Whole-genome sequencing analysis reveals the population history of *Mus musculus* in Madagascar

Kazumichi Fujiwara[1], Marie C Ranorosoa[2], Satoshi D Ohdachi[3,4], Satoru Arai[5], Yuki Sakuma[4], Hitoshi Suzuki[4], Naoki Osada[1,*]

[1]Graduate School of Information Science and Technology, Hokkaido University, Sapporo, 060-0814, Japan

[2]Ecole supérieur des Sciences Agronomiques ,Université d'Antananarivo, Antananarivo, 101, Madagascar

[3] Institute of Low Temperature Science, Hokkaido University, Sapporo, 060-0819, Japan

[4]Graduate School of Environmental Science, Hokkaido University, Sapporo, 060-0810, Japan

[5]Center for Surveillance, Immunization, and Epidemiologic Research, National Institute of Infectious Diseases, Toyama 1-23-1, Shinjuku, Tokyo 162-8640, Japan

*Correspondence to:

Naoki Osada, nosada@ist.hokudai.ac.jp

Word count for main text (excluding references, tables, and figures): 5,030 words

Abstract

In Madagascar, the house mouse (*Mus musculus*) is thought to have colonized along with humans and is now one of the most successfully colonized rodents on the island. In this study, we determined the whole-genome sequences of the Madagascar house mouse captured from the wild. We examined the evolutionary history of its population regarding the mitochondrial and autosomal genomes. We confirmed that in the mitochondrial genomes of Madagascar house mice, a monophyletic clade forms a basal origin within the species. An analysis of autosomal genomic sequences indicates that the Madagascar house mouse population is genetically a member of *M. m. castaneus* (CAS). It also contains genetic elements of *M. m. domesticus* (DOM) resulting from ancient hybridization. The signature of a strong population bottleneck 1000–3000 years ago was observed in the mitochondrial and autosomal genomic data. We also show that the divergence of the Madagascar population from the CAS population occurred approximately 50,000–99,000 years ago. Madagascar house mice show strong genetic affinity to many CAS samples across a wide range of Indian Ocean coastal regions. However, our results suggest that they would not have originated directly from the Indonesian islands, where Austronesian-speaking people in Madagascar originated. Because the ancient hybridization signature with DOM did not appear in the Indonesian and other CAS samples, we propose that Madagascar house mice were not directly brought by Austronesian-speaking people but came from somewhere around the Middle East or South Asia soon after the colonization of initial farmers.

Introduction

The house mouse (*Mus musculus*) originated from the northern part of the Indian subcontinent (Boursot *et al.*, 1996; Din *et al.*, 1996) and is a rodent that has exploded in population size and successfully colonized a wide range of continents, large islands, and even small remote islands around the world. Hypothetically, there are three major subspecies of house mice: the South Asian subspecies (*M. m. castaneus*: CAS), the North Eurasian subspecies (*M. m. musculus*: MUS), and the West European subspecies (*M. m. domesticus*: DOM). Rodents, such as mice and rats, are small commensal mammals that spread worldwide along with humans. Therefore, analyzing murine colonization patterns would provide insight into the history of human migration (Prager *et al.*, 1998; Duplantier *et al.*, 2002; Searle *et al.*, 2009; Suzuki *et al.*, 2013, 2015; Jing *et al.*, 2014; Phifer-Rixey *et al.*, 2018; Li *et al.*, 2021).

Madagascar is the fourth largest island in the world. It separated from the African continent around 160 million years ago and is about 300 km from the coast of the African continent at its shortest distance. The origin of humans living on the island of Madagascar, inferred from linguistic (Dahl, 1951, 1988; Adelaar, 1995), archaeological (Dewar and Wright, 1993; Burney *et al.*, 2004), and genetic (Hurles *et al.*, 2005) evidence, appears to be a dual origin from East Africa and Southeast Asia (in particular, Borneo Island). In addition, archaeological evidence suggests that the earliest human settlement occurred approximately 1500–2000 years ago (MacPhee and Burney, 1991; Burney *et al.*, 2004). However, recent archaeological excavations suggest a human presence on the island of Madagascar much earlier than previously thought (Dewar *et al.*, 2013; Hansford *et al.*, 2018; Douglass *et al.*, 2019) but it remains an open question. More recently, studies using the human genome showed that they migrated from Borneo approximately 2000–3000 years ago and from the east coast of Africa around 1500 years ago (Pierron *et al.*, 2017). Subsequent intermittent human settlement on the island of Madagascar led to forming a large human population on the entire island approximately 1000 years ago (Battistini and Verin, 1972; Burney *et al.*, 2004). At that time, a large commercial network

connecting Asia and the Mediterranean Sea had formed along the coastal regions of the Indian Ocean (Verin and Wright, 1999), and traders from the Arabian Peninsula also sailed to the coast of Africa, including Madagascar (Allibert, 1988; Liszkowski, 2000).

Rodents have successfully colonized the Madagascar island, and there are at least 23 species of rodents currently inhabiting it today (Rakotondravony and Randrianjafy, 1998). Of these, the black rat (*Rattus rattus*) is the most abundant rodent on the island (Goodman, 1995), whereas the house mouse is found in human settlements all over the island, though it is less abundant than the black rat. In a previous study analyzing limited mitochondrial control regions, the cytochrome *b* gene, and flanking tRNAs, Madagascar house mice were genetically similar to samples from Yemen, in the southern part of the Arabian Peninsula (Duplantier *et al.*, 2002; Sakuma *et al.*, 2016). The mtDNA analysis revealed that the Yemeni house mouse mtDNA lineage forms a cluster of another potential subspecies, *M. m. gentilulus* (GEN), which is distinct from the three major subspecies (Prager *et al.*, 1998; Suzuki *et al.*, 2013). The mtDNA lineage in Madagascar constitutes a “narrow” monophyletic group, suggesting a recent and probably single origin (Duplantier *et al.*, 2002; Sakuma *et al.*, 2016). Suzuki et al. (2013) further showed that the Madagascar-Yemen mitochondrial lineage was in the most basal place to *M. musculus*. In contrast, microsatellite studies suggested that the Madagascar mouse is CAS (Hardouin *et al.*, 2015), and the phenotype of the Madagascar house mouse shows that the tail is as long as the head and body (Rakotondravony and Randrianjafy, 1998), which is a characteristic of DOM (Orsini *et al.*, 1983; Marshall, 1998).

Previous studies on the genetic background of the house mouse in Madagascar are primarily based on mtDNA with limited nuclear genome analysis. This study performed high-quality whole-genome sequencing of five wild-caught *M. musculus* samples from the Madagascar islands. Although the whole-genome sequences of wild mice in European and Asian regions have been partly presented in previous studies (Harr *et al.*, 2016; Fujiwara *et al.*, 2021), there are still many undefined regions throughout the world. We determined the genetic position of these Madagascar wild house mice within

88 the global population of *M. musculus*. We estimated the divergence time resulting from the
89 colonization of these wild mice in Madagascar to determine their genetic backgrounds and population
90 dynamics. Our results shed light on the prehistoric and migration history of house mice and humans in
91 Madagascar.

Materials and Methods

Materials

Five specimens of the Madagascar wild house mouse were collected from the Parc Botanique et Zoologique de Tsimbazaza, Madagascar, which were also analyzed in the study by Sakuma et al. 2016 (Sakuma *et al.*, 2016). The other genomic sequence data of *M. musculus* were obtained from previous studies (Harr *et al.*, 2016; Fujiwara *et al.*, 2021). We used seven samples from the western Mediterranean mouse (*Mus spretus*) obtained from publicly available data (Harr *et al.*, 2016). These publicly available data were downloaded from the DDBJ (PRJDB11027) and the European Nucleotide Archive (PRJEB9450, PRJEB11742, PRJEB14167, PRJEB2176, and PRJEB11897), representing the data of Fujiwara et al. (2021) and Harr et al. (2016), respectively. The *M. spretus* samples were used for the outgroup population for downstream analysis. The complete mitochondrial sequence of *M. spretus* (NC_025952) was added for the mitochondrial lineage analysis. See Supplementary Table 1 for detailed information on the samples used in this study.

Mapping genomic reads and variant calling

For the five Madagascar wild house mouse samples, paired-end reads 150 bp in length were generated by using the DNBSEQ platform. The cleaned reads were quality checked by the FastQC tool (Andrews, 2010). The basic strategy for mapping genomic read pairs and single-nucleotide variant (SNV) calling was identical to that described in Fujiwara et al. (2021). The autosomal SNVs were filtered by Variant Quality Score Recalibration and Mappability unique scores (Pockrandt *et al.*, 2020) as described in Fujiwara et al. (2021). The mitochondrial genome sequences of the Madagascar mouse samples were reconstructed by a *de-novo* assembly software, Novoplasty v.4.2.1 (Dierckxsens *et al.*, 2017), using the house mouse mitochondrial genome (NC_005089) as a seed sequence. To distinguish the sex of the samples, we calculated the read coverages on the sex chromosomes. We used “depth” command from samtools to count the coverage of each sample in non-pseudoautosomal sites of the X and Y chromosomes that passed the mappability filter. The ratio of X to Y chromosome coverages

exhibited a clear bimodal distribution in which the modes were 1.03–1.09 and 148.24–268.16. Three out of five samples with the higher ratios were judged as male samples (Supplementary Table 2). The synonymous and nonsynonymous SNVs were annotated with the house mouse gene annotation data version GRCm38.101 (ftp://ftp.ensembl.org/pub/release-101/gtf/mus_musculus/) using the SnpEff and SnpSift programs (Cingolani, Patel, *et al.*, 2012; Cingolani, Platts, *et al.*, 2012). All Madagascar wild house mouse samples were examined for kinship using KING software (Manichaikul *et al.*, 2010) with the option "--kinship." None of the samples from Madagascar showed a kinship relationship. Therefore, a total of 133 samples of mice, five novel samples from Madagascar, and 128 publicly available samples, were used in this study.

Mitochondrial lineage analysis

For mitochondrial genome analysis, *M. spretus* was used for the outgroup. The mitochondrial sequences of all samples were aligned using MUSCLE (Edgar, 2004a, 2004b) implemented in MEGA7 (Kumar *et al.*, 2016), and all D-Loop regions and gapped sites were removed. To construct the maximum likelihood (ML) tree of the mitochondrial genome, we used IQ-TREE v.1.6.12 (Nguyen *et al.*, 2015). We also estimated the best substitution model using the ModelFinder (Kalyaanamoorthy *et al.*, 2017) function implemented in IQ-TREE and used the "TIM2+F+R3" model for our calculations. The bootstrapping values were also computed with 1000 replications for constructing ML trees. We used the program, BEAST v1.8.4 (Drummond and Rambaut, 2007) to estimate the time to most recent common ancestors (tMRCA) using mitogenome sequences (16,038 bp), the HKY+G substitution model, and the strict clock model as reported previously (Li *et al.*, 2021). The Markov chain Monte Carlo simulation was run for 10 million generations (burn-in 10%) and sampled every 10,000 generations. Tracer v1.6 software (Rambaut *et al.*, 2018) was used to assess the convergence of the chains and ensured effective sample size (ESS) values above 200 for most parameters. The trees were summarized using TreeAnnotator v1.8.4 software (http://beast.community/treeannotator) with the settings "Maximum clade credibility tree" and "Mean heights" and were displayed using FigTree

v1.4.3 software (http://tree.bio.ed.ac.uk/software/figtree/). The time-dependent evolutionary rates of mtDNA (Ho *et al.*, 2011) were considered as previously described (Li *et al.*, 2021). The evolutionary rates of 2.4×10^{-8} substitutions/site/year and 1.1×10^{-7} substitutions/site/year were used for the older (> 100,000 years) and younger divergences (< 20,000 years ago), respectively.

Population structure analysis of the Madagascar house mouse

We performed SNV filtering by using vcftools omitting indels and multiallelic SNVs for downstream analysis. The data for filtered SNVs, which contain only biallelic autosomal SNVs, were then converted to a PLINK (Purcell *et al.*, 2007) format. In general practice, the SNVs in linkage disequilibrium are excluded from the population structure analysis. However, these SNVs were not excluded in this study because the house mouse samples were highly structured at the subspecies level, and excluding these SNVs would result in a very small number of SNVs left for analysis.

Principal component analysis (PCA) was performed using the "smartpca" program implemented in Eigensoft (Patterson *et al.*, 2006). In order to calculate eigenvalues, the default parameter settings for smartpca were used, that we did not remove outlier samples.

The f_4-statistics were computed using the AdmixTools (Patterson *et al.*, 2012) qpDstat program with "f4 mode" and "printsd" options. The representative CAS and DOM populations were the least admixed samples of Indian CAS and German DOM, as chosen in the previous study (Fujiwara *et al.*, 2021). The configuration of the f_4-statistics is represented as f_4(A, B; C, D), where "A" to "D" represents each population. If the value of f_4(A, B; C, D) is not statistically significantly different from 0, the allele frequency difference between A and B would be independent on the allele frequency difference between C and D. In contrast, if the value of f_4(A, B; C, D) is statistically significantly positive, the gene flow may exist between A and C or B and D. Furthermore, if the value of f_4(A, B; C, D) is statistically significantly negative, the gene flow may exist between A and D or B and C. The outgroup f_3-statistics were computed by the qp3pop program of AdmixTools using the "outgroupmode" option. In the outgroup f_3-statistics scenario, the statistics of f_3(A, B; C), with target C as an outgroup

population, show the genetic drift from C to the common ancestor of A and B, and reveal the closeness between the population A and B.

A neighbor-joining tree was constructed for autosomes according to a pairwise distance matrix. Intermediate hybrid samples are excluded for computing distance matrix and constructing neighbor-joining tree, because they significantly change the topology of the phylogenetic tree. This sample exclusion condition is according to the method described by Fujiwara et al. (2021) (Fujiwara *et al.*, 2021). We computed all pairwise identity-by-state (IBS) distance matrices by using PRINK1.9 software using the "--distance square 1-ibs" option (Purcell *et al.*, 2007). To construct and visualize a neighbor-joining tree (Saitou and Nei, 1987) from the distance matrix, we used the "ape" package in R (Paradis *et al.*, 2004). The Neighbor-Net tree (Bryant and Moulton, 2004) was computed using gdsfmt, SeqArray, SNPRelate (Zheng *et al.*, 2012, 2017), and phangorn (Schliep, 2011) packages in R.

Demographic inference analysis

Pairwise sequentially Markovian coalescent (PSMC) analysis was performed for the samples (Li and Durbin, 2011). To obtain the consensus autosomal genome sequences of individual samples, the "mpileup" command from samtools was used with the "-C 50, -O, -D 60, -d 10" option as the input file for PSMC. The PSMC command option values (-t and -p) were considered according to the default settings suggested by the PSMC software. The time interval parameters were set to "4+25*2+4+6" with 25 iterations. The value, 0.57×10^{-8} per site per generation, were used for the mutation rate, and we set generation time as 1 year (Milholland *et al.*, 2017). To validate the variance of estimated population size, N_e, we performed 100 bootstrap replications for each representative subspecies sample.

The multiple sequentially Markovian coalescent (MSMC) (Schiffels and Durbin, 2014) and its second version (MSMC2: https://github.com/stschiff/msmc2) (Malaspinas *et al.*, 2016; Schiffels and Wang, 2020) were used to estimate N_e changes and population separation history. In our MSMC/MSMC2 analysis, we performed estimations using a phased haplotype sequence as input data.

For phasing, ShapeIt4 software (Delaneau *et al.*, 2019) was used to generate phased haplotype data. The "Mapping Data for G2F1 Based Coordinates" (Liu *et al.*, 2014) from "Mouse Map Converter (http://cgd.jax.org/mousemapconverter/)" was downloaded to provide the recombination rate input file for MSMC/MSMC2. Mappability of the reference mouse genome is considered, and only unique sequence positions were used in the calculations (Fujiwara *et al.*, 2021). To estimate the population separation history for CAS and Madagascar, or DOM and Madagascar, we used two haplotypes from each population (four haplotypes total) to calculate the relative cross coalescent rate (rCCR) for given population pairs. According to Shiffels et al. (2014) (Schiffels and Durbin, 2014), the rCCR variable ranges between 0 and 1 (in some cases it is unavoidable to calculate greater than 1), and the value close to 1 indicates that two populations were not differentiated. Heuristically, the half value of rCCR (i.e., rCCR = 0.5) is assumed as the time when the two populations separated. In the bootstrapping method of MSMC2, the original input data was cut into blocks of 5Mbp each, which were then randomly sampled to create a 3Gbp-length pseudo genome. Calculations were performed on a total of 20 of these pseudo genomes. For the Madagascar house mouse samples, we used all five individuals (ten haplotypes) for estimating the effective population size using MSMC to clarify the recent bottleneck event. To visualize the MSMC/MSMC2 data, we assumed a generation time of 1 year and a mutation rate of 0.57×10^{-8} per site per generation (Milholland *et al.*, 2017).

The demographic parameter estimation was performed using the composite likelihood method as implemented in FastSimCoal2 (Excoffier *et al.*, 2013). To reduce the effect of natural selection, exonic and flanking regions within 10 kb from exons were excluded from our analysis. The sites used in this analysis included a total of 469,631,319 sites. We measured joint minor allele frequencies of CAS-Madagascar and DOM-Madagascar populations and used the data for estimation. For each model, we simulated 100,000 genealogies and ran 40 expectation-maximization cycles. We repeated the above process 1000 times with different seed values and adopted the demographic parameters with the highest log-likelihood. Bootstrapping for FastSimCoal2 is demonstrated by 100 replicates of the

217 highest ML model parameters. The confidence interval for each demographic parameter is calculated
218 based on the bootstrapping results. We used the same values of mutation rate and generation time as
219 the MSMC/MSMC2 analysis.

Results

Genetic diversity of Madagascar Mus musculus

We analyzed five whole-genome sequences of *M. musculus* from the Madagascar island. The samples were collected from the Parc Botanique et Zoologique de Tsimbazaza, Madagascar, and nearby areas in 2015 (Sakuma *et al.*, 2016). We also used the 128 whole-genome sequences of *M. musculus* from previous studies (121 samples of *M. musculus*, 7 samples of *M. spretus*). Supplementary Table 1 contains a list of samples with sample ID and locations used in this study. The number of filtered SNVs in five Madagascar house mouse samples against the GRCm38 (C57BL/6J) reference genome was 24,883,782 in total with an overall transition/transversion ratio of 2.10. The average per-sample nucleotide diversity (heterozygosity) of the Madagascar sample was 0.0031. Table 1 presents the detailed metrics for each mouse sample.

Present population structure analysis

To define the genetic features of the Madagascar house mouse population, we first performed a PCA using autosomal 86,288,314 SNVs from 126 *M. musculus* samples (Figure 1). In the PCA plot, three clusters corresponding to MUS, CAS, and DOM genetic components, and individuals lie on the intermediate shows inter-subspecies hybrid samples. The PC1 (eigenvector 1) indicates the genetic difference between MUS and CAS, and the PC2 (eigenvector 2) indicates the genetic difference between the DOM and CAS. The five Madagascar samples positioned near the CAS cluster revealed a slightly admixed genetic component of the DOM. There was little variation in eigenvalues among the Madagascar mouse population. Next, we constructed an autosomal genetic tree using the neighbor-joining method with the *M. spretus* population set as the outgroup (Figure 2). The tree showed that the Madagascar population formed a single cluster within the CAS cluster. Neighboring samples of the Madagascar samples were the Nepalese samples and the Indian samples except for the mountainous areas. Although the Madagascar samples were close to the Indonesian samples in the PCA plot, the tree showed that the samples in the two populations belonged to different clusters within the CAS

clade. Furthermore, we constructed a Neighbor-Net network by using the genetic distances of all samples (Supplementary Figure 1). The Neighbor-Net network showed similar results to the neighbor-joining tree, showing that the Madagascar wild house mouse samples are within the CAS samples clusters. Fujiwara et al. (2021) reported that Indian and German samples experienced the smallest amount of gene flow among subspecies and would be the representative populations for CAS and DOM, respectively. Therefore, samples from these populations were used to calculate the f_4-statistics for the Madagascar population. To detect gene flow of the Madagascar population with CAS or DOM, we used 102,858,288 SNVs to compute f_4(SPR, CAS; DOM, MDG) and f_4(SPR, DOM; CAS, MDG), where MDG and SPR represent the Madagascar and *M. spretus* population, respectively. In both cases, the f_4-statistics scores indicate significantly positive values. The Z-score was 52.43 for f_4(SPR, CAS; DOM, MDG) and 26.25 for f_4(SPR, DOM; CAS, MDG), suggesting that the Madagascar population was genetically closer to CAS than to DOM, with a non-negligible amount of gene flow with DOM. Supplementary Table 3 presents the detailed results. To get a clearer picture of how the mouse population in Madagascar has propagated throughout history, we used the outgroup f_3 statistics, f_3(MDG, X; SPR), where X represents non-Madagascar samples. The results showed that individuals genetically close to the Madagascar population (high f_3 values) were located around the coastal side of the South Asian regions and Indonesian islands but not specifically genetically close to samples from Nepal or Indonesia (Figure 3).

Mitochondrial lineage analysis

In addition to the above results inferred using autosomal data, we clarified the genetic background of Madagascar samples using mitochondrial genomes. Phylogenetic trees based on maximum likelihood of inference using the complete mitochondrial genome revealed that the five individuals from Madagascar formed a distinct cluster (Supplementary Figure 2). The Madagascar samples were found in the most basal cluster of all *M. musculus* mitochondrial haplotypes, but a bootstrap support was not sufficiently high (0.64). The tMRCA of the mitochondrial genome of the

Madagascar population was estimated using BEAST software. This analysis revealed that the Madagascar population expanded about 3,000 years ago (Figure 4b).

Inference of past demographic history

PSMC analysis was performed to estimate the demographic pattern of the Madagascar mouse population, including changes in the effective population size. All mouse samples from Madagascar experienced similar effective population size changes over the past 1,000,000 years (Figure 5a). It is clear that the Madagascar population experienced a rapid increase in effective population size from 200,000 to 300,000 years ago, peaked around 100,000 years ago, and has been slowly declining since then over the Last Glacial Period. The PSMC plots for the representative individuals of CAS and DOM are also shown in Figure 5a. According to the PSMC plot, the DOM, CAS, and Madagascar populations experienced similar population histories until about 400,000–500,000 years ago. Then, the CAS and Madagascar populations experienced different population histories from the DOM, and they experienced similar population histories until about 100,000 years ago. The CAS population has experienced a rapid decrease in effective population size afterward. The Madagascar population has also experienced a similar decrease, but slower rate than the CAS population.

To reconstruct the CAS-Madagascar and DOM-Madagascar population divergence history, we applied two methods: MSMC and FastSimCoal2. For MSMC analysis, four haplotypes (two haplotypes from each) were used for the CAS-Madagascar and DOM-Madagascar analyses. Figure 5b shows the population separation histories (rCCR) of CAS-Madagascar and DOM-Madagascar. The timing of rCCR = 0.5 between CAS and Madagascar was approximately 50,000 years ago (95% Confidence Interval (95% CI): 48,907–50,011) and between DOM and Madagascar was approximately 225,000 years ago (95% CI: 224,641–226,546). Notably, the DOM-Madagascar rCCR plot in Figure 5b shows a slight increase of rCCR from about 60,000–100,000 years ago. In order to confirm whether this slight increase is unique to the Madagascar population, we calculated the rCCR between the DOM and Indonesian populations, where the Austronesian-speaking people were supposed to migrated from.

We found that DOM-Indonesia did not exhibit any of these increases (Figure 5c). Furthermore, the rCCR of DOM-CAS (CAS from least admixture sample of India) was computed as a control. The rCCR of DOM-CAS did not show any specific increase between 60,000–100,000 years ago and showed almost the same pattern as the rCCR of DOM-Indonesia.

We also performed MSMC analysis using 10 haplotypes from the Madagascar population to analyze the relatively recent effective population size (Figure 6). The plot exhibits a historical bottleneck event of the Madagascar population around 1,000–3,000 years ago.

We further performed coalescent simulations using FastSimCoal2 with the joint allele frequency spectrum data. The FastSimCoal2 simulations enabled us to estimate the population separation time of CAS-Madagascar or DOM-Madagascar, assuming a constant migration rate over time between the populations after the population split. The results showed the divergence between CAS and Madagascar was about 99,000 years ago (95% CL: 98,102–100,417), and the divergence between DOM and Madagascar was approximately 114,000 years ago (95% CL: 112,941–115,220). Table 2 presents the detailed estimated demographic parameters.

Discussion

Genetic background of the Madagascar wild house mouse population

Previous studies have shown that the wild *M. musculus* exhibit a very large effective population size and nucleotide diversity compared to humans. The Madagascar house mice analyzed in this study also showed a higher nucleotide diversity of 0.31% compared with humans (0.08%–0.12%); however, this value is less than half of the nucleotide diversity in CAS (0.74%–0.79%), slightly higher than the nucleotide diversity in MUS (0.18%–0.25%), and within the range of the nucleotide diversity in DOM (0.07%–0.35%). The relatively low nucleotide diversity within the *M. musculus* population may be related to the fact that the Madagascar house mouse population experienced a bottleneck when they were introduced onto the island.

Based on the PCA plot, f_4-statistics test, and neighbor-joining phylogenetic trees, we conclude that the sampled Madagascar population genetically belongs to CAS as a subspecies. This result is consistent with and supports a previous microsatellite study (Hardouin *et al.*, 2015). However, PCA, f_4-statistic test, and MSMC analysis demonstrated that the Madagascar population has experienced ancient admixture events with the DOM subspecies. The recent quantitative trait loci study of genes associated with hybrid fitness of *M. musculus* identified regions causing hybrid male sterility in CAS and DOM hybrids on autosomes, the X chromosome, and the pseudoautosomal region (PAR) of sex chromosomes (White *et al.*, 2012). According to the study conducted by White et al., a fairly high percentage of male individuals in the hybrid second generation (F_2), exhibited phenotypes associated with infertility, indicating that CAS and DOM hybrids have difficulty in constructing long-lasting hybrids. Nevertheless, the Madagascar population exhibits the DOM-like CAS genetic feature, indicating that inter-subspecific hybridization would be rare but possible, or hybrid incompatibility was not so strong around 60,000–100,000 years ago.

The mitochondrial genome analysis indicates that the mitochondrial lineages of the Madagascar population consist of a monophyletic group with low nucleotide diversity and that these

lineages have a recent and unique origin. MSMC analysis using 10 haplotypes from five samples of the Madagascar population showed that the Madagascar population experienced a bottleneck event about 1000–3000 years ago. This result is consistent with the mitochondrial analysis and suggests that *M. musculus* may have entered Madagascar during this period. These pattern supports the previous mitochondrial DNA study by Duplantier et al. (Duplantier *et al.*, 2002; Sakuma *et al.*, 2016). However, the study by Duplantier et al. (2002), which analyzed 539 nucleotide sites in D-Loop of the mitochondrial genome, differs from our results in that the MUS is located at the basal in the phylogenetic tree. We created a phylogenetic tree with approximately 16,000 bp and showed the mitochondrial genealogy with higher resolution than previous analyses. However, it should be noted that the bootstrap value supporting the basal position of Madagascar mitochondrial lineage was 0.64, which is not sufficiently high to draw a conclusion.

The MSMC results indicate that the Madagascar and CAS populations diverged about 50,000 years ago, whereas the DOM population diverged as far back as 225,000 years ago. However, there was a small but clear increase of rCCR between the DOM and Madagascar populations around 60,000–100,000 years ago. This pattern may represent an admixture event between DOM and CAS in the ancestral population of the Madagascar house mice. To ensure that the admixture event was unique to the Madagascar wild house mouse population, we computed and compared rCCR for DOM-Indonesia. We selected the Indonesian sample for comparison because humans from Indonesian Borneo Island are considered one of the ancestral origins of the Malagasy people. We thought if DOM-Indonesia showed the same shape of rCCR as DOM-Madagascar, the Madagascar wild house mouse population might be of Indonesian origin, and the increase in rCCR around 60,000–100,000 years ago would be shared by the Indonesian CAS. However, according to the rCCR plot in Figure 5c, the DOM-Indonesia exhibited a different rCCR decline than DOM-Madagascar.

FastSimCoal2 analysis showed that the divergence from CAS was about 99,000 years and that from DOM was about 114,000 years, indicating that, as in MSMC, the divergence from the CAS

359 population occurred relatively recently compared with that for the DOM population. However, when
360 the MSMC results were compared with the FastSimCoal2 results, there was a difference in the
361 estimated divergence time. The MSMC model does not assume any specific demographic models, but
362 the FastSimCoal2 model assumes a continuous gene flow rate between the two populations. The
363 different assumptions may cause a discrepancy between the results of MSMC and FastSimCoal2. In
364 any case, it is safe to assume that the CAS and Madagascar populations likely diverged in the range of
365 50,000–99,000 years and the DOM and Madagascar populations diverged in the range of 115,000–
366 225,000 years.

367 *Possible migration history of the Madagascar wild house mouse population*

368 Since the human population in Madagascar seems to be dual origin from the Indonesian
369 Borneo Island in Southeast Asia and the east coast of the African continent, the house mice were
370 probably introduced to Madagascar island associated with the migration of humans.

371 First, we consider when the ancestors of the Madagascar house mice entered the island.
372 Although the molecular evolution rate of mitochondria, the mutation rate in nuclear genomes, and the
373 generation time of mice are debatable issues and estimated time would be highly vulnerable with these
374 assumptions, our mitochondrial genealogy and MSMC analysis consistently showed that they
375 experienced a strong population bottleneck around 1000–3000 years ago. Although the estimated time
376 range is wide, it is equivalent to the commonly accepted timing of migration when Austronesian-
377 speaking people from the Indonesian islands arrived in Madagascar (BCE 300–CE 500) (Dewar and
378 Wright, 1993; Burney *et al.*, 2004), as well as the estimated timing using human genetic data (Hurles
379 *et al.*, 2005). Since commensal animals, such as house mice, would not suddenly expand their
380 population size without a proper type of human activity, such as agriculture, we suggest that the
381 Madagascar house mice migrated to the island simultaneously or soon after the first Austronesian-
382 speaking farmers arrived on the island.

383 Second, where did the Madagascar house mice originate, and did the Austronesian-speaking

people bring house mice from their homeland? Our various analyses consistently show that the Madagascar samples belong to CAS, but were not specifically close to Indonesian samples. In particular, the Madagascar population showed a particular signature containing an ancient admixture with DOM around 60,000–100,000 years ago (Figure 5c), which is not observed in the Indonesian samples. In addition, based on the analysis of f_3 statistics, they instead show affinity to a wide range of CAS samples from the coastal regions of the Indian Ocean, such as present-day Bangladesh, India, and Sri Lanka. These broad affinities may represent more recent admixture events through trading across the Indian Ocean.

Based on the above evidence, we propose the following two hypotheses for the history of introducing the wild house mouse to Madagascar island. Based on our results supporting a non-Indonesian origin of the Madagascar population, the Madagascar wild house mouse may have been brought to the island by Austronesian-speaking people who passed through the coastal Indian Ocean route to Madagascar. Although ancient migrants could directly travel from Indonesian islands to Madagascar with their nautical technology, it does not necessarily exclude the possibility that they temporarily stayed in coastal areas and picked up local mice of which the mitochondrial lineage is currently extinct. The second possibility is that an ancient Madagascar house mouse population migrated from the neighboring coastal areas due to trade or other exchanges shortly after the first migration of the Austronesian-speaking people. Since there are many commensal animals prevalent on the island, we do not necessarily assume that all of them came to the island accompanied with the Austronesian-speaking people. Because the ancient admixture between DOM and CAS did not likely occur in Southeast Asia, they may have come from somewhere around the South Asia or Middle East.

Based on current evidence, there is no genetic or archaeological trace of a human migration route to Madagascar island via South Asian coastal line, we consider the latter hypothesis, that the Madagascar house mice came from the South Asia or Middle East, is more likely. The previous study by Sakuma et al. (2016) showed that the mitochondrial genome of the Madagascar sample is

409 genetically close to the Yemen samples (Sakuma *et al.*, 2016), suggesting that the Madagascar house
410 mice might have originated from a region around the Middle East. In that case, because Yemen is not
411 in the distribution range of typical CAS, it is possible that the distribution of the house mice subspecies
412 in the Middle East approximately 1000–3000 years ago would be different from the current one, and
413 the direct ancestor of the Madagascar house mice may have gone extinct on the continent.
414 Unfortunately, our samples did not fully cover these regions, and further studies will clarify the detailed
415 history of the Madagascar house mice and its relationship with human activity.
416

Conclusion

This study determined the whole-genome sequences of five Madagascar wild house mouse populations and analyzed their genetic backgrounds. Until today, mice from Madagascar have been treated as DOM or *M. m. gentilulus*, but genomic evidence indicates that their genetic feature belongs to CAS. However, we observed fragments of DOM that may have been admixed 60,000–100,000 years ago. In addition, mitochondrial and nuclear genome evidence indicates that they experienced a bottleneck approximately 1000–3000 years ago, coincides with the arrival of Austronesian humans to Madagascar. However, the genetic relationship between the Madagascar and Indonesian samples does not appear to be very close. Based on the results of this study, we propose that the mice arrived in Madagascar through trade right after the first Austronesians arrived on Madagascar. In the future, mouse samples from the Middle East and Borneo will be needed better to define the genetic backgrounds of the Madagascar mouse population. Overall, this study provides insight into the rodents that have successfully migrated to Madagascar, which has only been partially understood.

Data Availability

The short-read whole-genome sequencing data generated in this study have been submitted to the DDBJ BioProject database (https://www.ddbj.nig.ac.jp/bioproject/) under the accession number PRJDB11969. The complete mitochondrial genome sequences are submitted to DDBJ database under the accession number LC644158–LC644162.

Competing Interest Statement

The authors declare no competing interests.

Acknowledgements

This work was partly supported by the MEXT KAKENHI (grant 18H05511 to N. O.). We would like to express my gratitude to Dr. Chihiro Tanaka, who belongs to Animal Care and Exhibition Section, Yagiyama Zoological Park, Sendai, Japan, for her support of the sample collection project. We would also like to express my deepest gratitude to Drs. Kimiyuki Tsuchiya and Hajanirina Ramino for sample collection in Madagascar.

Figure Legends

Figure 1

Principal Component Analysis plot of *M. musculus* using autosomal single nucleotide variants (SNVs). The points in the figure represent the eigenvalues of individual samples collected from the countries shown in the right. The upper left, lower left, and right vertices of triangle represent *Mus musculus domesticus* (DOM), *M. m. castaneus* (CAS), and *M. m. musculus* (MUS) genetic components, respectively. The proportion of variance for each eigenvalue is shown in parentheses on the labels of *x*-axis and *y*-axis. The arrow shows the position of Madagascar population.

Figure 2

The neighbor-joining tree inferred using pairwise genetic distances of *M. musculus.* The seven samples of *M. spretus* (SPR) were used as the outgroups. The color shows the subspecies: red (*M. m. musculus*: MUS), green (*M. m. castaneus*: CAS), blue (*M. m. domesticus*: DOM), and cyan (*M. musculus* - Madagascar). The hybrid samples were excluded for constructing distance tree.

Figure 3

The f_3 statistics test results assuming an outgroup f_3 (MDG, X; SPR) scenario. The right plot shows the top 20 of the outgroup f_3 statistics values in which individuals of regional origin are genetically close to the Madagascar wild house mouse population. The horizontal line associated with each point indicates the standard error. The higher values indicate genetic closeness to the Madagascar house mouse population. The left world map shows the f_3 statistics value heat map focused on the South Asian region.

Figure 4

Divergence time estimates (million years ago) of the five phylogroups of the *M. musculus* subspecies based on entire mitochondrial genome sequences (15,181 bp) and a Bayesian-relaxed molecular clock of 2.4×10^{-8} substitutions/site/year (**a**). Blue bars represent the 95% highest posterior density interval. See Li et al. (2021) for the details of the subclades. Given the time-dependency of the mtDNA

evolutionary rate, the mean time of the most recent common ancestor of the five haplotypes from Madagascar was estimated with a molecular clock of 1.1×10^{-7} substitutions/site/year (**b**).

Figure 5

The Pairwise Sequentially Markovian Coalescent (PSMC) plot and the Multiple Sequentially Markovian Coalescent (MSMC) plot of the inter-population divergence of *M. musculus*. The green, blue, and black lines represent the inferred effective population size transition of CAS, DOM, and Madagascar (MDG) mouse populations, respectively (**a**). The *x*-axis represents generations before the present scaled by the mutation rate 0.57×10^{-8} per site per generation and the y-axis represents the number of the inferred effective population size. The lines with lighter colors represent 100 replications of the bootstrapping results. The figure shows the relative cross coalescent rate (rCCR) between the CAS-MDG and DOM-MDG inter-populations (**b**). The 20 replications bootstrapping results are represented by multicolor. Comparative plot of rCCR for DOM-CAS (MDG) and DOM-CAS (Indonesia, IDN) scenarios with DOM-CAS (India, IND) as a control. (**c**).

Figure 6

The Multiple Sequentially Markovian Coalescent plot of five Madagascar samples using 10 haplotypes. The *x*-axis represents generations before the present scaled by a mutation rate of 0.57×10^{-8} per site per generation. The *y*-axis represents the number of the inferred effective population size of Madagascar house mice. The MSMC plot with multiple haplotypes can demonstrate the recent inferred effective population size.

Reference

Adelaar A (1995). Asian roots of the Malagasy; A linguistic perspective. Bijdragen Tot De Taal- Land- En Volkenkunde / J Humanit Soc Sci Southeast Asia 151: 325–356.

Allibert C (1988). Les contacts entre l'Arabie, le Golfe persique, l'Afrique orientale et Madagascar: confrontation des documents écrits, des traditions orales et des données archéologiques récentes. MOM Éditions 16: 111–126.

Andrews S (2010). FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Battistini R, Verin P (1972). Man and the environment in Madagascar. In: Monographiae Biologicae. Springer Netherlands: Dordrecht, p 311.

Boursot P, Din W, Anand R, Darviche D, Dod B, Deimling FV, et al. (1996). Origin and radiation of the house mouse: mitochondrial DNA phylogeny. J Evolution Biol 9: 391–415.

Bryant D, Moulton V (2004). Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21: 255–265.

Burney DA, Burney LP, Godfrey LR, Jungers WL, Goodman SM, Wright HT, et al. (2004). A chronology for late prehistoric Madagascar. J Hum Evol 47: 25–63.

Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. (2012). Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. Frontiers Genetics 3: 35.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6: 80–92.

Dahl OC (1951). Malgache et maanjan: une comparaison linguistique. Egede Instituttet: Oslo.

Dahl OC (1988). Bantu substratum in Malagasy. In: Linguistique de Madagascar et des Comores, Etudes Océan indien. Vol 9, pp 91–132.

Delaneau O, Zagury J-F, Robinson MR, Marchini JL, Dermitzakis ET (2019). Accurate, scalable and integrative haplotype estimation. Nat Commun 10: 5436.

515 Dewar RE, Radimilahy C, Wright HT, Jacobs Z, Kelly GO, Berna F (2013). Stone tools and foraging
516 in northern Madagascar challenge Holocene extinction models. Proc National Acad Sci 110:
517 12583–12588.

518 Dewar RE, Wright HT (1993). The culture history of Madagascar. J World Prehist 7: 417–466.

519 Dierckxsens N, Mardulyn P, Smits G (2017). NOVOPlasty: de novo assembly of organelle genomes
520 from whole genome data. Nucleic Acids Res 45: e18–e18.

521 Din W, Anand R, Boursot P, Darviche D, Dod B, Jouvin-Marche E, et al. (1996). Origin and
522 radiation of the house mouse: clues from nuclear genes. J Evolution Biol 9: 519–539.

523 Douglass K, Hixon S, Wright HT, Godfrey LR, Crowley BE, Manjakahery B, et al. (2019). A critical
524 review of radiocarbon dates clarifies the human settlement of Madagascar. Quaternary Sci Rev
525 221: 105878.

526 Drummond AJ, Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees.
527 BMC Evol Biol 7: 214.

528 Duplantier J-M, Orth A, Catalan J, Bonhomme F (2002). Evidence for a mitochondrial lineage
529 originating from the Arabian peninsula in the Madagascar house mouse (*Mus musculus*). Heredity
530 89: 154–158.

531 Edgar RC (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space
532 complexity. BMC Bioinformatics 5: 113.

533 Edgar RC (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput.
534 Nucleic Acids Res 32: 1792–1797.

535 Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013). Robust demographic
536 inference from genomic and SNP data. Plos Genet 9: e1003905.

537 Fujiwara K, Kawai Y, Moriwaki K, Takada T, Shiroishi T, Saitou N, et al. (2021). Insights into *Mus*
538 *musculus* subspecies population structure across Eurasia revealed by whole-genome sequence
539 analysis. bioRxiv: 2021.02.05.429881.

540 Goodman SM (1995). Rattus on Madagascar and the dilemma of protecting the endemic rodent
541 fauna. Conserv Biol 9: 450–453.

542 Hansford J, Wright PC, Rasoamiaramanana A, Pérez VR, Godfrey LR, Errickson D, et al. (2018).
543 Early Holocene human presence in Madagascar evidenced by exploitation of avian megafauna.
544 Sci Adv 4: eaat6925.

545 Hardouin EA, Orth A, Teschke M, Darvish J, Tautz D, Bonhomme F (2015). Eurasian house mouse
546 (*Mus musculus* L.) differentiation at microsatellite loci identifies the Iranian plateau as a
547 phylogeographic hotspot. BMC Evol Biol 15: 26.

548 Harr B, Karakoc E, Neme R, Teschke M, Pfeifle C, Pezer Ž, et al. (2016). Genomic resources for
549 wild populations of the house mouse, *Mus musculus* and its close relative *Mus spretus*. Scientific
550 Data 3:160075.

551 Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, et al. (2011). Time-
552 dependent rates of molecular evolution. Mol Ecol 20: 3087–3101.

553 Hurles ME, Sykes BC, Jobling MA, Forster P (2005). The dual origin of the Malagasy in island
554 Southeast Asia and East Africa: evidence from maternal and paternal lineages. Am J Hum
555 Genetics 76: 894–901.

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

559 Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A von, Jermiin LS (2017). ModelFinder: fast
560 model selection for accurate phylogenetic estimates. Nat Methods 14: 587–589.

561 Kumar S, Stecher G, Tamura K (2016). MEGA7: molecular evolutionary genetics analysis version
562 7.0 for bigger datasets. Mol Biol Evol 33: 1870–1874.

563 Li H, Durbin R (2011). Inference of human population history from individual whole-genome
564 sequences. Nature 475: 493–496.

565 Li Y, Fujiwara K, Osada N, Kawai Y, Takada T, Kryukov AP, et al. (2021). House mouse *Mus*
566 *musculus* dispersal in East Eurasia inferred from 98 newly determined complete mitochondrial
567 genome sequences. Heredity 126: 132–147.

568 Liszkowski HD (2000). Mayotte et les Comores: escales sur la route des Indes aux XVe et XVIIIe
569 siècles. Editions du Baobab: Mamoudzou, Mayotte, France.

570 Liu EY, Morgan AP, Chesler EJ, Wang W, Churchill GA, Villena FP-M de (2014). High-resolution
571 sex-specific linkage maps of the mouse reveal polarized distribution of crossovers in male
572 germline. Genetics 197: 91–106.

573 MacPhee RDE, Burney DA (1991). Dating of modified femora of extinct dwarf Hippopotamus from
574 Southern Madagascar: Implications for constraining human colonization and vertebrate extinction
575 events. J Archaeol Sci 18: 695–706.

576 Malaspinas A-S, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, et al. (2016). A genomic
577 history of Aboriginal Australia. Nature 538: 207–214.

578 Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M (2010). Robust relationship
579 inference in genome-wide association studies. Bioinform Oxf Engl 26: 2867–73.

580 Marshall J (1998). Identification and scientific names of Eurasian house mice and their European
581 allies, subgenus *Mus* (Rodentia, Muridae). National Museum of Natural History, Washington.

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

584 Nguyen L-T, Schmidt HA, Haeseler A von, Minh BQ (2015). IQ-TREE: A fast and effective
585 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32: 268–
586 274.

587 Orsini P, Bonhomme F, Britton-Davidian J, Croset H, Guerasimov S, Thaler L (1983). Le complexe
588 d'espèces du genre *Mus* en Europe centrale et orientale. II Critères d'identification, répartition,
589 caractéristiques écologiques. Z f Säugetierkunde 48: 86–99.

590 Paradis E, Claude J, Strimmer K (2004). APE: analyses of phylogenetics and evolution in R
591 language. Bioinformatics 20: 289–290.

592 Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. (2012). Ancient admixture in
593 human history. Genetics 192: 1065–1093.

594 Patterson N, Price AL, Reich D (2006). Population structure and eigenanalysis. Plos Genet 2: e190.

595 Phifer-Rixey M, Bi K, Ferris KG, Sheehan MJ, Lin D, Mack KL, et al. (2018). The genomic basis of
596 environmental adaptation in house mice. PLoS Genet 14: e1007672.

Pierron D, Heiske M, Razafindrazaka H, Rakoto I, Rabetokotany N, Ravololomanga B, et al. (2017). Genomic landscape of human diversity across Madagascar. Proc National Acad Sci 114: E6498–E6506.

Pockrandt C, Alzamel M, Iliopoulos CS, Reinert K (2020). GenMap: Ultra-fast computation of genome mappability. Bioinformatics 36: 3687–3692.

Prager EM, Orrego C, Sage RD (1998). Genetic variation and phylogeography of Central Asian and other house mice, including a major new mitochondrial lineage in Yemen. Genetics 150: 835–861.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genetics 81: 559–575.

Rakotondravony A, Randrianjafy R (1998). Espèces introduites et endémiques de Madagascar: description, répartition, biologie et écologie. In: Zehrer W, Rafanomezana S (eds) Proceedings of Symposium “Rongeurs et lutte antimurine à Madagascar”, Tamatave, 6–9 Décembre 1994, DPV/ GTZ pub: Antananarivo, Madagascar, pp 51–76.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biol 67: 901–904.

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–25.

Sakuma Y, Ranorosoa MC, Kinoshita G, Shimoji H, Tsuchiya K, Ohdachi SD, et al. (2016). Variation in the coat-color-controlling genes, *Mc1r* and *Asip*, in the house mouse *Mus musculus* from Madagascar. Mamm Study 41: 131–140.

Schiffels S, Durbin R (2014). Inferring human population size and separation history from multiple genome sequences. Nat Genet 46: 919–925.

Schiffels S, Wang K (2020). MSMC and MSMC2: The multiple sequentially markovian coalescent. In: Dutheil JY (ed) Statistical Population Genomics, Methods in Molecular Biology. Humana: New York, NY, pp 147–166.

Schliep KP (2011). phangorn: phylogenetic analysis in R. Bioinformatics 27: 592–593.

625 Searle JB, Jamieson PM, Gndz slam, Stevens MI, Jones EP, Gemmill CEC, et al. (2009). The diverse
626 origins of New Zealand house mice. Proc Royal Soc B Biological Sci 276: 209–217.

627 Suzuki H, Nunome M, Kinoshita G, Aplin KP, Vogel P, Kryukov AP, et al. (2013). Evolutionary and
628 dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic
629 sampling of mitochondrial DNA. Heredity 111: 375–390.

630 Suzuki H, Yakimenko LV, Usuda D, Frisman LV (2015). Tracing the eastward dispersal of the
631 house mouse, *Mus musculus*. Genes Environ 37: 20.

632 Verin P, Wright H (1999). Madagascar and Indonesia: new evidence from archaeology and
633 linguistics. Bull Indo Pac Pre Hi 18.

634 White MA, Stubbings M, Dumont BL, Payseur BA (2012). Genetics and evolution of hybrid male
635 sterility in house mice. Genetics 191: 917–934.

636 Zheng X, Gogarten SM, Lawrence M, Stilp A, Conomos MP, Weir BS, et al. (2017). SeqArray—a
637 storage-efficient high-performance data format for WGS variant calls. Bioinformatics 33: 2251–
638 2257.

639 Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012). A high-performance
640 computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics
641 28: 3326–3328.

642

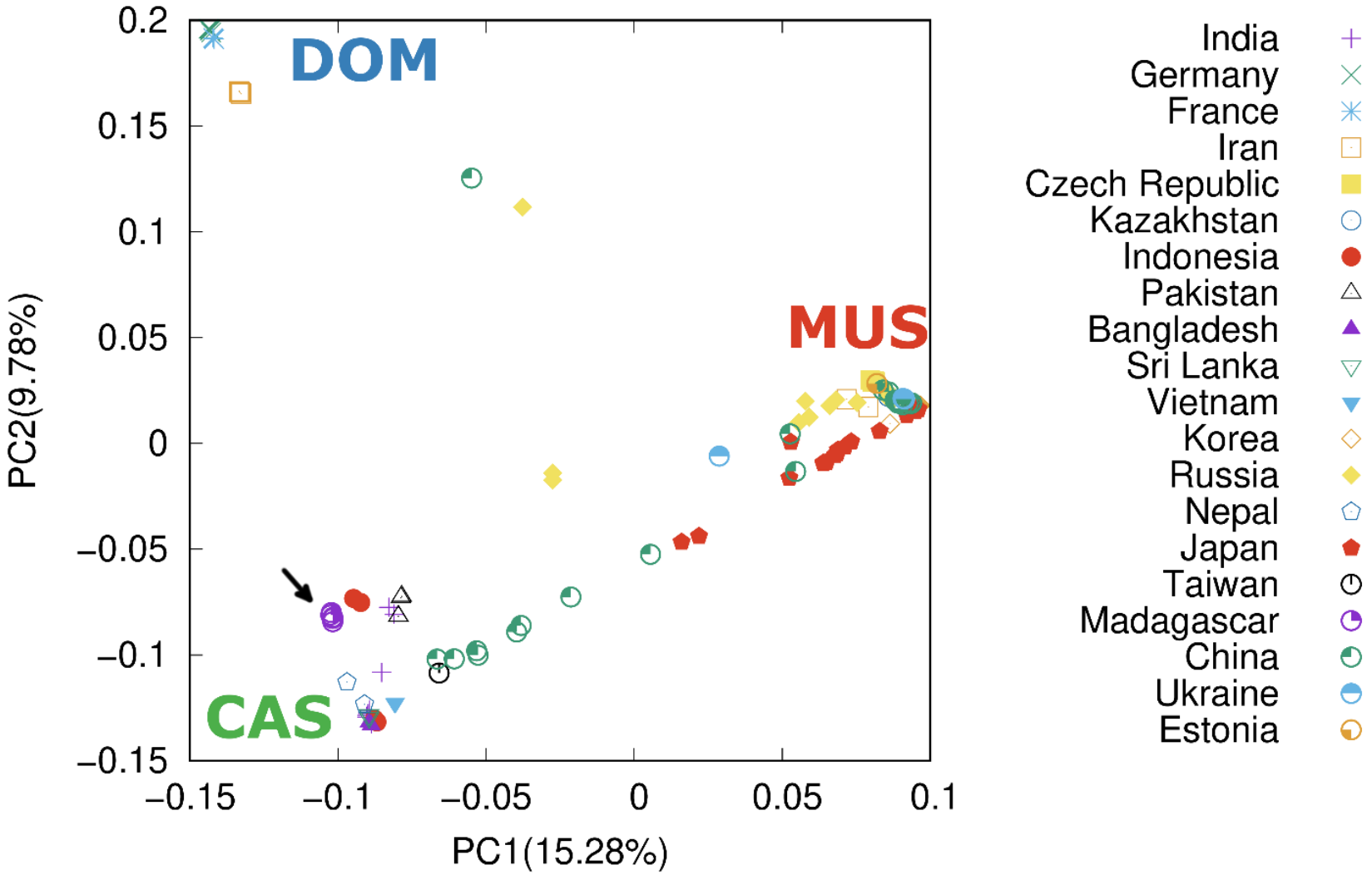
DOM
MUS
CAS
PC2(9.78%)
PC1(15.28%)
0.2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.15
−0.1
−0.05
0
0.05
0.1
India
Germany
France
Iran
Czech Republic
Kazakhstan
Indonesia
Pakistan
Bangladesh
Sri Lanka
Vietnam
Korea
Russia
Nepal
Japan
Taiwan
Madagascar
China
Ukraine
Estonia

Figure 1.

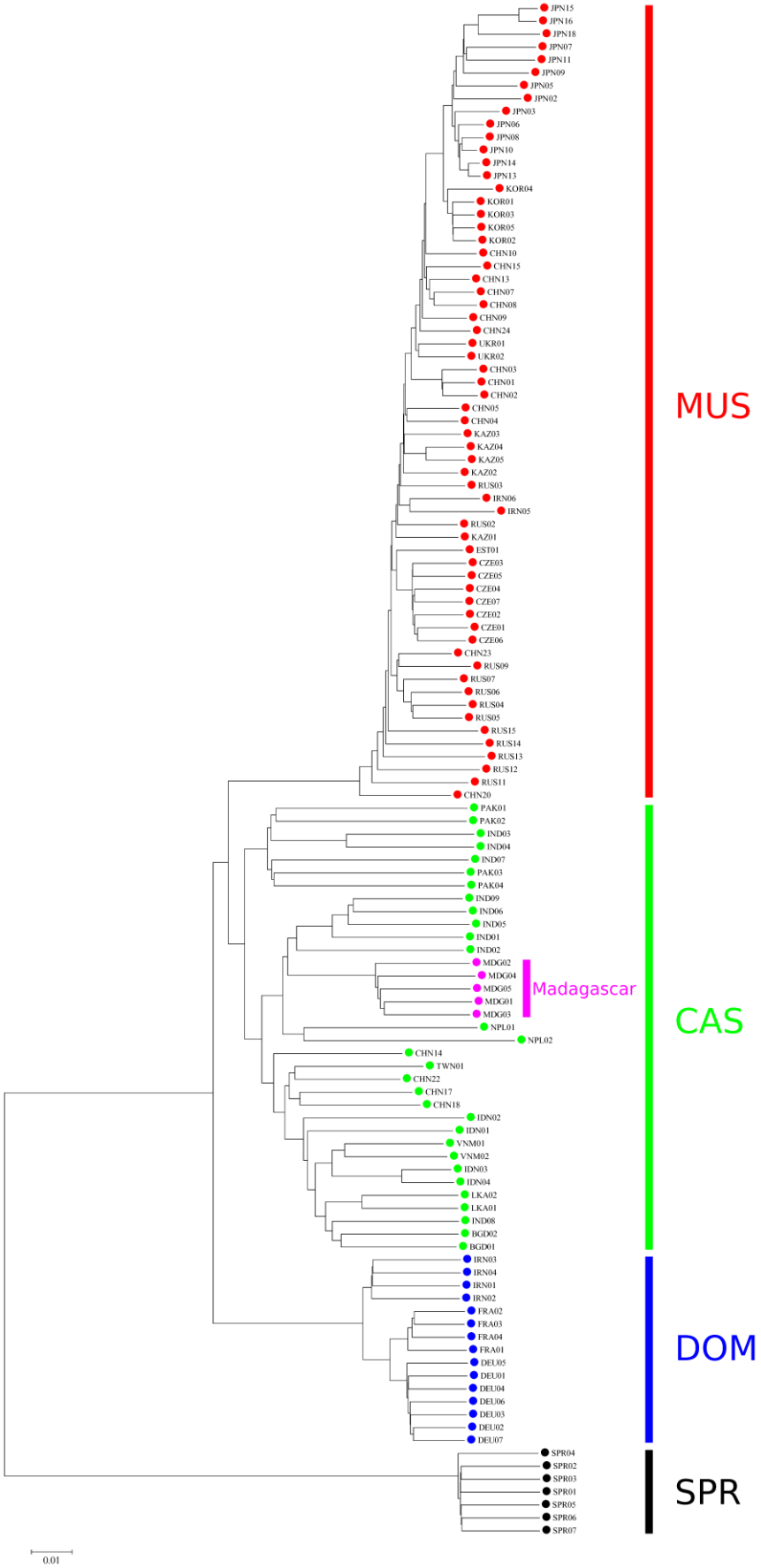

Figure 2.

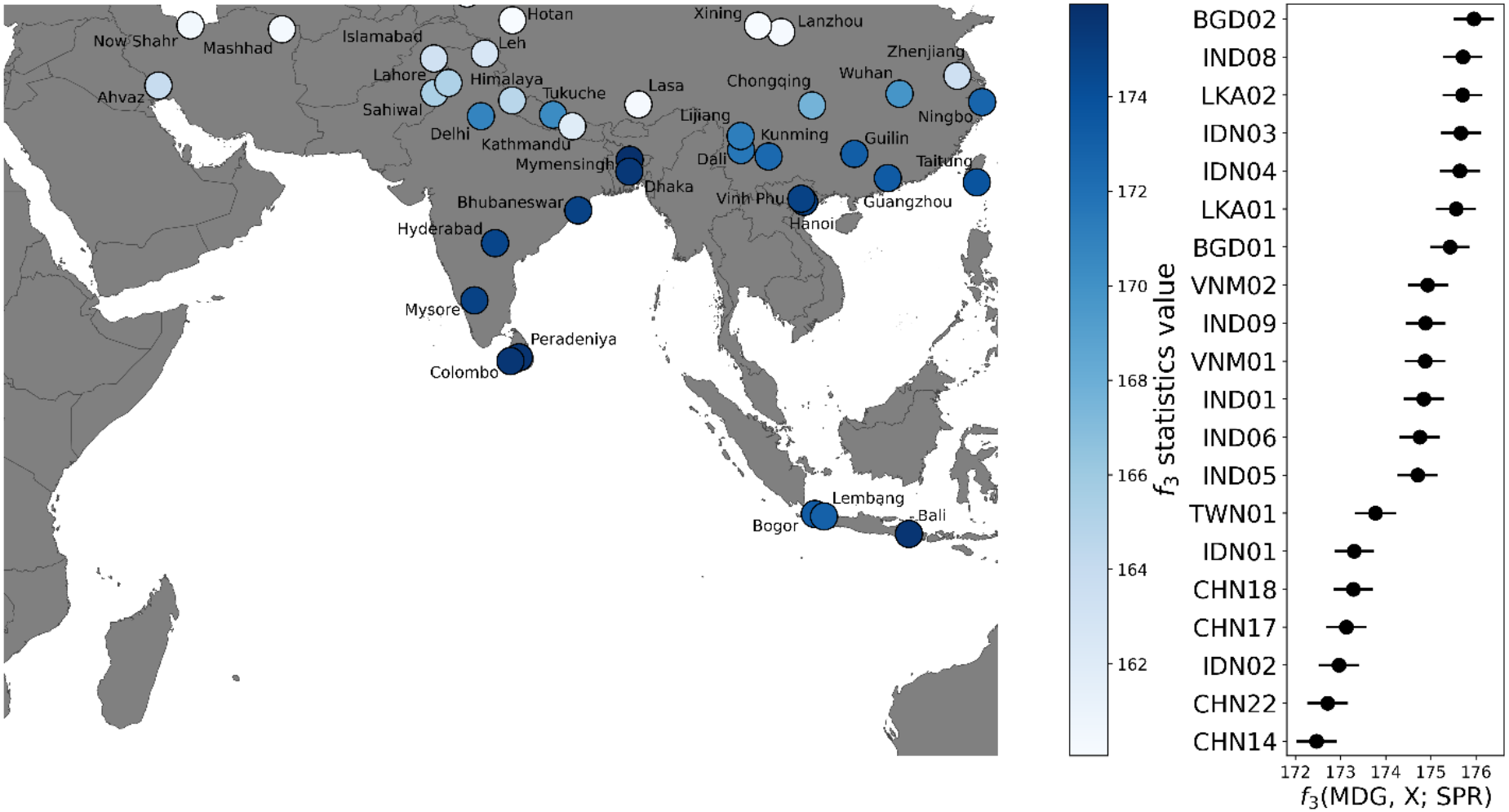
Hotan
Xining
Lanzhou
Now Shahr
Mashhad
Islamabad
Leh
Zhenjiang
Lahore
Himalaya
Wuhan
Ahvaz
Tukuche
Lasa
Chongqing
Sahiwal
Ningbo
Lijiang
Delhi
Kunming
Guilin
Kathmandu
Dali
Taitung
Mymensingh
Dhaka
Vinh Phu
Guangzhou
Bhubaneswar
Hanoi
Hyderabad
Mysore
Peradeniya
Colombo
Lembang
Bali
Bogor
174
172
170
168
166
164
162
f_3 statistics value
BGD02
IND08
LKA02
IDN03
IDN04
LKA01
BGD01
VNM02
IND09
VNM01
IND01
IND06
IND05
TWN01
IDN01
CHN18
CHN17
IDN02
CHN22
CHN14
172
173
174
175
176
f_3(MDG, X; SPR)

Figure 3.

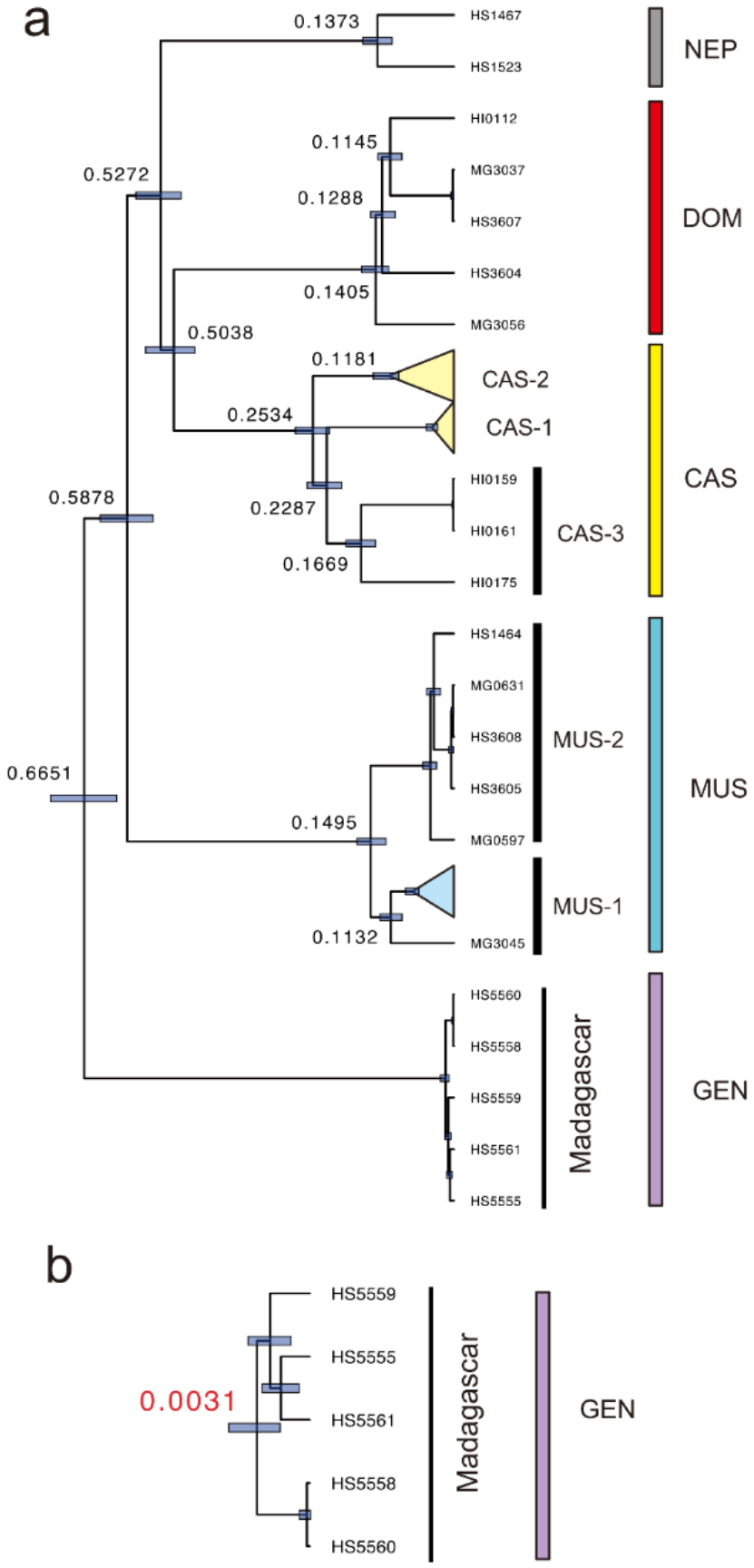

Figure 4.

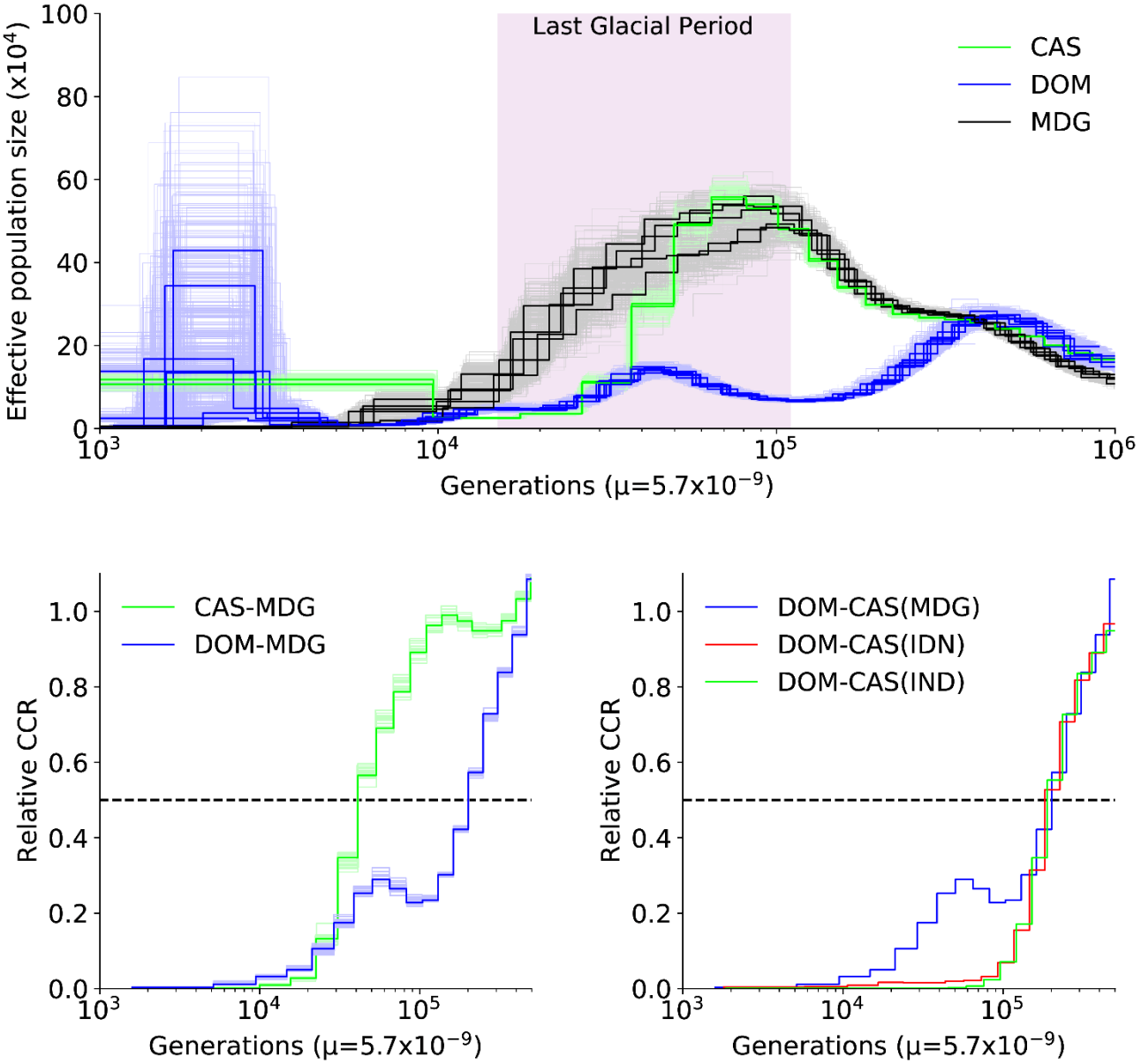

Figure 5.

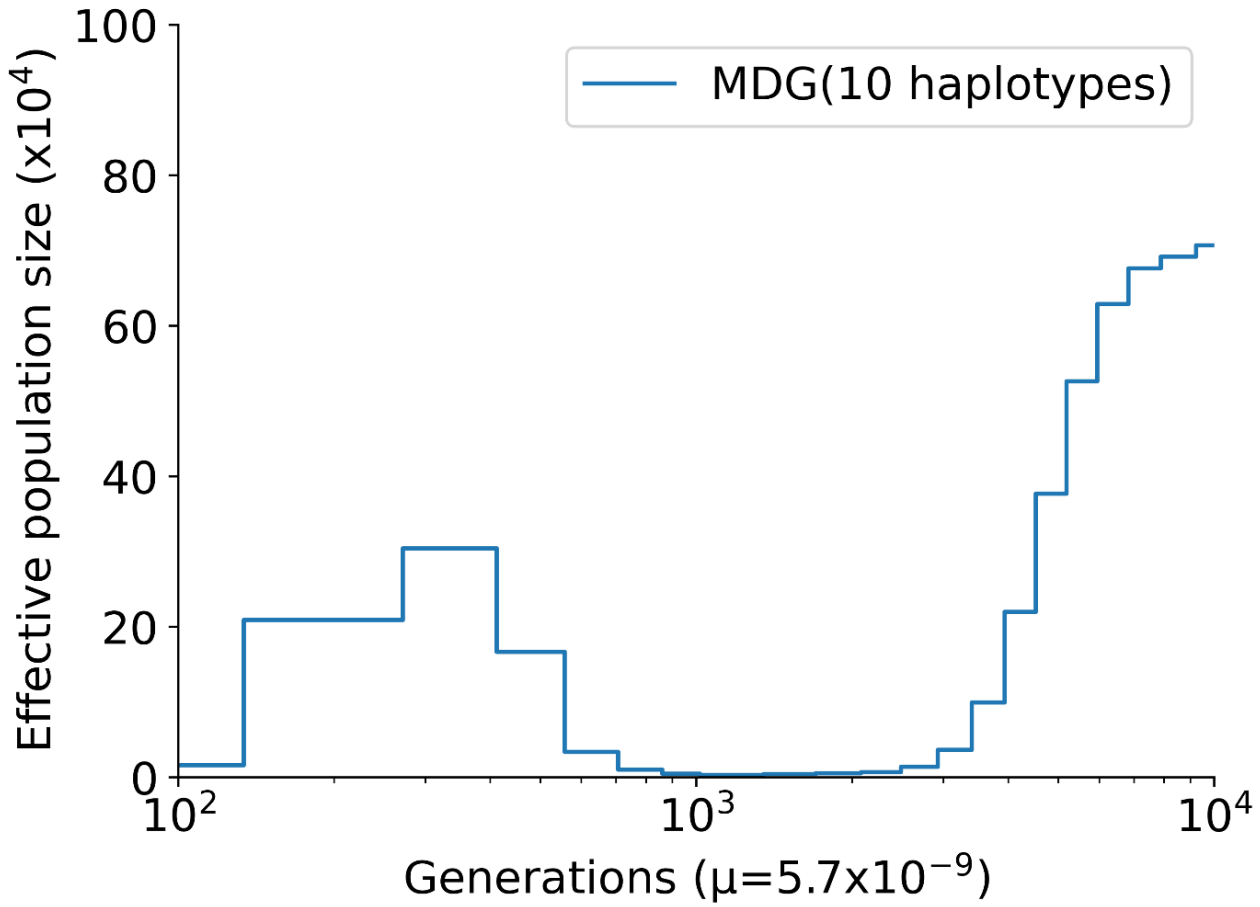

Figure 6.

Table 1. Basic Statistics of Malagasy House Mouse Samples

Sample Code	Sample ID	Location	mtDNA Haplogroup	Median Coverage	Ts/Tv	Per-Sample Nucleotide Diversity	Nonsynonymous	Synonymous	Nonsynonymous/Synonymous
MDG01	HS5555	Madagascar: Tsimbazaza	GEN	26.0	2.107	0.003091476	85684	209312	0.40936019
MDG02	HS5558	Madagascar: Tsimbazaza	GEN	27.0	2.099	0.003602944	88291	216901	0.407056676
MDG03	HS5559	Madagascar: Tsimbazaza	GEN	27.0	2.104	0.003338113	85936	214002	0.401566341
MDG04	HS5560	Madagascar: Tsimbazaza	GEN	27.0	2.107	0.002250803	81478	199493	0.408425358
MDG05	HS5561	Madagascar: Tsimbazaza	GEN	26.0	2.104	0.003375003	87030	215782	0.403323725

Table 2. The estimated demographic parameters.

Parameter	CAS - Madagascar	DOM - Madagascar
N_{Ancestor}	940,229 (951,825 - 957,181)	970,580 (980,184 - 985,150)
N_{CAS}	243,144 (241,141 - 246,463)	
N_{DOM}		74,307 (73,775 - 74,965)
N_{MDG}	117,614 (117,387 - 119,603)	107,180 (106,976 - 108,607)
T_{DIV}	98,513 (98,102 - 100,417)	113,861 (112,941 - 115,220)
$N_{\text{MDG}}m_{\text{MDG->CAS}}$	0.0045 (0.0064 - 0.0129)	
$N_{\text{MDG}}m_{\text{MDG->DOM}}$		0.0005 (0.0009 - 0.0022)
$N_{\text{CAS}}m_{\text{CAS->MDG}}$	0.0400 (0.0334 - 0.0426)	
$N_{\text{DOM}}m_{\text{DOM->MDG}}$		0.1005 (0.0974 - 0.1008)

*95% bootstrap confidence intervals are shown in the parentheses.