spretus*, respectively. To determine the sexes of the 37 samples, we counted the number of mapped reads to sex chromosomes as described in our previous studies [12]. The determined sexes for these samples can be found in Supplementary Table 1.

*Autosome re-sequencing analysis*

We conducted a principal component analysis on 163 *M. musculus* samples to investigate population structure. This analysis was performed using smartpca from the Eigensoft software with default parameters, except that outliers were not excluded [59].

To further analyze population structure and admixture, we employed ADMIXTURE with the number of clusters $K = 3$ for the 163 *M. musculus* samples. Before admixture inference, we pruned SNPs under linkage disequilibrium (LD) using PLINK with the option "--indep-pairwise 50 5 0.5". In downstream analyses, the majority of ancestral proportions determined by ADMIXTURE analysis define the subspecies of each sample.

To quantify the fraction of genomes contributed by the subspecies *musculus* (α) in the East Asian samples, we estimated $f_4$-ratio statistics [60]. The estimation of $f_4$ statistics was carried out using the "patterson_d" function in the Scikit-allel Python package. Assuming the source of *musculus* genomic components was more closely related to the Korean *musculus* samples than to Kazakh *musculus* samples, we estimated the α value for sample X as equation (1):

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

, where KAZ, SPR, KOR, and IND represent the samples from Kazakhstan (5 samples), *M. spretus*

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

*Mitochondrial genome assembly*

Complete mitochondrial sequences of all 37 samples were *de novo* assembled using GetOrganelle software [61] with maximum extension rounds set to 10 and *k*-mer parameters set to "21, 45, 65, 85, 105, 127". The qualities of assembled mitochondrial genomes were visually checked using Bandage software [62]. All sample sequences were aligned using Clustal Omega with the "--auto" option. All D-Loop regions and gapped sites were removed from the alignments. A maximum-likelihood phylogenetic tree was constructed using IQ-TREE2 software [63] with the bootstrapping of 1000 replications. Before constructing a phylogenetic tree, we performed ModelFinder software (Kalyaanamoorthy et al. 2017) implemented in IQ-TREE2 to determine the best substitution models. According to ModelFinder, "TPM2+F+I+G4" was chosen as the best substitution model using the Bayesian Information Criterion.

*Y chromosome re-sequencing*

We re-sequenced the Y chromosomes using raw read data obtained by whole genome sequencing. The list of male samples that were used for the Y genotyping is shown in Supplementary Table S1. The SNVs and Indels were called by GATK4 [64] HaplotypeCaller with "-ERC GVCF" option to obtain genomic variant calling files (gVCF). All sample gVCF were then genotyped by GATK4 GenotypeGVCFs with "-all-sites" option. We used GenMap software to calculate mappability scores with "-K 30" and "-E 2" options, and filtered out the positions with mappability score <1.

For the phylogenetic analysis, we restricted our analysis to the short arm of this Y chromosome (the first 3.4 Mb length), because the vast majority of the long arm of Y is composed of amplicon sequences. We also excluded SNVs that were labeled as "LowQual" by the GATK4 HaplotypeCaller and those with read depths ≤3. The obtained VCF file of the Y chromosome was converted to Fasta, and Nexus

files. The maximum likelihood phylogenetic tree of the Y chromosome was constructed by IQ-TREE software [63] with the bootstrapping of 1000 replications. According to ModelFinder, "TIM2+F+R2" was chosen as the best substitution model using Bayesian Information Criterion.

*Copy number estimation of Sly/Slx*

We downloaded all the paralogous sequences of *Sly* (ENSMUSG00000101155) and *Slx* (ENSMUSG00000095063) on the sex chromosomes, including introns, from Ensembl Release 102. The sequences were then aligned using MAFFT [65,66]. Following the alignment, we reconstructed a phylogenetic tree using MEGA X software [67], employing a maximum likelihood method with the GTR (General Time Reversible) model. From this alignment, we identified tag-SNP sites with cluster-specific alleles.

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

*Deletion detection from sequence data*

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

*Construction of genome-wide genealogies using RELATE*

We estimated the genome-wide genealogies using RELATE v.1.1.9 [47]. The input files were converted from variant call format (.vcf) to haplotype format (.haps) using the RelateFileFormats command with the "--mode ConvertFromVcf" option implemented in the RELATE package. Non-biallelic SNVs were removed from the analysis, and the *M. spretus* genome was used to assign derived mutations in *M. musculus*. A germline mutation rate of $5.7 \times 10^{-9}$ per bp per generation [68] and a generation time of 1 year, and an initial effective population size parameter of 120,000 were used for estimation. We then used the EstimatePopulationSize.sh script in the RELATE package to re-estimate the effective population size changes over time. We set the threshold to remove uninformative trees at

0.5.

## Data availability

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

## Acknowledgments

We would like to express our sincere gratitude to the late Dr. Kazuo Moriwaki for collecting valuable wild house mouse samples from around the world. We thank the members of the 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.)

References


1. Macholán, M., Baird, S.J.E., Munclinger, P. & Piálek, J. *Evolution of the House Mouse*, (Cambridge University Press, Cambridge, 2012).

2. GÜndüz, İ. *et al.* Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology* **10**, 2023-2029 (2001).

3. Searle, J.B. *et al.* Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society B: Biological Sciences* **276**, 201-207 (2009).

4. Searle, J.B. *et al.* The diverse origins of New Zealand house mice. *Proceedings of the Royal Society B: Biological Sciences* **276**, 209-217 (2009).

5. Gabriel, S.I., Stevens, M.I., Mathias, M.d.L. & Searle, J.B. Of Mice and 'Convicts': Origin of the Australian House Mouse, Mus musculus. *PLOS ONE* **6**, e28622 (2011).

6. Jones, E.P., Eager, H.M., Gabriel, S.I., Jóhannesdóttir, F. & Searle, J.B. Genetic tracking of mice and other bioproxies to infer human history. *Trends in Genetics* **29**, 298-308 (2013).

7. Li, Y. *et al.* House mouse Mus musculus dispersal in East Eurasia inferred from 98 newly determined complete mitochondrial genome sequences. *Heredity* (2020).

8. Geraldes, A. *et al.* Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology* **17**, 5349-5363 (2008).

9. Boursot, P. *et al.* Origin and radiation of the house mouse: mitochondrial DNA phylogeny. *Journal of Evolutionary Biology* **9**, 391-415 (1996).

10. Salcedo, T., Geraldes, A. & Nachman, M.W. Nucleotide Variation in Wild and Inbred Mice. *Genetics* **177**, 2277-2291 (2007).

11. Phifer-Rixey, M., Harr, B. & Hey, J. Further resolution of the house mouse (Mus musculus) phylogeny by integration over isolation-with-migration histories. *BMC Evolutionary Biology* **20**, 120 (2020).

12. Fujiwara, K. *et al.* Insights into Mus musculus Population Structure across Eurasia Revealed by Whole-Genome Analysis. *Genome Biol Evol* **14**(2022).

13. Boursot, P., Auffray, J.C., Britton-Davidian, J. & Bonhomme, F. The Evolution of House Mice. *Annual Review of Ecology and Systematics* **24**, 119-152 (1993).

14. Din, W. *et al.* Origin and radiation of the house mouse: clues from nuclear genes. *Journal of Evolutionary Biology* **9**, 519-539 (1996).

15. White, M.A., Steffy, B., Wiltshire, T. & Payseur, B.A. Genetic Dissection of a Key Reproductive Barrier Between Nascent Species of House Mice. *Genetics* **189**, 289-304 (2011).

16. White, M.A., Stubbings, M., Dumont, B.L. & Payseur, B.A. Genetics and Evolution of Hybrid Male Sterility in House Mice. *Genetics* **191**, 917-934 (2012).

17. Orth, A., Adama, T., Din, W. & Bonhomme, F. [Natural hybridization between two subspecies of the house mouse, Mus musculus domesticus and Mus musculus castaneus, near Lake Casitas, California]. *Genome* **41**, 104-10 (1998).

18. Ďureje, Ľ., Macholán, M., Baird, S.J.E. & Piálek, J. The mouse hybrid zone in Central Europe: from morphology to molecules. *Folia Zoologica* **61**, 308-318, 11 (2012).

19. Jing, M. *et al.* Phylogeography of Chinese house mice (Mus musculus musculus/castaneus): distribution, routes of colonization and geographic regions of hybridization. *Molecular Ecology* **23**, 4387-4405 (2014).

20. Sakai, T. *et al.* Origins of mouse inbred strains deduced from whole-genome scanning by polymorphic microsatellite loci. *Mammalian Genome* **16**, 11-19 (2005).

21. Takada, T. *et al.* The ancestor of extant Japanese fancy mice contributed to the mosaic genomes of classical inbred strains. *Genome Research* **23**, 1329-1338 (2013).

22. Yonekawa, H. *et al.* Origins of laboratory mice deduced from restriction patterns of mitochondrial DNA. *Differentiation* **22**, 222-6 (1982).

23. Yonekawa, H. *et al.* Relationship Between Laboratory Mice and the Subspecies *Mus musculus domesticus* Based on Restriction Endonuclease Cleavage Patterns of Mitochondrial DNA. *The Japanese journal of genetics* **55**, 289-296 (1980).

24. Yonekawa, H. *et al.* Hybrid origin of Japanese mice "Mus musculus molossinus": evidence from restriction analysis of mitochondrial DNA. *Molecular Biology and Evolution* **5**, 63-78 (1988).

25. Kuwayama, T., Nunome, M., Kinoshita, G., Abe, K. & Suzuki, H. Heterogeneous genetic make-up of Japanese house mice (Mus musculus) created by multiple independent introductions and spatio-temporally diverse hybridization processes. *Biological Journal of the Linnean Society* **122**, 661-674 (2017).

26. Terashima, M. *et al.* Phylogeographic origin of Hokkaido house mice (Mus musculus) as indicated by genetic markers with maternal, paternal and biparental inheritance. *Heredity* **96**, 128-138 (2006).

27. Hanihara, K. Dual Structure Model for the Population History of the Japanese. *Japan Review*, 1-33 (1991).

28. Osada, N. & Kawai, Y. Exploring models of human migration to the Japanese archipelago using genome-wide genetic data. *Anthropological Science* **129**, 45-58 (2021).

29. McColl, H. *et al.* The prehistoric peopling of Southeast Asia. *Science* **361**, 88-92 (2018).

30. Gakuhari, T. *et al.* Ancient Jomon genome sequence analysis sheds light on migration patterns of early East Asian populations. *Communications Biology* **3**, 437 (2020).

31. Kanzawa-Kiriyama, H. *et al.* Late Jomon male and female genome sequences from the Funadomari site in Hokkaido, Japan. *Anthropological Science* **127**, 83-108 (2019).

32. Morgan, A.P. & Pardo-Manuel de Villena, F. Sequence and Structural Diversity of Mouse Y Chromosomes. *Molecular Biology and Evolution* **34**, 3186-3204 (2017).

33. Ellis, P.J.I., Bacon, J. & Affara, N.A. Association of Sly with sex-linked gene amplification during mouse evolution: a side effect of genomic conflict in spermatids? *Human Molecular Genetics* **20**, 3010-3021 (2011).

34. Scavetta, R.J. & Tautz, D. Copy Number Changes of CNV Regions in Intersubspecific Crosses of the House Mouse. *Molecular Biology and Evolution* **27**, 1845-1856 (2010).

35. Cocquet, J. *et al.* A Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the Mouse. *PLOS Genetics* **8**, e1002900 (2012).

36. Huerta-Sánchez, E. *et al.* Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194-197 (2014).

37. Song, Y. *et al.* Adaptive Introgression of Anticoagulant Rodent Poison Resistance by Hybridization between Old World Mice. *Current Biology* **21**, 1296-1301 (2011).

38. Pardo-Diaz, C. *et al.* Adaptive Introgression across Species Boundaries in Heliconius Butterflies. *PLOS Genetics* **8**, e1002752 (2012).

39. Enard, D. & Petrov, D.A. Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. *Cell* **175**, 360-371.e13 (2018).

40. Fujiwara, K. *et al.* Whole-genome sequencing analysis of wild house mice (Mus musculus) captured in Madagascar. *Genes Genet Syst* **97**, 193-207 (2022).

41. Bishop, C.E., Boursot, P., Baron, B., Bonhomme, F. & Hatat, D. Most classical Mus musculus domesticus laboratory mouse strains carry a Mus musculus musculus Y chromosome. *Nature* **315**, 70-72 (1985).

42. Tucker, P.K., Lee, B.K., Lundrigan, B.L. & Eicher, E.M. Geographic origin of the Y chromosomes in "old" inbred strains of mice. *Mamm Genome* **3**, 254-61 (1992).

43. Nagamine, C.M. *et al.* The musculus-type Y Chromosome of the laboratory mouse is of Asian origin. *Mammalian Genome* **3**, 84-91 (1992).

44. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* **10**, 1523 (2019).

45. Best, S., Tissier, P.L., Towers, G. & Stoye, J.P. Positional cloning of the mouse retrovirus restriction gene Fvl. *Nature* **382**, 826-829 (1996).

46. Boso, G. *et al.* Patterns of Coevolutionary Adaptations across Time and Space in Mouse Gammaretroviruses and Three Restrictive Host Factors. *Viruses* **13**, 1864 (2021).

47. Speidel, L., Forest, M., Shi, S. & Myers, S.R. A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics* **51**, 1321-1329 (2019).

48. Lawal, R.A., Arora, U.P. & Dumont, B.L. Selection shapes the landscape of functional variation in wild house mice. *BMC Biology* **19**, 239 (2021).

49. Racimo, F., Sankararaman, S., Nielsen, R. & Huerta-Sánchez, E. Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics* **16**, 359-371 (2015).

50. Zhao, M.-M. *et al.* Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *Signal Transduction and Targeted Therapy* **6**, 134 (2021).

51. Wynn, E.H., Sánchez-Andrade, G., Carss, K.J. & Logan, D.W. Genomic variation in the vomeronasal receptor gene repertoires of inbred mice. *BMC Genomics* **13**, 415 (2012).

52. Dulac, C. & Torello, A.T. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature Reviews Neuroscience* **4**, 551-562 (2003).

53. Isogai, Y. *et al.* Molecular organization of vomeronasal chemoreception. *Nature* **478**, 241-245 (2011).

54. Kodama, S., Nunome, M., Moriwaki, K. & Suzuki, H. Ancient onset of geographical divergence, interpopulation genetic exchange, and natural selection on theMc1rcoat-colour gene in the house mouse (Mus musculus). *Biological Journal of the Linnean Society* **114**, 778-794 (2015).

55. Nunome, M. *et al.* Detection of recombinant haplotypes in wild mice (Mus musculus) provides new insights into the origin of Japanese mice. *Mol Ecol* **19**, 2474-89 (2010).

56. Suzuki, H. *et al.* Evolutionary and dispersal history of Eurasian house mice Mus musculus clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity* **111**, 375-390 (2013).

57. Moriwaki, K., Miyashita, N., Suzuki, H., Kurihara, Y. & Yonekawa, H. Genetic Features of Major Geographical Isolates of Mus musculus. (eds Potter, M., Nadeau, J.H. & Cancro, M.P.) 55-61 (Springer, 1986).

58. Suzuki, H., Shimada, T., Terashima, M., Tsuchiya, K. & Aplin, K. Temporal, spatial, and ecological modes of evolution of Eurasian Mus based on mitochondrial and nuclear gene sequences. *Molecular Phylogenetics and Evolution* **33**, 626-646 (2004).

59. Patterson, N., Price, A.L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet* **2**, e190 (2006).

60. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065-1093 (2012).

61. Jin, J.-J. *et al.* GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**, 241 (2020).

62. Wick, R.R., Schultz, M.B., Zobel, J. & Holt, K.E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350-3352 (2015).

63. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268-274 (2015).

64. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).

65. Katoh, K. & Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-80 (2013).

66. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-66 (2002).

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

68. Milholland, B. *et al.* Differences between germline and somatic mutation rates in humans and mice. *Nature Communications* **8**, 15183 (2017).

## Figure Legends

*Figure 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 (A: East Asian samples except Japanese archipelago, B: 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.

*Figure 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).

*Figure 3: Copy number evolution of Sly* and *Slx*

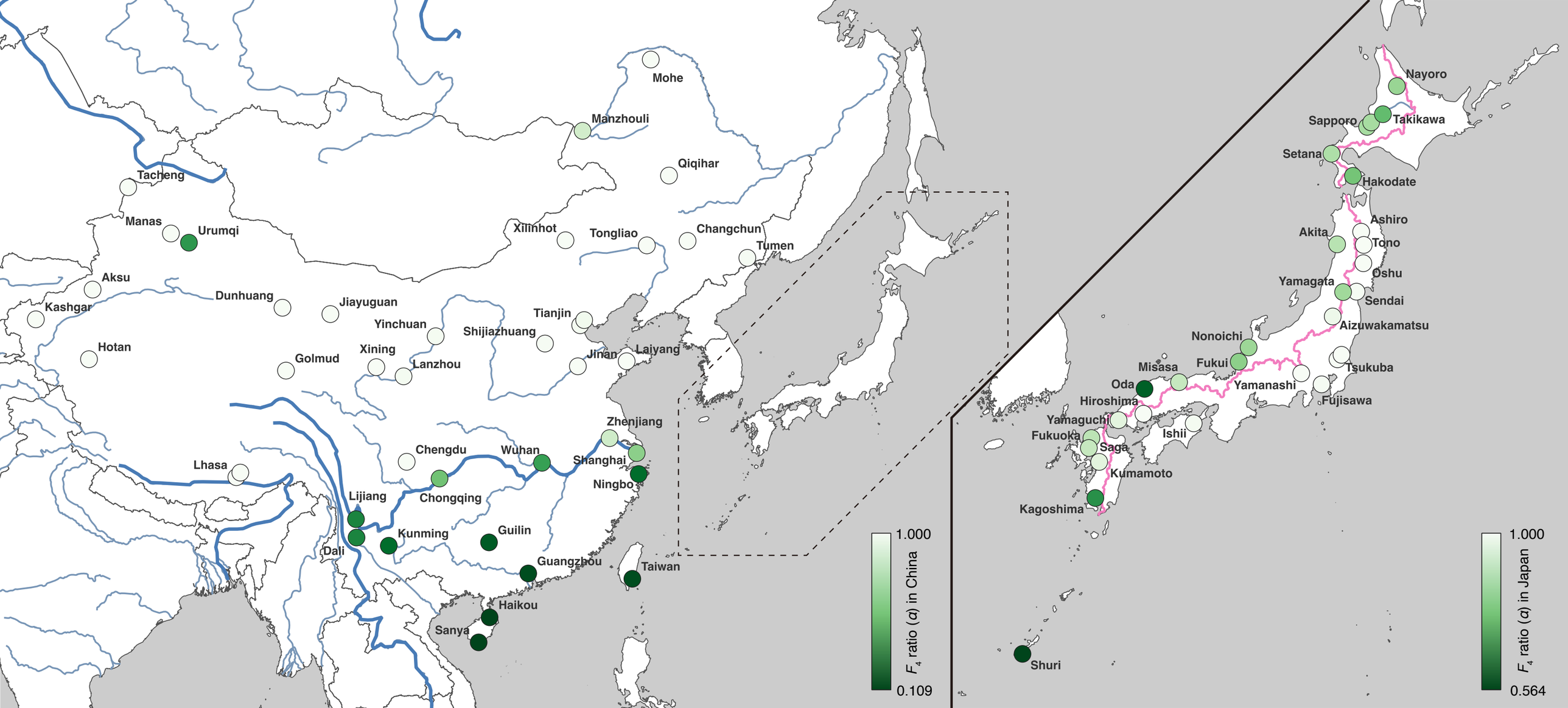
The phylogenetic tree of *Sly*/*Slx* genes on the mouse reference genome, GRCm38 (a). Bootstrap values are described at each node. Estimated copy numbers of *Sly* and *Slx* for each sample (b). Different colors represent the copy numbers of different cluster. The samples were aligned according to the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

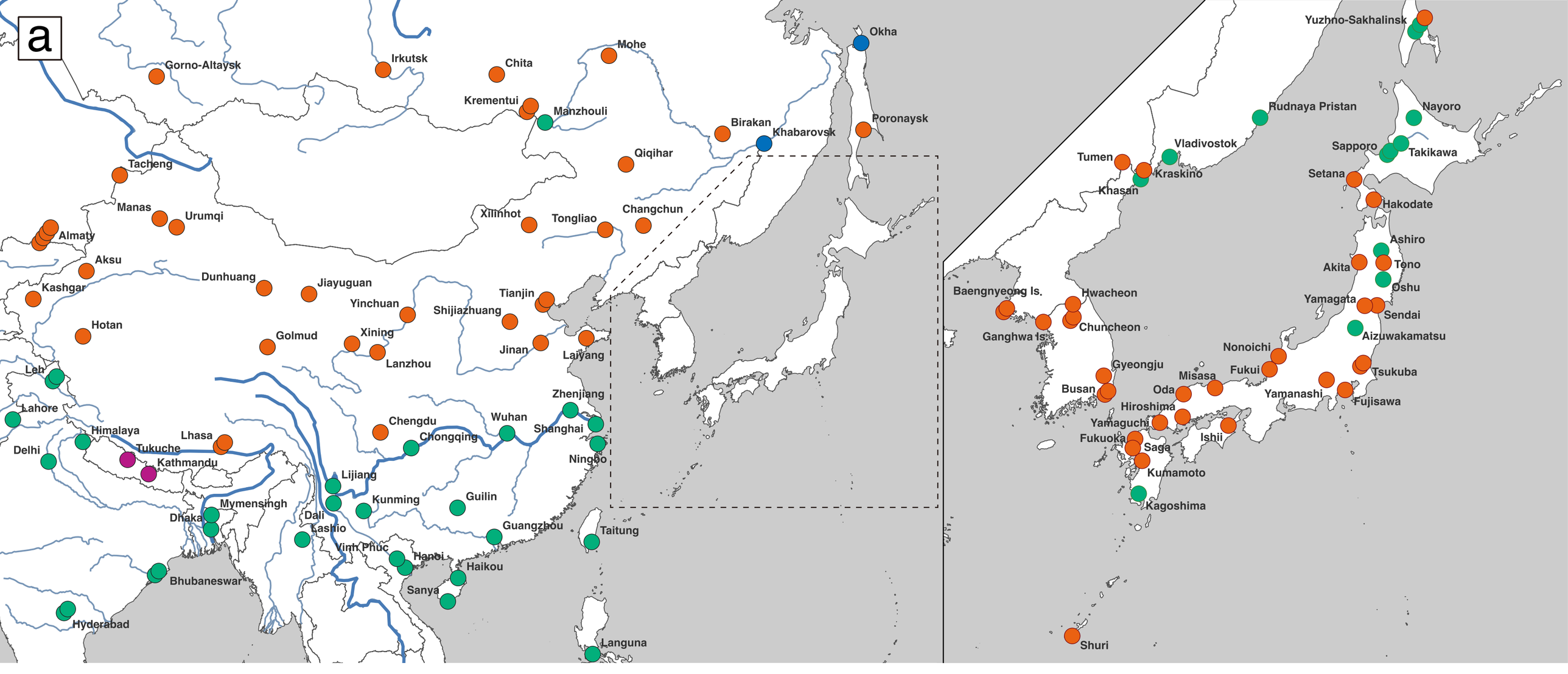
*Figure 4: Haplotype structure of nonsynonymous SNVs and genealogies of wild house mouse Vmn2r65 and Vmn2r70.*

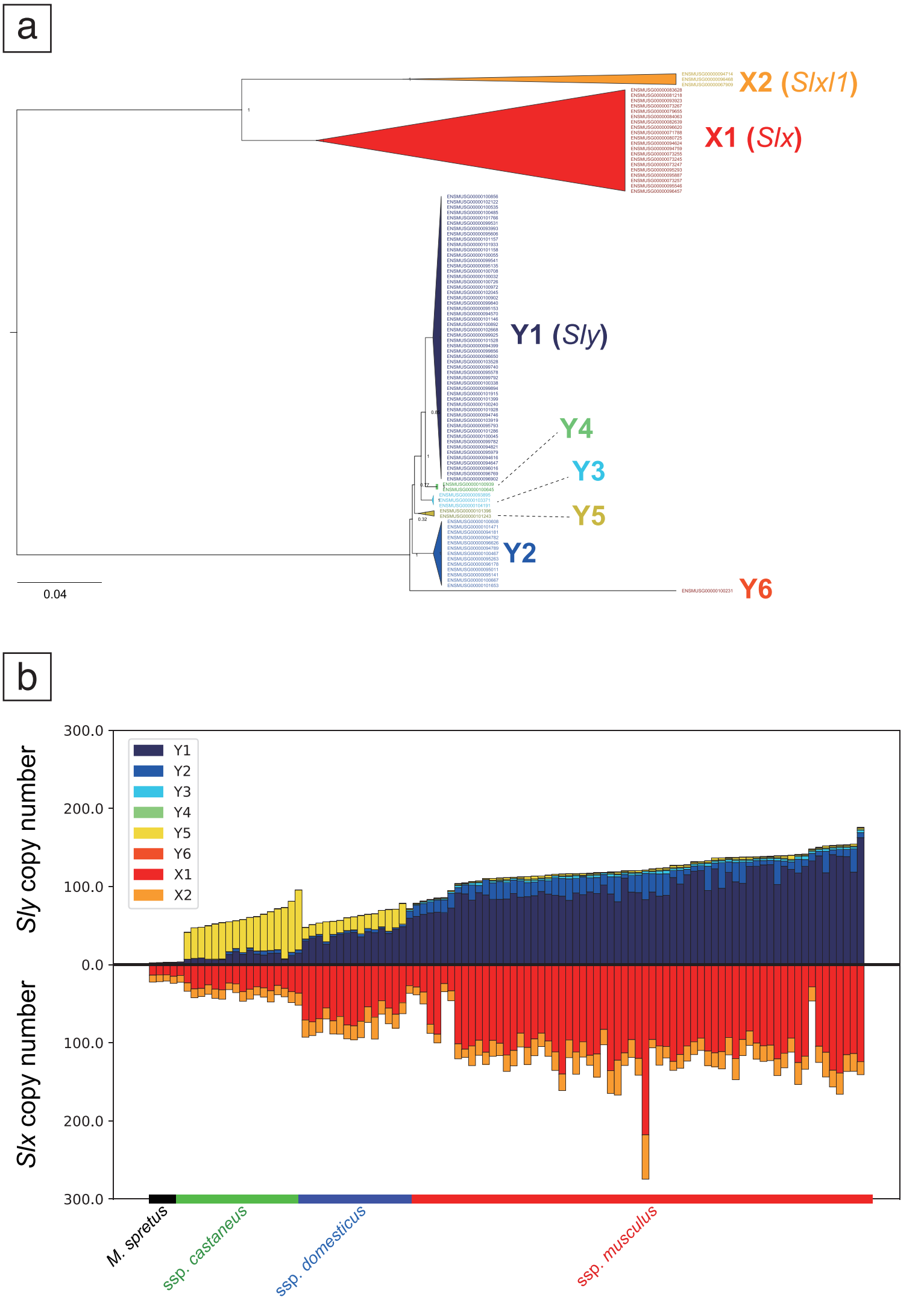
The upper panels show the segregation pattern of nonsynonymous SNVs. The sites in the red rectangles represent the focal sites that showed *castaneus*-ancestry enrichment in the Japanese (JPN) samples. The lower panels show the inferred genealogies around the focal sites. The branches with the derived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

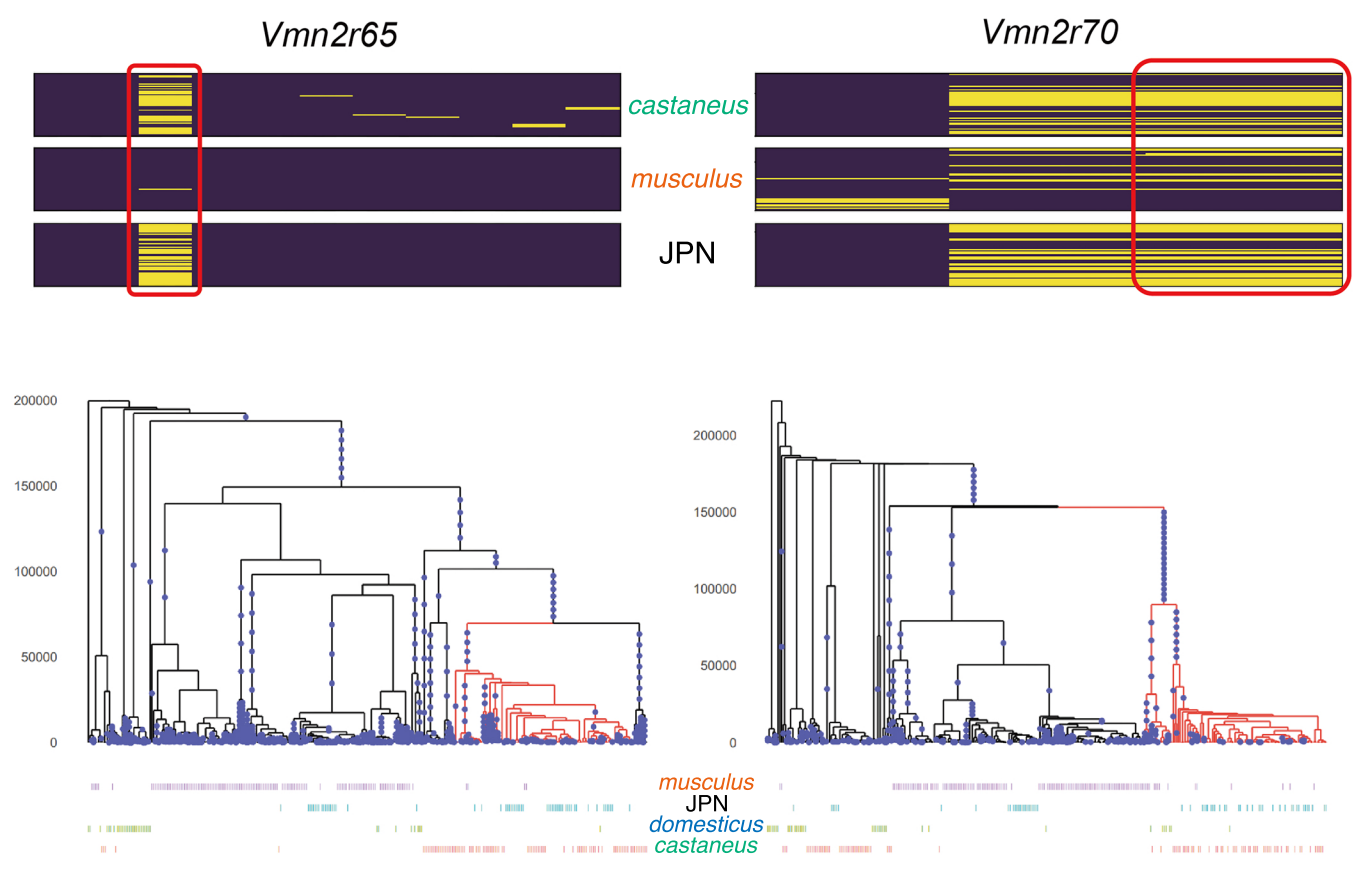
*Figure 5: Distribution of $F_{ST}$ in 20kb-length-windows on chromosome 7.*

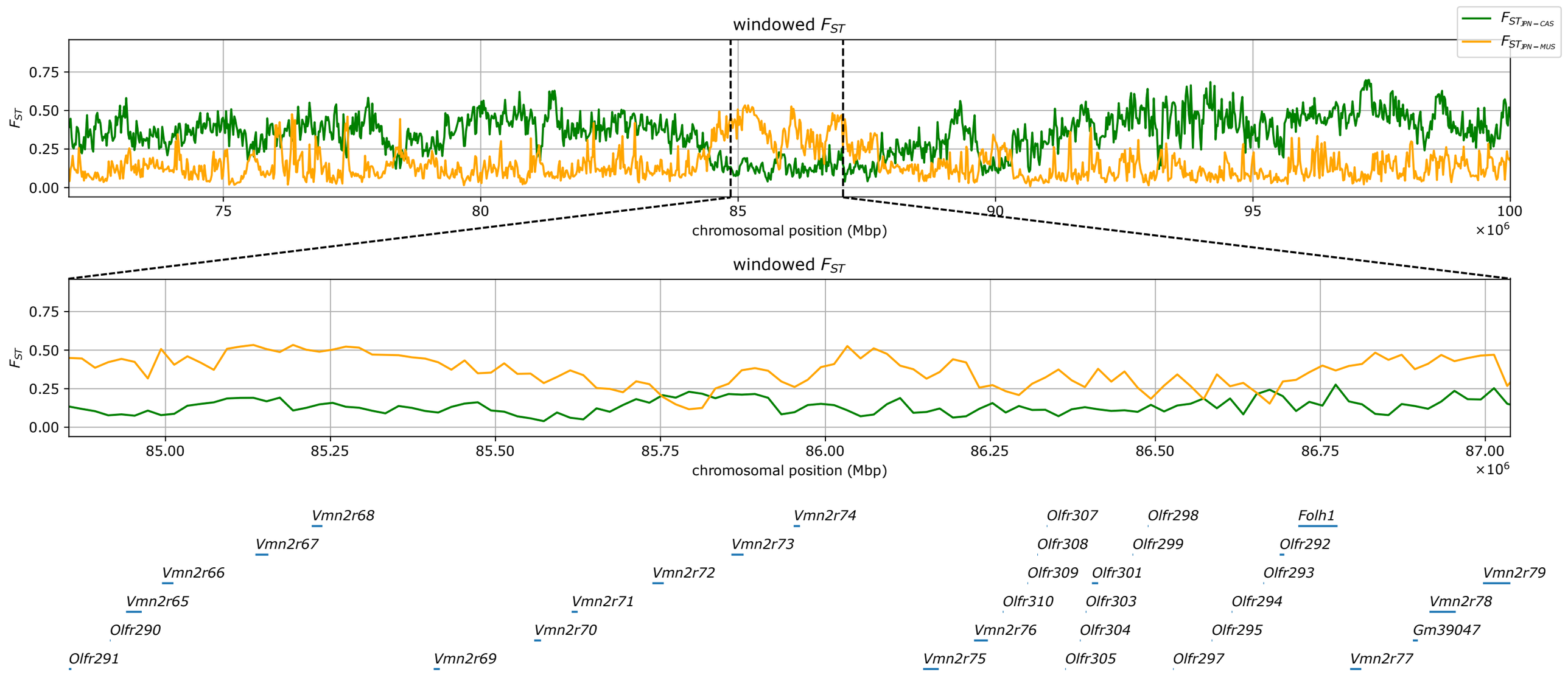
The yellow and green lines indicate $F_{ST}$ between the Japanese and subspecies *musculus* samples and $F_{ST}$ between the Japanese and subspecies *castaneus* samples, respectively. The lower panel shows the pattern in the olfactory/vomeronasal genes in this cluster (chr7:84,853,553-87,037,968, GRCm38).


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 ($\alpha$) 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 ($\alpha$) in Japan
0.564


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


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