Whole-genome sequencing analysis of wild house mice (Mus musculus) captured in Madagascar

Kazumichi Fujiwara1, Marie C. Ranorosoa2, Satoshi D. Ohdachi3,4, Satoru Arai5, Yuki Sakuma4, Hitoshi Suzuki4 and Naoki Osada1*

1Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido 060-0814, Japan
2Ecole Supérieure des Sciences Agronomiques, Université d’Antananarivo, Antananarivo 101, Madagascar
3Institute of Low Temperature Science, Hokkaido University, Sapporo, Hokkaido 060-0819, Japan
4Graduate School of Environmental Science, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan
5Center for Surveillance, Immunization, and Epidemiologic Research, National Institute of Infectious Diseases, Shinjuku-ku, Tokyo 162-8640, Japan

(Received 12 July 2022, accepted 3 August 2022; J-STAGE Advance published date: 19 November 2022)

In Madagascar, the house mouse (Mus musculus) is widely believed to have colonized with human activities and is now one of the most abundant rodents on the island. However, its genetic background at the genomic level remains unclear, and clarifying this would help us to infer the timing of introduction and route of migration. In this study, we determined the whole-genome sequences of five Madagascar house mice captured from an inland location in Madagascar. We examined the genetic background of samples by analyzing the mitochondrial and autosomal genomes. We confirmed that the mitochondrial genome lineages of collected samples formed a single clade placed at one of the most basal positions in the Mus musculus species. Autosomal genomic sequences revealed that these samples are most closely related to the subspecies M. m. castaneus (CAS), but also contain a genetic component of the subspecies M. m. domesticus (DOM). The signature of a strong population bottleneck 1,000–3,000 years ago was observed in both mitochondrial and autosomal genomic data. In a comparison with global samples of M. musculus, the Madagascar samples showed strong genetic affinity to many CAS samples across a wide range of Indian Ocean coastal and insular regions, with divergence time estimated as around 4,000 years ago. These findings support the proposition that the ancestors of these animals started to colonize the island with human agricultural activity and experienced a complex history during their establishment.

Key words: house mouse, Mus musculus, Madagascar, whole-genome sequence, population genetics

INTRODUCTION

House mice (Mus musculus) are small commensal mammals that spread worldwide along with human activities. They are thought to originate from the northern part of the Indian subcontinent (Boursot et al., 1993, 1996; Din et al., 1996) and to have experienced explosive population size growth after the initial colonization. Three major subspecies of house mice have been recognized so far (Didion and de Villena, 2013; Fujiwara et al., 2022): the southern Asian subspecies (M. m. castaneus: CAS),
the northern Eurasian subspecies \((M. m. musculus: \text{MUS})\) and the western European subspecies \((M. m. domesticus: \text{DOM})\). Previous studies analyzing colonization patterns of house mice have provided insights into the history of human activity, such as migration, spread of agriculture and trading (Prager et al., 1998; Duplantier et al., 2002; Britton-Davidian et al., 2007; Searle et al., 2009a; Suzuki et al., 2013, 2015; Jing et al., 2014; Phifer-Rixey et al., 2018; Li et al., 2021; Fujiwara et al., 2022).

Madagascar is the fourth largest island in the world. It separated from the African continent and from the Indian subcontinent around 160 and 80 million years ago, respectively (Torsvik et al., 2000). The island is now about 300 km from the coastal line of the African continent at its shortest distance. Inferences from linguistic (Dahl, 1951, 1988; Adelaar, 1995), archaeological (Battistini and Verin, 1972; Dewar and Wright, 1993; Burney et al., 2004) and genetic (Hurles et al., 2005) evidence suggested that Malagasy people originate from both East Africa and Southeast Asia (in particular, Borneo). When humans first arrived on Madagascar is still an open and controversial issue. Although previous archaeological evidence has suggested that the first human settlement on the island occurred 1,500–2,000 years ago (MacPhee and Burney, 1991; Burney et al., 2004; Gemmery et al., 2011), recent excavations have provided evidence that humans became active on the island earlier than this (Dewar et al., 2013; Hansford et al., 2018; Douglass et al., 2019). More recently, studies using human genome data showed that ancient Malagasy people migrated from Borneo approximately 1,000–3,000 years ago and from the east coast of Africa around 700–1,500 years ago (Brucato et al., 2016; Pierron et al., 2017). A large commercial network connecting Asia and the Mediterranean Sea along the coastal regions of the Indian Ocean was established at least 1,000 years ago (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 Madagascar, and there are at least 23 species of rodents currently inhabiting the island (Duplantier et al., 2002). After Rattus rattus, \(M. m. musculus\) is the second most common rodent found in human settlements throughout the island (Goodman, 1995). Duplantier et al. (2002) have reported that the tail of Madagascar house mice is as long as the head and body, which is a typical trait of DOM (Orsini et al., 1983). The Madagascar house mouse has therefore been thought to be a member of DOM (Boursot et al., 1993). In contrast, previous studies analyzing partial regions of mitochondrial genomes (control regions, the cytochrome \(b\) gene and tRNAs) found that Madagascar house mice were genetically close to samples from Yemen, in the southern part of the Arabian Peninsula (Duplantier et al., 2002; Sakuma et al., 2016). The mtDNA analysis suggested that the Yemeni house mice form another potential subspecies, \(M. m. gentilulus\) (GEN), which is distinct from the other three major subspecies (Prager et al., 1998; Suzuki et al., 2013). The mtDNA sequences of Madagascar samples collected from different localities constituted a single cluster with Yemeni samples (Duplantier et al., 2002; Sakuma et al., 2016). Suzuki et al. (2013) further compared the mitochondrial sequence of Yemeni GEN with those of other wild house mice and found that the mitochondrial lineage of the Yemeni GEN was in the most basal position to \(M. m. musculus\) among the other \(M. m. musculus\) subspecies. However, a study of microsatellite variation in the nuclear genome suggested that Madagascar mice are genetically close to CAS (Hardouin et al., 2015). These incongruences make the origin of Madagascar house mice highly controversial.

Previous studies on the genetic background of the house mouse in Madagascar are primarily based on mtDNA or limited nuclear genome analysis. To increase the resolution of genetic analyses, we performed high-quality whole-genome sequencing of five \(M. m. musculus\) samples captured in Madagascar and combined the data with a previously published worldwide house mouse genomic dataset (Harr et al., 2016; Fujiwara et al., 2022). We revealed the genetic features of these Madagascar wild house mice and estimated the divergence time between the continental and Madagascar samples. Our study provides insights into the prehistoric migration patterns of house mice and humans in Madagascar.

RESULTS

Genetic diversity of Madagascar house mouse We analyzed five whole-genome sequences of captured \(M. m. musculus\) from Madagascar that were partly analyzed in a previous study (Sakuma et al., 2016). These Madagascar samples were jointly genotyped with 128 publicly available samples, including seven \(M. spretus\) samples (Harr et al., 2016; Fujiwara et al., 2022). Supplementary Table S1 presents a list of samples used in this study with sample IDs and approximate captured locations. The number of filtered SNVs in five Madagascar samples was 24,883,782 with an overall transition/transversion ratio of 2.10. The average per-sample nucleotide diversity (heterozygosity) of the Madagascar samples was 0.0031. The ratio of nonsynonymous to synonymous mutations in the CAS, DOM, MUS and Madagascar samples was calculated. Each subspecies showed a distinct distribution of the ratios. DOM, MUS and CAS showed high (0.3156–0.3301), intermediate (0.2955–0.3040) and low values (0.2842–0.2953), respectively. Madagascar samples showed a similar level of ratios to CAS (0.2865–0.2905). The histogram is shown in Supplementary Fig. S1. Supplementary Table S2 presents the detailed results for each Madagascar sample.
The derived site frequency spectrum (DSFS) obtained from the Madagascar samples is shown in Fig. 1, using *M. spretus* to polarize the direction of mutations. The sites were classified into synonymous, nonsynonymous and noncoding sites (Fig. 1). The DSFS of Madagascar samples showed a deficiency of sites with low allele frequency, and an excess of sites with intermediate and high allele frequency. This pattern implies that the Madagascar populations underwent severe population decline or recent population bottleneck. We also observed a slight excess of nonsynonymous polymorphisms compared with synonymous polymorphisms in the singleton class, indicating the segregation of slightly deleterious mutations in low frequency polymorphisms.

**Population structure** To obtain an insight into the genetic features of Madagascar house mice, we first performed a principal component analysis (PCA) using 86,288,314 autosomal SNVs of 126 *M. musculus* samples (Fig. 2). In the PCA plot, three clusters on the vertices of a triangle correspond to MUS, CAS and DOM genetic components, and individuals lying between the vertices would be inter-subspecies hybrid samples (Fujiwara et al., 2022). Principal component (PC)1 and 2 represent the genetic differentiation between MUS and the others, and between DOM and the others, respectively. The Madagascar samples were positioned near the CAS cluster with a slightly admixed genetic component of DOM. There was little variation in PC scores among the Madagascar samples. The linkage disequilibrium (LD) pruned version of PCA was not considerably different from Fig. 2 (Supplementary Fig. S2). We also performed ADMIXTURE analysis to estimate ancestral genetic components in the Madagascar samples (Supplementary Fig. S3), assuming three (*K* = 3) and four (*K* = 4) ancestral populations. In both results, the Madagascar samples were found to have a major CAS ancestry and a minor DOM ancestry.

Next, we constructed an autosomal genetic tree using the neighbor-joining method with the *M. spretus* samples as the outgroup (Fig. 3). Only the samples that had a major subspecies component greater than 80% in the ADMIXTURE result (*K* = 3) were used in the phylogenetic analysis (see Methods for details). The Madagascar samples formed a single cluster within the CAS cluster. We also constructed a neighbor-net network (Supplementary Fig. S4), which had a similar pattern to the neighbor-joining tree, showing that the Madagascar samples are within the CAS cluster. To obtain a clearer picture, we used the outgroup *f*₃ statistics, *f*₃(MDG, X; SPR), where MDG, X and SPR represent Madagascar, non-Madagascar *M. musculus* and *M. spretus* samples, respectively.

![Fig. 1. Derived allele frequency spectrum of five Madagascar samples. Light gray, dark gray and black colors represent nonsynonymous, synonymous and noncoding, respectively. The neutral expected frequencies for constant population size are indicated by striped bars.](https://example.com/fig1)
results showed that individuals genetically close to the Madagascar samples (high $f_3$ values) were located around the coastal side of the South Asian regions and Indonesian islands (Fig. 4).

We further investigated the pattern of genetic admixture in Madagascar samples using $f_4$ statistics. Following Fujiwara et al. (2022), we used Indian (IND03 and IND04) and German (DEU01–DEU07) samples as the reference populations for CAS and DOM, respectively. We computed $f_4$(SPR, CAS; DOM, MDG) and $f_4$(SPR, DOM; CAS, MDG), obtaining 0.0054 (Z-score: 52.43) and 0.0066 (Z-score: 26.25), respectively. In both cases, the $f_4$ statistics were significantly positive values, suggesting that the Madagascar samples were genetically closer to CAS than to DOM, but with a non-negligible amount of gene flow with DOM.

Mitochondrial lineage analysis We investigated the genetic background of Madagascar samples using complete mitochondrial genomes. Phylogenetic trees based on maximum likelihood inference revealed that the five individuals from Madagascar formed a single cluster (Fig. 5A, Supplementary Fig. S5). The Madagascar samples were situated in the most basal lineage of all $M.\text{musculus}$ mitochondrial haplotypes, but bootstrap support was not sufficiently high (0.64, Supplementary Fig. S5). The time to the most recent common ancestor (tMRC) of Madagascar mitochondrial lineages was estimated using the BEAST v1.8.4 software and we obtained the age of 3,100 years ago (Fig. 5B).

Inference of past demographic history Pairwise sequentially Markovian coalescent (PSMC) analysis was performed to infer the demographic history. All Madagascar samples showed almost the same effective population size trajectories over the past 1,000,000 years (Fig. 6). The ancestral population of Madagascar samples showed a rapid increase from 300,000 to 200,000 years ago, peaked around 100,000 years ago, and has been slowly declining since then over the Last Glacial Period. PSMC plots for representative individuals of CAS and DOM are also shown in Fig. 6. CAS and Madagascar populations showed similar population histories before about 50,000 years ago. The CAS population exhibited a rapid decrease in effective population size after 50,000 years ago. The Madagascar samples also showed a similar decline, but with a slower rate than the CAS population. We also performed multiple sequentially Markovian coalescent (MSMC) analysis using Madagascar samples to analyze the relatively recent effective population size (Fig. 7). The plot exhibits a historical bottleneck event for the Madagascar samples around 1,000–3,000 years ago.

Estimation of divergence time between populations To estimate when the Madagascar samples diverged from the other populations, we performed a cross-population MSMC analysis. We evaluated cross
Fig. 3. Neighbor-joining tree using pairwise genetic distances of *M. musculus*. The seven samples of *M. spretus* (SPR) were used as the outgroup. The color shows the subspecies: red (*M. m. musculus*: MUS), green (*M. m. castaneus*: CAS), blue (*M. m. domesticus*: DOM) and magenta (*M. musculus*: Madagascar (MDG)). Hybrid samples were excluded from tree construction. The scale shows the identity-by-state (IBS) distance.
coalescent rates between Madagascar samples and 13 other samples consisting of German DOM samples, Indian CAS samples and CAS samples from coastal and insular regions of the Indian Ocean. The latter refers to samples from India (Bhubaneswar, Hyderabad, Mysore), Bangladesh (Dhaka, Mymensingh), Sri Lanka (Colombo, Peradeniya) and Indonesia (Bali, Java). The divergence time between German DOM and Madagascar was approximately 225,000 years ago with a 95% confidence interval (CI) of 224,641–226,546. While an Indian CAS, which was sampled at an inland mountainous area, showed a divergence time of approximately 50,000 years ago (95% CI: 48,907–50,011) from Madagascar, the other CAS samples from the coastal and island regions of the Indian Ocean showed a much more recent (~4,000 years ago) divergence time from Madagascar samples (Fig. 8). As a control, we performed the analysis for two samples from Bali and found that they did not show the signature of population divergence (Supplementary Fig. S6).

**DISCUSSION**

The house mouse is an animal that has now colonized a wide range of areas including remote islands. This creature has been living commensally with humans for at least 10,000 years (Cucchi et al., 2005) and is thought to have been introduced to islands, including Madagascar, by human activities over thousands or hundreds of years. Other than Madagascar, New Zealand (Searle et al., 2009a), Madeira (Gündüz et al., 2001) and the British Isles (Searle et al., 2009b) are known for their studies of the introduction of house mice to islands. In this study, we present genetic features of Madagascar house mice by comparing them with global samples and provide insights into how they have been introduced into and established on the island. However, we should make the caveat that this study was conducted using a limited number of samples captured in a small region of Madagascar, and these samples do not necessarily represent the entire Madagascar house mouse population. A large-scale sampling across the island would be desirable to draw a clearer picture in the future. We should also note that, although we showed population size trajectories inferred using PSMC and MSMC in Figs. 6 and 7, apparent effective population size may change with many factors, such as population structure, migration and admixture, and may not reflect true effective population sizes (Mazet et al., 2016). Indeed, estimated effective population size ($N_e$) values should be viewed as the inverse of coalescent rate in piecewise time.

**Genetic background of the Madagascar wild house mouse samples** Previous studies have shown that wild *M. musculus* exhibit a very large effective population size and nucleotide diversity compared to humans. The Madagascar house mouse samples analyzed in this study also showed a higher nucleotide diversity (0.31%) than humans (0.08–0.12%) (Perry et al., 2012; Prado-Martinez et al., 2013; Arbiza et al., 2014). However, given that the Madagascar samples were clustered with CAS, whose
Fig. 5. Phylogenetic tree of *M. musculus* entire mitochondrial genome sequences (15,000 bp) with a Bayesian-relaxed molecular clock. Numbers represent the divergence time of haplotypes in million years, and the blue bars represent the 95% highest posterior density interval. (A) Divergence times of five haplogroups of *M. musculus* were estimated using a mutation rate of $2.4 \times 10^{-8}$ substitutions/site/year. See Li et al. (2021) for the details of the subclades. (B) 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 $1.1 \times 10^{-7}$ substitutions/site/year.
nucleotide diversity is around 0.74–0.79% (Geraldes et al., 2008; Phifer-Rixey et al., 2014; Harr et al., 2016), the relatively low nucleotide diversity of the Madagascar samples is probably related to the population bottleneck when they were brought to the island. The DSFS of Madagascar samples also agrees with the population bottleneck scenario. However, the ratios of nonsynonymous to synonymous mutations in Madagascar samples were comparable with those in CAS. The nonsynonymous/synonymous ratio is often interpreted as an indicator of the efficacy of natural selection against slightly deleterious mutations, because the effect of genetic drift should become stronger in small populations (Akashi et al., 2012). The observed pattern agrees with cases in humans and primates, where recent population size changes did not strongly affect the genetic load of individuals (Simons et al., 2014; Osada et al., 2015).

Based on the results of the PCA plot, ADMIXTURE, $f_4$ statistics, phylogenetic tree and phylogenetic network, we concluded that the nuclear genomes of Madagascar samples largely consisted of a CAS genetic component. This result is consistent with a previous microsatellite study (Hardouin et al., 2015). However, the analyses demonstrated that our Madagascar samples have experienced admixture events with DOM. A recent study of quantitative trait loci identified multiple regions causing hybrid male sterility in CAS and DOM hybrids (White et al., 2012). According to that study, a fairly high percentage of male individuals in the second-generation hybrids ($F_2$) exhibited phenotypes associated with infertility, indicating that gene introgression between CAS and DOM is difficult. Nevertheless, the Madagascar samples exhibit the genetic feature of CAS–DOM hybrids (DOM-like CAS), indicating that inter-subspecific hybridization is rare but possible.

Mitochondrial genome analysis indicated that the mitochondrial lineages of the Madagascar samples formed a monophyletic group with low nucleotide diversity and that these lineages have a recent origin, with a tMRCA of approximately 3,000 years ago. This pattern is consistent with the results of Duplantier et al. (2002), who analyzed samples from different locations in Madagascar. 

Fig. 6. PSMC plot of *M. musculus*. Green, blue and black lines represent the inferred effective population size transition of CAS, DOM and Madagascar (MDG) mouse populations, respectively. The x-axis represents years before the present, scaled by $g = 1$ and $\mu = 5.7 \times 10^{-9}$, and the y-axis represents the number of the inferred effective population size. Lines with lighter colors represent 100 replications of the bootstrapping results.

Fig. 7. MSMC plot of five Madagascar samples using 10 haplotypes. The x-axis represents years before the present, scaled by $g = 1$ and $\mu = 5.7 \times 10^{-9}$. The y-axis represents the inferred effective population size of Madagascar house mice.
except that Yemeni samples were also clustered with the Madagascar samples in Duplantier et al. (2002). The MSMC analysis showed that our Madagascar samples experienced a population bottleneck event about 1,000–3,000 years ago. These results suggest that *M. musculus* was first brought to the island of Madagascar during this period.

We showed that the mitochondrial lineage of Madagascar samples was placed in one of the most basal positions of the entire *M. musculus* species. However, the study by Duplantier et al. (2002), which analyzed 539 nucleotide sites in the D-loop of the mitochondrial genome, showed that the MUS clade was located at the basal position of all *M. musculus*. We created a phylogenetic tree with whole mitochondrial genomes and showed the mitochondrial genealogy with higher resolution than previous analyses, but the bootstrap value supporting the basal position was 0.64, which is still not sufficiently high to conclude with high confidence which mitochondrial lineage is the most basal.

Collectively, our study showed highly complex genetic features of the Madagascar samples. The nuclear genomes are mostly derived from CAS, but with a significant amount of admixture with DOM. In contrast, the mitochondrial lineage of Madagascar samples has diverged highly from both CAS and DOM lineages. These findings indicate that the genetic background of Madagascar house mice is unique among the other worldwide populations.

**Divergence of Madagascar population from other CAS populations** We examined which CAS samples were closely related to the Madagascar samples. Although the neighbor-joining tree and neighbor-net network showed that the Madagascar samples clustered with Indian and Nepalese samples, $f_1$ statistics and MSMC analysis showed that the CAS samples in a wide range of the coastal and insular regions of the Indian Ocean are genetically closer to the Madagascar samples than the inland Indian samples. Considering that the inclusion of admixed samples to tree and network constructions potentially results in a distorted pattern, we concluded that the results of $f_1$ statistics and MSMC were likely to be more reliable. In particular, the model of MSMC assumes that the coalescent time of genomic fragments between chromosomes differs across the genome and so the estimated divergence time should be more robust to gene flow and admixture events. The relatively recent divergence time (~4,000 years ago) between the Madagascar and the Indian Ocean coastal and insular populations suggests that the migration of the Madagascar population coincided with human activity in the Neolithic period.

Notably, there was a small but clear increase in the relative cross coalescent rate (rCCR) between the DOM–Madagascar populations around 50,000–70,000 years ago (Fig. 8). This pattern is likely due to an admixture event between the ancestral population of German DOM and the Madagascar population. However, it should be noted that the peak timing (50,000–70,000 years ago) does not necessarily represent the actual timing of the admixture event. Since we used German DOM samples as representatives of DOM, if an unsampled or extinct DOM population was the source population, the peak timing would instead represent the divergence time between the ancestral population of German DOM and the unsampled/extinct DOM. Unfortunately, our dataset did not cover a large number of DOM samples, and we were not able to examine the timing of historical admixture events.

![Fig. 8. Relative cross coalescent rate (rCCR) between the CAS (Indian Ocean)-MDG, CAS-MDG and DOM-MDG inter-populations. The x-axis represents years before the present, scaled by $g = 1$ and $\mu = 5.7 \times 10^{-9}$. The rCCR value of 0.5 is assumed heuristically as the time when the two populations separated.](image)
Possible history of the Madagascar wild house mouse samples

The ancestors of present-day Malagasy people are thought to have dual origins on the Indonesian island of Borneo in Southeast Asia and on the east coast of Africa, and wild house mice are thought to have been introduced commensurately with the migration of these human ancestors who arrived on the island of Madagascar. Here, we would like to discuss the potential migration time and route of Madagascar house mice to the island.

First, we consider when the ancestors of the Madagascar house mouse were introduced onto the island. Although the rate of mitochondrial molecular evolution, nuclear genome mutation rate and generation time in wild house mice are controversial issues, and these assumptions make the estimated timing highly vulnerable, our analyses consistently indicate a strong population bottleneck around 1,000–3,000 years ago. Note that, in principle, population admixture would not create a spurious signal of a population bottleneck. However, the timing of the bottleneck is compatible with the generally accepted date of arrival in Madagascar of Austronesian-speaking peoples from the Indonesian Islands (BCE 300–CE 500) (Dewar and Wright, 1993; Burney et al., 2004; Hurles et al., 2005). Furthermore, cross-population MSMC analyses indicate that the divergence between the Madagascar samples and those from the coastal and insular areas of South and Southeast Asia surrounding the Indian Ocean occurred approximately 4,000 years ago. Since commensal animals, such as house mice, would not suddenly expand their population size without an appropriate type of human activity like agriculture, we suggest that the Madagascar house mice migrated to the island at the same time as or soon after the first Austronesian-speaking farmers arrived on the island.

Second, it is difficult to answer the question of where the ancestors of the Madagascar samples migrated from. Because our \( f_2 \) and MSMC analysis indicated that the Madagascar samples have genetic affinities with many samples from coastal and insular regions of the Indian Ocean, it would be difficult to precisely determine their origin. In addition, our samples did not cover Yemen, where previous mitochondrial studies have suggested an origin, or Borneo, where Austronesian-speaking peoples are supposed to have migrated from. The original homeland of the Madagascar samples therefore still remains an open question. Another interesting question is whether there were multiple introductions of house mice to Madagascar. Previous mitochondrial analyses of rats from multiple locations in Madagascar indicated that there is considerable genetic diversity within Madagascar, originating from multiple sites in the Indian Ocean (Tollenaere et al., 2010). In this study, we show that Madagascar house mice possess both CAS and DOM genetic traits, which strongly supports the multiple origin scenario. If samples can be collected from multiple sites throughout Madagascar, a more definitive view of Madagascar house mice may be obtained in the future.

In summary, this study determined the whole-genome sequences of five Madagascar wild house mice and analyzed their genetic backgrounds. Until now, mice from Madagascar have been treated as DOM or GEN, but we showed that their genetic features are mostly derived from CAS with a non-negligible amount of introgression from DOM. In addition, mitochondrial and nuclear genome data indicated that they experienced a strong population bottleneck approximately 1,000–3,000 years ago, coinciding with the arrival of Austronesian farmers in Madagascar. In the future, mouse samples from Middle East and Borneo, as well as more samples from across Madagascar, should clarify the in-depth genetic background and history of the Madagascar mouse population. Overall, we have revealed unique and highly complex genetic features of Madagascar house mice and provided insights into the evolutionary history of rodents that have successfully migrated to Madagascar, which has only been partially understood.

MATERIALS AND METHODS

Biological samples and genomic data

Five specimens of the Madagascar wild house mouse were collected in the Parc Botanique et Zoologique de Tsimbazaza, Madagascar and nearby areas, from which a partial mitochondrial genome and coat color-controlling gene were sequenced in the study of Sakuma et al. (2016). The other mitochondrial and nuclear genomic sequence data of *M. musculus* as well as the western Mediterranean mouse (*M. spretus*) were obtained from previous studies (Harr et al., 2016; Li et al., 2021; Fujiwara et al., 2022). These data were downloaded from the DDBJ (PRJDB11027) and the European Nucleotide Archive (PRJEB9450, PRJEB11742, PRJEB14167, PRJEB2176 and PRJEB11897). The reference complete mitochondrial sequence of *M. spretus* (NC_025852) was also used for the mitochondrial genome analysis. See Supplementary Table S1 for detailed information on the samples used in this study.

Mapping genomic reads and variant calling

For the five Madagascar samples, 150-bp-length paired-end reads were sequenced using the DNBSEQ platform. The reads were quality-checked using the FastQC v.0.11.9 tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). All cleaned reads were mapped to the house mouse reference genome sequence GRCm38 (mm10) using the bwa-mem v.0.7.17-r1188 (Li and Durbin, 2009) algorithm with “-M” option and piped to the Samblaster tool (https://www.bioinformatics.babraham.ac.uk/projects/
Mitochondrial lineage analysis and SNP-Sift programs (Cingolani et al., 2012a, 2012b). annotation data version GRCm38.101 (ftp://ftp.ensembl.org/pub/release-101/gtf/mus_musculus/) using the SnpEff and SnpSift programs (Cingolani et al., 2012a, 2012b).

Mitochondrial lineage analysis The mitochondrial genome sequences of Madagascar samples were reconstructed by 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. The mitochondrial sequences of all samples were aligned using MUSCLE (Edgar, 2004a, 2004b) implemented in MEGA7 (Kumar et al., 2016). All D-loop regions and gapped sites were removed, resulting in 16,038 bp length. To construct the maximum likelihood (ML) tree of mitochondrial genomes, we used IQ-TREE v.1.6.12 (Nguyen et al., 2015). The best substitution model, “TIM2+F+R3”, was selected using the ModelFinder (Kalyaanamoorthy et al., 2017) function implemented in IQ-TREE. Bootstrapping values were computed with 1,000 replications.

The divergence time among mitochondrial genomes was estimated using BEAST v1.8.4 (Drummond and Rambaut, 2007) with the HKY+G substitution and the strict clock models (Li et al., 2021). The Markov chain Monte Carlo simulation was run for 10 million steps including 10% burn-in steps, and samples were recorded for every 10,000 steps. 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). Evolutionary rates of $2.4 \times 10^{-8}$ substitutions/site/year and $1.1 \times 10^{-7}$ substitutions/site/year were used for the older (before 100,000 years ago) and younger divergences (after 20,000 years ago), respectively.

Population structure analysis of the Madagascar house mouse We filtered our variant data to exclude indels and multiallelic SNVs using VCFTools. The filtered autosomal SNVs were then converted to a PLINK v.1.9 (Purcell et al., 2007) format. We polarized a mutation direction of each SNV using M. spretus as the outgroup species and calculated DSFS. In general practice, the SNVs in LD are excluded from population structure analysis. However, because the house mouse samples were genetically highly structured at the subspecies level, and excluding these SNVs would result in a very small number of SNVs for the downstream analysis, we performed PCA with and without LD pruning but presented the results without the LD pruning as primary results. PCA was performed using the smartpca program implemented in Eigensoft v.7.2.1 (Patterson et al., 2006) using the default parameter settings, except that we did not remove outlier samples automatically.

A neighbor-joining tree was reconstructed using a pairwise distance matrix generated from autosomal data
Demographic inference  Pairwise sequentially Markovian coalescent (PSMC) (Li and Durbin, 2011) analysis was performed for all five samples from Madagascar. To create the input file for PSMC, we used the “mpileup” command of samtools with the “-C” option, and used vcftools.pl program in the PSMC package with the “-D 60 -d 10” option, to obtain the consensus autosomal genome sequence with GRCm38 (mm10) as the reference genome. The tree was visualized using the Ape package in R (Paradis et al., 2004). Using the same samples, a Neighbor-Net tree (Bryant and Moulton, 2004) was computed using the phangorn package (Schliep, 2011) with gdsfmt, SeqArray and SNPRelate packages (Zheng et al., 2012, 2017) in R.

We computed $f_5$ and $f_4$ statistics using AdmixTools v.7.0 (Patterson et al., 2012). The $f_5$ statistics were computed using the qpDstat program with “f4 mode” and “printsd” options. The Indian CAS (IND03 and IND04) and German DOM (DEU01–DEU07) samples were used as reference populations. These samples showed the least amount of admixture with other subspecies based on the $f_4$ statistics in the previous study (Fujiwara et al., 2022). The configuration of the $f_5$ statistics is represented as $f_5(A; B; C; D)$, where A to D represent each population. The outgroup $f_4$ statistics were computed by the qp3pop program of AdmixTools using the “outgroup-mode” option. In the outgroup $f_4$-mode, the statistics of $f_4(A; B; C)$, where C represents an outgroup population, show the genetic drift from C to the common ancestor of A and B and represent the genetic relatedness between the populations A and B.

Data Archiving  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 the DDBJ database under the accession numbers LC644158–LC644162.

We thank two anonymous reviewers for helpful comments. This work was partly supported by MEXT KAKENHI (grant 18H05511 to N. O. and grant 18H05508 to H. S.). We would like to express our gratitude to Dr. Chihiro Tanaka, who belongs to the Animal Care and Exhibition Section, Yagiyama Zoological Park, Sendai, Japan, for her support of the sample collection project. We would also like to express our deepest gratitude to Drs. Kimiyuki Tsuchiya and Hajanirina Ramino for sample collection in Madagascar.

REFERENCES

Adelaar, A. (1995) Asian roots of the Malagasy: a linguistic perspective. Bijdragen tot de Taal- Land- en Volkenkunde 151, 325–356.

Akashi, H., Osada, N., and Ohta, T. (2012) Weak selection and protein evolution. Genetics 192, 15–31.

Allibert, C. (1988) Les contacts entre l’Arabie, le Golfe perique, l’Afrique orientale et Madagascar: confrontation des documents ecrits, des traditions orales et des donnees archeologiques recentes. Travaux de la Maison de l’Orient 16, 111–126.

Arbiza, L., Gottipati, S., Siepel, A., and Keinan, A. (2014) Contrasting X-linked and autosomal diversity across 14 human populations. Am. J. Hum. Genet. 94, 827–844.

Battistini, R., and Verin, P. (1972) Man and the environment in Madagascar. In Biogeography and Ecology in Madagascar. (eds.: Battistini, R., and Richard-Vindard, G.), pp. 311–337. Springer, Netherlands.

Boursot, P., Auffray, J.-C., Britton-Davidian, J., and Bonhomme, F. (1993) The evolution of the house mouse. Annu. Rev. Ecol. Syst. 24, 119–152.

Boursot, P., Din, W., Anand, R., Darviche, D., Dod, B., von Deimling, F., Talwar, G. P., and Bonhomme, F. (1996) Origin of...
Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, Bronson, F. H. (1979) The reproductive ecology of the house mouse. *Q. Rev. Biol.* 54, 265–299.

Brucato, N., Kusuma, P., Cox, M. P., Pierron, D., Purnomo, G. A., Adelara, A., Kivisild, T., Letellier, T., Sudoyo, H., and Ricaut, F.-X. (2016) Malagasy genetic ancestry comes from an historical Malay trading post in Southeast Borneo. *Mol. Biol. Evol.* 33, 2396–2400.

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

Burney, D. A., Burney, L. P., Godfrey, L. R., Jungers, W. L., Goodman, S. M., Wright, H. T., and Timothy Jull, A. J. (2004) A chronology for late prehistoric Madagascar. *J. Hum. Evol.* 47, 25–63.

Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., and Lu, X. (2012b) Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front. Genet.* 3, 35.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. (2012a) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3.

Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L., and Abecasis, G. R., Auffray, J.-C., Britton-Davidian, J., Catalan, J., and Bonhomme, F. (2013) Deconstructing *Mus musculus*: advances in understanding ancestry, *Mamm. Genome* 24, 1–20.

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017) NOVO-Plasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45, e18.

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

Douglas, K., Hixon, S., Wright, H. T., Godfrey, L. R., Crowley, B. E., Manjakahery, B., Rasolondrainy, T., Crossland, Z., and Radimilahy, C. (2019) A critical review of radiocarbon dates clarifies the human settlement of Madagascar. *Quat. Sci. Rev.* 221, 105878.

Drummond, A. J., and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.

Dupalantier, J.-M., Orth, A., Catalan, J., and Bonhomme, F. (2002) Evidence for a mitochondrial lineage originating from the Arabian Peninsula in the Madagascar house mouse (*Mus musculus*). *Hereditas* 89, 154–158.

Edgar, R. C. (2004a) Muscle: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.

Edgar, R. C. (2004b) Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

Faust, G. G., and Hall, I. M. (2014) SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 30, 2503–2505.

Fujisawa, K., Kawai, Y., Takada, T., Shirouishi, T., Saitou, N., Suzuki, H., and Osada, N. (2022) Insights into *Mus musculus* population structure across Eurasia revealed by whole-genome analysis. *Genome Biol. Evol.* 14, eva068.

Geraldès, A., Basset, P., Gibson, B., Smith, K. L., Harr, B., Yu, H.-T., Bulatova, N., Ziv, Y., and Nachman, M. W. (2008) Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Mol. Ecol.* 17, 5349–5363.

Gomery, D., Ramanivosa, B., Faure, M., Guérin, C., Kerloch, P., Sénéga, F., and Randrianantenaina, H. (2011) Les plus anciennes traces d’activités anthropiques de Madagascar sur des ossements d’hippopotames subfossiles d’Anjohibe (Province de Mahajanga). Comptes Rendus Palevol 10, 271–278.

Goodman, S. M. (1995) *Rattus* on Madagascar and the dilemma of protecting the endemic rodent fauna. *Conserv. Biol.* 9, 450–453.

Gündüz, Í., Auffray, J.-C., Britton-Davidian, J., Catalan, J., Ganem, G., Ramalhinho, M. G., Mathias, M. L., and Searle, J. B. (2001) Molecular studies on the colonization of the Madeiran archipelago by house mice. *Mol. Ecol.* 10, 2023–2029.

Hansford, J., Wright, P. C., Rasomiamaranana, A., Pérez, V. R., Godfrey, L. R., Errickson, D., Thompson, T., and Turvey, S. T. (2018) Early Holocene human presence in Madagascar evidenced by exploitation of avian megafauna. *Sci. Adv.* 4, eaat6925.

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

Harr, B., Karakoc, E., Neme, R., Teschke, M., Pfeifle, C., Pezer, Ž., Babiker, H., Linnenbrink, M., Montero, I., Scavetta, R., et al. (2016) Genomic resources for wild populations of the house mouse, *Mus musculus* and its close relative *M. spreitus*. *Sci. Data* 3, 160075.

Hoe, S. Y. W., Lanfer, R., Bromham, L., Phillips, M. J., Soubrier, J., Rodrigo, A. G., and Cooper, A. (2011) Time-dependent rates of molecular evolution. *Mol. Biol. Evol.* 20, 3087–3101.

Hurles, M. E., Sykes, B. C., Jobling, M. A., and Forster, P. (2005) The dual origin of the Malagasy in island Southeast Asia and East Africa: evidence from maternal and paternal lineages. *Am. J. Hum. Genet.* 76, 894–901.

Jing, M., Yu, H.-T., Bi, X., Lai, Y.-C., Jiang, W., and Huang, L. (2014) Phylogeography of Chinese house mice (*Mus musculus castaneus*): distribution, routes of colonisation and radiations. *Front. Genet.* 5, 1–20.
nization and geographic regions of hybridization. Mol. Ecol. 23, 4387–4405.
Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589.
Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transformer. Bioinformatics 25, 1754–1760.
Li, H., and Durbin, R. (2011) Inference of human population history from individual whole-genome sequences. Nature 475, 493–496.
Li, Y., Fujiwara, K., Osada, N., Kawai, Y., Takada, T., Kryukov, G. A., Abe, K., Yonekawa, H., Shiroishi, T., Moriwaki, K., et al. (2021) House mouse Mus musculus dispersal in East Eurasia inferred from 98 newly determined complete mitochondrial genome sequences. Heredity 126, 132–147.
Liszkowski, H. D. (2000) Mayotte et les Comores: escales sur la route des Indes aux XVe et XVIIIe siècles. Editions du Baobab, Mamoudzou, Mayotte, France.
Liu, E. Y., Morgan, A. P., Chesser, E. J., Wang, W., Churchill, G. A., and Pardo-Manuel de Villena, F. (2014) High-resolution sex-specific linkage maps of the mouse reveal polarized distribution of crossovers in male germline. Genetics 197, 91–106.
MacPhee, R. D. E., and Burney, D. A. (1991) Dating of modified femora of extinct dwarf Hippopatamus from southern Madagascar: implications for constraining human colonization and vertebrate extinction events. J. Archaeol. Sci. 18, 695–706.
Malaspinas, A.-S., Westaway, M. C., Muller, C., Sousa, V. C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J. Y., Crawford, J. E., et al. (2016) A genomic history of Aboriginal Australia. Nature 538, 207–214.
Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W.-M. (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26, 2867–2873.
Mazet, O., Rodriguez, W., Grusea, S., Boitard, S., and Chikhi, L. (2016) On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? Heredity 116, 362–371.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.
Milhammad, B., Dong, X., Zhang, L., Hao, X., Suh, Y., and Vijg, J. (2017) Differences between germline and somatic mutation rates in humans and mice. Nat. Commun. 8, 15183.
Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.
Orcisí, P., Bonhomme, F., Britton-Davidian, J., Croset, H., Guerazimov, S., and Thaler, L. (1983) Le complexe d’espèces du genre Mus en Europe centrale et orientale. II Critères d’identification, éparition, caractéristiques écologiques. Z. Saugetierkd. 48, 86–99.
Osada, N., Hettiarachchi, N., Adeyemi Babarinde, I., Saitou, N., and Blancher, A. (2015) Whole-genome sequencing of six Mauritian Cynomolgus macaques (Macaca fascicularis) reveals a genome-wide pattern of polymorphisms under extreme population bottlenecks. Genome Biol. Evol. 7, 821–830.
Paradis, E., Claude, J., and Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20, 289–290.
Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Geschoreck, T., Webster, T., and Reich, D. (2012) Ancient admixture in human history. Genetics 192, 1065–1093.
Patterson, N., Price, A. L., and Reich, D. (2006) Population structure and eigenanalysis. PloS Genet. 2, e190.
Perry, G. H., Melsted, P., Marioni, J. C., Wang, Y., Bainer, R., Pickrell, J. K., Michelini, K., Zehr, S., Yoder, A. D., Stephens, M., et al. (2012) Comparative RNA sequencing reveals substantial genetic variation in endangered primates. Genome Res. 22, 602–610.
Phiher-Rixey, M., Bi, K., Ferris, K. G., Sheehan, M. J., Lin, D., Mack, K. L., Keeble, S. M., Suzuki, T. A., Good, J. M., and Nachman, M. W. (2018) The genomic basis of environmental adaptation in house mice. PloS Genet. 14, e1007672.
Phiher-Rixey, M., Bomhoff, M., and Nachman, M. W. (2014) Genome-wide patterns of differentiation among house mouse subspecies. Genetics 198, 283–297.
Pierron, D., Heiske, M., Razafindrazaka, H., Rakoto, I., Rabetokotany, N., Rakotozafy, L. M.-A., Rakotomalala, M. M., Razafiarivony, M., Rasoloferatra, B., et al. (2017) Genomic landscape of human diversity across Madagascar. Proc. Natl. Acad. Sci. USA 114, E6498–E6506.
Pockrandt, C., Alzamela, M., Iliopoulos, C. S., and Reinert, K. (2020) GenMap: ultra-fast computation of genome mappability. Bioinformatics 36, 3687–3692.
Prado-Martinez, J., Sudmant, P. H., Kidd, J. M., Li, H., Kelley, J. L., Lorent-Galdos, B., Veeramah, K. R., Werner, A. E., O’Connor, T. D., Santpere, G., et al. (2013) Great ape genetic diversity and population history. Nature 499, 471–475.
Prager, E. M., Orrego, C., and Sage, R. D. (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, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.
Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904.
Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Sakuma, Y., Ranorosoa, M. C., Kinoshita, G., Shimoji, H., Tsujiyama, K., Ohdachi, S. D., Arai, S., Tanaka, C., Ramino, H., and Suzuki, H. (2016) Variation in the coat-color-adaptation in house mice. J. L. Mam. Biol. Evol. 4, 131–140.
Schiffels, S., and Durbin, R. (2014) Inferring human population size and separation history from multiple genome sequences. Nat. Genet. 46, 919–925.
Schiffels, S., and Wang, K. (2020) MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. Methods Mol. Biol. 2090, 147–166.
Schliep, K. P. (2011) phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593.
Searle, J. B., Jamieson, P. M., Gündüz, İ., Stevens, M. I., Jones, E. P., Gemmill, C. E. C., and King, C. M. (2009a) The diverse origins of New Zealand house mice. Proc. Biol. Sci. 276, 209–217.

Searle, J. B., Jones, C. S., Gündüz, İ., Scascitelli, M., Jones, E. P., Herman, J. S., Rambau, R. V., Noble, L. R., Berry, R. J., Giménez, M. D., et al. (2009b) Of mice and (Viking?) men: phylogeography of British and Irish house mice. Proc. Biol. Sci. 276, 201–207.

Simons, Y. B., Turchin, M. C., Pritchard, J. K., and Sella, G. (2014) The deleterious mutation load is insensitive to recent population history. Nat. Genet. 46, 220–224.

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

Suzuki, H., Shimada, T., Terashima, M., Tsuchiya, K., and Aplin, K. (2004) Temporal, spatial, and ecological modes of evolution of Eurasian Mus based on mitochondrial and nuclear gene sequences. Mol. Phylogenet. Evol. 33, 626–646.

Suzuki, H., Yakimenko, L. V., Usuda, D., and Frisman, L. V. (2015) Tracing the eastward dispersal of the house mouse, Mus musculus. Genes Environ. 37, 20.

Tollenaere, C., Brouat, C., Duplantier, J.-M., Rahalison, L., Rahelinirina, S., Pascal, M., Moné, H., Mouahid, G., Leirs, H., and Cosson, J.-F. (2010) Phylogeography of the introduced species Rattus ratus in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. J. Biogeogr. 37, 398–410.

Torsvik, T. H., Tucker, R. D., Ashwal, L. D., Carter, L. M., Jamtveit, B., Vidyardharan, K. T., and Venkataramana, P. (2000) Late Cretaceous India–Madagascar fit and timing of break-up related magmatism. Terra Nova 12, 220–224.

Vérin, P., and Wright, H. (1999) Madagascar and Indonesia: new evidence from archaeology and linguistics. Bulletin of the Indo-Pacific Prehistory Association 18, 35–42.

White, M. A., Stubbings, M., Dumont, B. L., and Payseur, B. A. (2012) Genetics and evolution of hybrid male sterility in house mice. Genetics 191, 917–934.

Zheng, X., Gogarten, S. M., Lawrence, M., Stilp, A., Conomos, M. P., Weir, B. S., Laurie, C., and Levine, D. (2017) SeqArray—a storage-efficient high-performance data format for WGS variant calls. Bioinformatics 33, 2251–2257.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28, 3326–3328.