Heterogeneous genetic makeup of Japanese house mice (*Mus musculus*) created by multiple independent introductions and spatiotemporally diverse hybridisation processes

Takashi Kuwayama[1], Mitsuo Nunome[2], Gohta Kinoshita[1], Kuniya Abe[3], and Hitoshi Suzuki[1]

[1]Graduate School of Environmental Science, Hokkaido University, North 10, West 5, Sapporo 060-0810, Japan
[2]Avian Bioscience Research Center, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Aichi 464-8601, Japan
[3]Technology and Development Team for Mammalian Genome Dynamics, BioResource Center, RIKEN Tsukuba Institute, Tsukuba, Ibaraki, Japan
*E-mail: htsuzuki@ees.hokudai.ac.jp

Running head: Natural history of Japanese wild mice

Key words: *Mus musculus*, nuclear gene haplotype, peopling of Japan, recombination, secondary contact

ABSTRACT

Current phylogenetic analyses of relatively long sequences of mitochondrial DNA (4,225 bp) in the Japanese house mouse *Mus musculus* provided the first evidence that both southern Asian subspecies of *Mus musculus castaneus* (CAS) and northern Asian subspecies of *M. m. musculus* (MUS) arrived in Japan through rapid population expansion events, from Southern China ~4,000 years ago and the Korean Peninsula ~2,000 years ago, respectively. Nuclear DNA haplotype structure analyses targeted a chromosome region with two different tract sizes, 1 Mb and 5 Mb, consisting of nine and six tandemly arranged markers, respectively, yielding a possible average fragment length of 170 kb of CAS haplotypes in the MUS background genome in northern Japan, providing a rough estimate of its elapsed time of 815 generations under an assumption of continued backcrossing. Less frequent and shortened CAS-like haplotypes specific to Japan were detected, suggestive of ancient introduction prior to the appearance of the South Chinese CAS in Japan. Our analyses also showed sporadic appearance of long fragments (2–5 Mb) from the west European subspecies *M. m. domesticus*, indicating contemporary stowaway introduction. Overall, multiple overseas introductions, and the time-lagged inter-subspecies genetic admixture, likely resulted in the heterogeneous state of Japanese wild mice.

INTRODUCTION

The house mouse (*Mus musculus*) is a particularly fascinating species for demographic and evolutionary studies, due in part to its extensive natural distribution and genetically differentiated subspecies, but also because of its long-term association with humans (Boursot et al., 1993; Terashima et al., 2006). To date, numerous molecular phylogenetic studies performed on *M. musculus*, using a variety of markers, have indicated that *M. musculus* comprised three major genetic groups: *M. m. castaneus* (CAS), *M. m. domesticus* (DOM) *M. m. musculus* (MUS), with respective parapatric distributions in the central, western peripheral and northern peripheral parts of the predicted range, which extends from the Middle East to India (Britton & Thaler, 1978; Sage, 1981; Bonhomme et al., 1984; Auffray et al., 1991; Boursot et al., 1993; Prager, Orrego & Sage, 1998; Kodama et al., 2015). The three subspecies are known to have colonised their respective ranges through prehistoric human activities; CAS from Iran to Indonesia and South China to Southeast Asia, DOM from the Middle East to western Europe and MUS from the nearby southern coastal area of the Caspian Sea to the northern part of Eurasia, including eastern Europe and the Japanese Islands (e.g. Bonhomme et al., 2007). In the modern age, DOM has a large distribution range covering Africa, Oceania and America, and can be easily introduced anywhere in the world (e.g. Gabriel et al., 2011). Reconstruction of the evolutionary history of the house mouse remains particularly difficult in areas such as the Japanese Islands, which have experienced multiple dispersal events in the past and present. We should therefore distinguish among lineage introductions in different time periods and source areas when making phylogeographic inferences regarding this species.

The Japanese Islands mark the endpoint of the eastward movement of *M. musculus*, which occurred along with that of prehistoric humans. Elucidation of the movement of this species to Japan would help clarify the colonisation history of *M. musculus* in East Asia, while also facilitating our understanding of prehistoric agricultural development in Asia. It has been shown that two distinct lineages of the subspecies groups, the South Asian subspecies CAS and North Eurasian subspecies MUS, were introduced to Japan in conjunction with human movement; these lineages are now minor and major components of the overall population, respectively (Terashima et al., 2006; Nunome et al., 2010). The relatively frequent presence of the

CAS lineage has been confirmed in the northern part of Japan (Hokkaido and northern Honshu), by a variety of phylogenetic markers, including allozymes, mtDNA and nuclear gene sequences (Minezawa, Moriwaki & Kondo, 1979; Yonekawa et al., 1988; Bonhomme et al., 1989; Terashima et al., 2006; Nunome et al., 2010). Japanese mice, with their intricate genetic compositions, would be an intriguing subject for the testing of a variety of phylogeographic inference methods.

Due to its lack of recombination and rapid evolutionary rate, mtDNA is a powerful genealogical marker. In the house mouse, there are five distinct lineages representing three major subspecies groups, CAS, DOM and MUS, and two regionally confined populations in Nepal (subspecies undefined) and Yemen (*M. m. gentilulus*) (Sakuma et al., 2016). Our recent work with the mitochondrial cytochrome *b* gene (*Cytb*) and the control region showed that, among four distinct phylogroups of CAS, designated as CAS-1, -2, -3 and -4 (Terashima et al., 2006; Suzuki et al., 2013), CAS-1 is the only lineage that has dispersed via human activity from its predicted source areas, such as eastern India, to a large part of East Asia, Southeast Asia and Indonesia, extending its lineage to the Japanese Islands. Notably, it has been shown that the CAS-1 lineage in northern Japan is related to those seen in South China, implying that South China is the predicted source area of the Japanese CAS-1 lineage (CAS-1a; Suzuki et al., 2013). The Japanese MUS lineage has been shown to be closely related to the specific subgroup of MUS that now occurs in the Korean Peninsula, nearby North China and Primorye, Russia (MUS-1c; Suzuki et al., 2013).

Although nuclear gene sequences are believed to be unsuitable for genealogical inference due to recombination and its slow evolutionary rate, our previous studies used small segments (500 bp) in seven or eight closely linked gene regions along a 200-kb chromosome region; construction of network trees with concatenated sequences yielded fine genealogical signals to assess the relationships between and within subspecies of *M. musculus* (Nunome et al., 2010; Kodama et al., 2015). These studies revealed that there are several local phylogroups within CAS in their predicted homeland, and only one of these was involved in the prehistoric dispersal event (to Southeast Asia, Indonesia, the Philippines and South China). Haplotype block analysis showed that the MUS subtype seen in Japan belongs to groups of mice from northern China and Korea, and differs from MUS mice from eastern Europe and Russia, including the easternmost parts of the country. Additionally, comparison of the level of genetic variability among the linked

genes showed mutually dependent relationships among the mice, in which they exchanged genetic materials in their home ranges during the course of their evolution (Kodama et al., 2015). Utilising phylogeographic signals generated by recombination events, which has proved to be an effective approach to assess the timespan of introgression events, by measuring the length of incorporated chromosome segments (Martinsen et al., 2001; Koopman et al., 2007; Nunome et al., 2010), helped us to uncover evidence of the historical hybridisation event of the two lineages in Japan. This in turn provided robust evidence for historical hybridisation between CAS and MUS. It is evident that DOM haplotypes are incorporated in mice collected from several places that have no specific relationship with each other (e.g. Kyowa, Hokkaido and Atsugi in Eastern Honshu). The DOM fragments are known to be > 200 kb and are considered to result from recent sporadic introductions occurring via modern human activity (Nunome et al., 2010). However, to improve our knowledge of stowaway introductions that presumably took place in the modern or contemporary age, it may be necessary to examine regions with target sizes longer than 200 kb.

We conducted a phylogeographic study to determine the source areas for Japanese CAS and MUS lineages using a relatively long mtDNA sequence marker (approximately 4 kb). We performed haplotype structure analyses targeting 1 Mb and 5 Mb chromosome regions on chromosome 8, consisting of nine (100-kb interval) and six (1-Mb interval) tandemly arranged markers, respectively. We used 24 mice from Japan and five laboratory strains as subspecies reference groups. The aims of this study were to 1) determine the dispersal events that led to Japanese CAS and MUS subtypes by examining a relatively long sequence of mtDNA, 2) assess hybridisation events in northern Japan, where the CAS and MUS lineages are known to have mingled and 3) test the efficiency of using markers, with and without recombination, for phylogeographic inference regarding this evolutionarily complicated species.

MATERIALS AND METHODS

Materials

Our sequencing analysis was mainly based on samples of house mouse genomic DNA stored at the National Institute of Genetics, Mishima, Japan. These samples were collected in a variety of other countries, mainly China, India and Russia, on expeditions

organised by Dr. K. Moriwaki during the period 1983–2003 (MG series, stored in the National Institute of Genetics; BRC Series, stored in the RIKEN BioResource Center), and by Dr. H. Ikeda during the period 1989–1992 (HI series, stored in Hokkaido University). We also used DNA samples stored at Hokkaido University (HS series). In addition, five strains – BFM/2Ms, BLG2/Ms, MSM/Ms, PGN2/Ms and PWK – were provided by RIKEN BRC. Wild mice from Germany (MB9) were kindly provided by Dr. P. Munclinger (Czech Republic). In total, 78 individual mice from 59 localities (Table 1, Fig. 1), including 28 individuals from 22 localities in Japan, were used for mtDNA analysis. Furthermore, 5 inbred strains and 31 wild-captured mice from 26 localities, including 24 individuals from 19 localities in Japan, were used for nuclear DNA analysis.

Sequence analyses

Polymerase chain reaction (PCR) was performed employing primers designed using the Ensemble Mouse Genome Database (http://www.ensembl.org/), the conditions listed in Supplemental Table S1 and the AmpliTaq Gold® 360 Master Mix kit (ABI). PCR products were sequenced using the PRISM Ready Reaction DyeDeoxy Terminator Cycle Sequencing Kit (ABI) and an ABI3130 automated sequencer. Sequence fragments obtained with different primers were assembled using the MEGA5 program (Tamura et al., 2011).

PCR and direct sequencing of the mtDNA marker *Cytb* (1,140 bp; Suzuki et al., 2004) were performed following previously described methods. We amplified a mtDNA region covering *Nd2*, *Co1* and seven adjacent tRNAs.

For nuclear gene analyses, we targeted a chromosome region of a 200-kb tract in the distal part of chromosome 8, in which eight markers are arranged at 10–20 kb intervals (Nunome et al., 2010). For this study, we extended the targeting region to both sides and designed two tracts of 1 Mb and 5 Mb. The former used nine markers intermittently arranged at 80–250 kb intervals (Fig. 2A): *Cdt1* (512 bp), *Cbfa2t3* (508 bp), *Acsf3* (506 bp), *Ankrd11* (615 bp), *Cdk10* (414 bp), *Fanca* (448 bp), *Tcf25* (600 kb), *Dbndd1* (550 bp) and *Rhou* (461 bp). The *Rhou* tract was covered with five markers of *Zcchc14* (548 bp) – *Cdt1* (the first marker of the 1-Mb tract), *Rhou* (the last marker of the 1-Mb tract), *Pgbd5* (507 bp), *Sipa1l2* (586 bp) and *Irf2bp2* (527 bp) – which were arranged at 0.8–1.2 Mb intervals. Sequences with more than one polymorphic site were

separated into alleles, mainly by the parsimony method (Karn et al., 2002). For loci with several heterozygous sites, alleles were determined using PHASE 2.1 software (Stephens, Smith & Donelly, 2001; Stephens & Donnelly, 2003). *Fanca*, *Tcf25* and *Dbndd1* sequences were obtained from databases for 14 samples (Nunome et al., 2010). We used PHASE 2.1 to infer unique haplotypes with respect to the 13 gene regions of *Zcchc14*, *Cdt1*, *Cbfa2t3*, *Acsf3*, *Ankrd11*, *Cdk10*, *Fanca*, *Tcf25*, *Dbndd1*, *Rhou*, *Pgbd5*, *Sipa1l2* and *Irf2bp2*. The sequences determined in this study were deposited in the DDBJ/EMBL/GenBank databases under the accession numbers LC228778–LC228932.

Networks of mitochondrial and nuclear gene sequences were constructed using the Neighbour-Net (NN) method, as implemented in SPLITS TREE (ver. 4.11.3) (Huson & Bryant, 2006). For the nuclear gene sequence analyses, we assigned alleles to the subspecies groups (*c*: CAS, *d*: DOM, *m*: MUS) based on two criteria: (i) clustering patterns in the network trees and (ii) geographic origins of individual mice, with particular emphasis on mice from within the inferred natural range of each subspecies (Nunome et al., 2010).

Assessment of historical demographical processes

The ARLEQUIN 3.5 program (Excoffier & Lischer, 2010) was used for population genetic analysis. Molecular diversity indices (mean pairwise difference, haplotype diversity: *Hd*; nucleotide diversity: π) were calculated for each sublineage. Neutrality of sequence variation was tested using Tajima's *D* (Tajima, 1989) and Fu's *F* tests (Fu, 1997). The pairwise mismatch distributions (Rogers & Harpending, 1992), which comprise the pairwise differences among all individuals of each clade, were compared using the simulated sudden expansion model, and population demographic parameters were estimated. Datasets of the mtDNA sequences (4,225 bp) were used to assess the temporal aspect of rapid expansion using the formula $t = \tau/2u$, where t is the time since expansion in generations, τ is a unit of mutational time, and u is the cumulative evolutionary rate per generation for the entire sequence (Rogers & Harpending, 1992; Rogers, 1995). The value of u was derived from the formula $u = \mu kg$, where μ is the evolutionary rate per site per year, k is the sequence length, and g is the generation time in years. Time since expansion in years, T (= tg), was estimated using the formula $T = \tau/2\mu k$. We used previously known estimates for the substitution rate per site per million years (myr) in rodents: 0.03, 0.11, 0.16 (Suzuki et al., 2015) and 0.39 (Herman & Searle,

2011) (Table 2).

Estimation of timing for hybridisation event start time

The probability P that a given haplotype did not change from its ancestor G generations ago is $P = (1 - r)^G$, where r is the recombination and mutation rate (the equation can be transformed to $G = -\ln(P)/r$) (Stephens et al., 1998; Koopman et al., 2007; Nunome et al., 2010). We took into account the overall rate of recombination in *M. musculus*, 0.52 cM/Mb (Jensen-Seaman et al., 2004), although recombinations are not uniform across the genome due to hotspots present at certain intervals, e.g. at 10–100 kb (Daly *et al*., 2001).

RESULTS

Genetic variation of mtDNA

We conducted phylogenetic analyses with relatively longer mtDNA sequences (4,225 kb) consisting of three gene regions (*Co1*, *Cytb* and *Nd2* and adjacent *tRNAs*), focusing on the two phylogroups (CAS-1a and MUS-1c) that are known to have reached Japan through prehistorical human movement (Suzuki et al., 2013). For these analyses, 78 mice were used in total (Table 1). An NN analysis on the CAS-1 haplotypes (n = 53), which are thought to have dispersed to wide areas of eastern Asia (Suzuki et al., 2013), exhibited a star-like structure, confirming our previous hypothesis (Fig. 1B). The concatenated network further supported previous *Cytb* analyses showing secondary emergence of the phylogroups comprising haplotypes from South China, Japan and south Sakhalin, with a star-like structure (CAS-1a; Suzuki et al., 2013). In the CAS haplotypes recovered from Japan, one exceptional case was the haplotype from Otaru (Locality 25 in Fig. 1A), Hokkaido, which was not included in the CAS-1a sub-phylogroup but was included in the remaining CAS-1 cluster, here termed CAS-1b (Fig. 1B). Two individuals from Sakhalin (Yuzhno-Sakhalinsk, Locality 12) occupied different parts of the network, CAS-1a and CAS-1b, the latter of which was identical to individuals from Primorye, Russia (Localities 13 and 15).

The 17 MUS haplotypes recovered from 26 individual mice were divided into two subclades, MUS-1 and MUS-2, in the NN network (Fig. 1C). In the MUS-1 phylogroup, 14 individuals from Korean and Japanese mice showed a close relationship, forming a sub-phylogroup that we have termed MUS-1c, as shown in our previous

study (Suzuki et al., 2013).

Mismatch distribution analysis did not disprove the sudden expansion model (Rogers & Harpending, 1992) for phylogroups CAS-1 (data not shown), CAS-1a, CAS-1b and MUS-1c (Fig. 1D). The neutrality tests (Tajima's *D* and Fu's F_S) were significantly negative for CAS-1, CAS-1a, CAS-1b and MUS-1c (Table 2). We estimated expansion times using possible substitution rates of 0.03, 0.11, 0.16 (Suzuki et al., 2015) and 0.39 (Herman & Searle, 2011) (Table 2). The expansion times of CAS-1, CAS-1a, CAS-1b and MUS-1c were estimated to be approximately 6,500, 4,000, 6,700 and 2,100 years ago, respectively, under an evolutionary rate of 0.11 substitutions/site/myr (Suzuki et al., 2015).

Assessment of genetic constitutions of nuclear genomes

We determined nucleotide sequences in 13 gene regions at the distal part of mouse chromosome 8. The network construction in each of the genes examined (approximately 500 bp) exhibited clustering patterns of apparent phylogeographic significance, allowing subspecies assignment in most cases (Fig. 2B). Alleles were designated as "*m*", "*c*" and "*d*", representing the subspecies groups MUS, CAS and DOM, respectively, followed by a number (e.g. *m1*, *m2*, etc.) (Table 3). In the *Pgbd5* and *Sipa1l2* networks, individuals from CAS and DOM territories shared the same allele. These alleles were labelled as unknown (*u*); this may conceivably point to a shared ancestral allele. In the *Cdt1* and *Cbfa2t3* networks, haplotypes (*u1*, filled arrows in Fig. 2B) recovered from Hokkaido (Onuma, Locality 30 and Hakodate, Locality 31) differed from haplotypes belonging to MUS and CAS (open arrows) commonly seen in Japan and were regarded as unknown haplotypes. The unknown haplotype (*u1*) for *Cbfa2t3* was seen in the mouse from Nepal (Kathmandu, Locality 60).

Haplotype assessment

Allelic combinations, with respect to the 13 gene regions of *Zcchc14*, *Cdt1*, *Cbfa2t3*, *Acsf3*, *Ankrd11*, *Cdk10*, *Fanca*, *Tcf25*, *Dbndd1*, *Rhou*, *Pgbd5*, *Sipa1l2* and *Irf2bp2*, were assessed with a Bayesian method using PHASE (Table 3). In Japan, CAS alleles were recovered from Hokkaido, northern Honshu and Kyushu, and DOM alleles were from Hokkaido and eastern Honshu (Table 3); notably, all of these were chimeric haplotypes, with combinations of MUS, CAS and DOM alleles.

We assessed inter-subspecies hybridisation events in wild mice with 1 Mb and 5 Mb tracts, focusing on the Japanese populations. The 1 Mb tract system with nine markers and 200 kb intervals revealed approximate lengths for the shortened CAS fragments in Hokkaido (152 kb, n = 25) and northern Honshu (211 kb, n = 9) of approximately 170 kb (n = 34), ranging from 100 to 400 kb (Table 3). The time elapsed after the hybridisation events began in northern Japan was calculated to be 815 generations, assuming a continuous heterozygous state for CAS fragments across generations. The predicted lengths of CAS fragments shown in mice from Atsugi (central Honshu) and Kagoshima (Kyushu) were 500 kb and 400 kb, respectively. The generation times for backcrossing events were calculated to be 250 and 300, respectively. Relatively large DOM fragments were recovered from sporadic localities in Kushiro (3–5 Mb), Kyowa (2 Mb) and Atsugi (5 Mb), and generation times after backcrossing started were calculated to be 45–28, 70 and 28, respectively.

DISCUSSION

Origins of the major and minor components of the mtDNA lineages

Japanese wild mice possess the mtDNA lineage of the south Asian subspecies, *M. m. castaneus* (CAS) as the second major component, together with that of the predominant lineage of the north Eurasian subspecies of *M. m. musculus* (MUS) (Yonekawa et al., 1988; Terashima et al., 2006; Suzuki et al., 2013). The CAS group, covering broad areas of East Asia and South Asia, is known to possess distinct sublineages of mtDNA (i.e., CAS-1–4); however, only one of these, CAS-1, is likely to have achieved the broad geographic range of Southeast Asia, Indonesia and East Asia, presumably in association with prehistorical human movement (Suzuki et al., 2013). In our previous study using *Cytb* (1,140 bp), we found prominent clusters of CAS-1 and its sub-lineage, CAS-1a, which was shared by mice from South China, Japan and southern Sakhalin (Suzuki et al., 2013). In this study, phylogenetic analyses using longer sequences (4,225 bp) provided evidence for rapid expansion of both CAS-1 and CAS-1a (Tables 1 and 2). This result indicates that the Japanese CAS mtDNA results from two rapid historical expansion events, the first occurring on the continent and peripheral islands of Sri Lanka and Indonesia (CAS-1) and the second occurring within the continental part of South China and the Japanese Islands (CAS-1a). Accounting for the

occurrence of a CAS-1a haplotype in Kyushu, southern Japan, near southern China, it is conceivable that the ancestral lineage of CAS-1a came from South China to the Japanese Islands, via Kyushu as the entry point.

We estimated the start of the expansion of CAS-1 and CAS-1a based on the obtained τ values, using four possible options for the mtDNA evolutionary rate – 0.03, 0.11, 0.16 and 0.39 (Table 2) – from studies on wood mice (Suzuki et al., 2015) and voles (Herman & Searle, 2011), thereby accounting for time-dependent rates (Ho & Larson, 2006). The use of relatively higher rates (e.g. 0.11 substitutions/site/myr) has been recommended when comparing recent divergence events, e.g. within the last 10,000 years, and lower rates (e.g. 0.03 substitutions/site/myr) for older periods, such as 130,000 years ago (Suzuki et al., 2015; Hanazaki et al., 2017). In fact, the time with the lowest estimated evolutionary rate, of 0.03 substitutions/site/myr, was 25,000 years ago (Table 2); assuming that the broad geographic expansion of CAS-1 included northern China (Suzuki et al., 2013) during the greatest glacial maximum is therefore unrealistic. Using time estimates of 1,100 years ago from the relatively higher rate of 0.39 substitutions/site/myr does not explain the rapid expansion events of CAS-1a mice in South China (Jing et al., 2014) and Japan (Suzuki et al., 2013). The use of an evolutionary rate of 0.11 substitutions/site/myr (Suzuki et al., 2015) likely provides more reasonable estimates for the expansion events of CAS-1 and CAS-1a; the first CAS-1 expansion occurred approximately 7,000 years ago and the second, predicted for CAS-1a, was approximately 4,000 years ago (Table 2). These estimates are consistent with the well-described ancient agricultural development that occurred in Southeast Asia and East Asia by 8,000 years ago (Khush, 1997; Londo et al., 2006; Zheng et al., 2009; Zhang et al., 2012; Larson et al., 2014; Fuller et al., 2014; Jing et al., 2014; Silva et al., 2015). Jing et al. (2014) examined CAS mtDNA sequences in mice from South China and provided a preferable estimate of 4,650–9,300 years ago for the start of the expansion event. Recent genetic and archaeological evidence has suggested that rice cultivation first emerged along the Pearl River (Guangxi province, here represented by mice from Guilin) in southern China (Huang et al., 2012) and developed along the upper Yangtze river (e.g. Yunnan province, here represented by mice from Kunming) by approximately 4,500 years ago (Fuller et al., 2014; Silva et al., 2015). Accordingly, this may be related to the recent finding that historical admixing between peoples of the

Asian continent and Japanese Islands must have occurred during the Jomon period 5,000–6,000 years ago (Nakagome et al., 2015).

Japanese wild mice possess a specific type of mtDNA sublineage belonging to *M. m. musculus* (MUS), termed MUS-1c, that is considered to have been introduced from the Korean Peninsula (Suzuki et al., 2013). Our major concern was to assess the timing of this introduction event, which has not yet been discovered in previous studies. Our current study using longer sequences indicated that the 12 Japanese haplotypes formed a cluster with 3 haplotypes from Korea in the network analysis (Fig. 1B). The results from neutrality tests and mismatch distributions for MUS-1c tend to support recent demographic expansion (Fig. 1D, Table 2). The use of an evolutionary rate of 0.11 substitutions/site/myr and the τ value obtained (τ = 2.0) suggest recent expansion events in the Korean Peninsula and Japanese Islands approximately 2,000 years ago (Table 2). This result supports the general assessment (e.g. Suzuki et al., 2013) that MUS-1c was introduced to Japan in association with the historical migration of the Yayoi People via the Korean Peninsula, which is believed to have occurred 2,000–3,000 years ago (e.g. Hanihara, 1991; Jinam, Kanzawa-Kiriyama & Saitou, 2015; Nakagome et al., 2015).

Estimation of timing of multiple hybridisation events in Japanese wild mice

We assessed the genomic consequences of inter-subspecies genetic hybridisation of the house mouse in Japan, namely admixing between CAS and MUS. Since the house mouse is subjected to stowaway introduction, we needed to determine the genomic components that would have been introduced during ancient and recent times. Generally, introduced nuclear genomic segments have been subjected to fragmentation through meiotic recombination from generation to generation (Stephens et al., 1998). We performed haplotype structure analysis by monitoring the lengths of the subspecies-specific fragments via the 1 Mb and 5 Mb tracts (Table 3).

A comparison of 425 sequences covering the gene array from *Cdt1* to *Rhou* (1 Mb tracts), where subspecies assignments were successful, showed that house mice in the Japanese Islands comprised three distinct components: MUS (75.8%), CAS (15.8%) and DOM (8.5%). Contrary to the result of the mtDNA analyses mentioned above, which did not indicate the appearance of DOM haplotypes, we detected substantial allelic sequences of DOM, which is currently dominant in West Europe, America and

Oceania.

The marked long DOM segments (2–5 Mb) were recovered from sampling localities near human dwellings, namely a port (Kushiro) and rice fields (Kyowa, Atsugi) (Table 3), implying contemporary stowaway DOM introductions in Japan, as has been previously predicted (Minezawa et al., 1979; Yonekawa et al., 1988; Bonhomme et al., 1989; Tsuda et al., 2007). Such DOM segments are considered to have been sporadically introduced into Japan by single individuals in each locality and thus are expected to be heterozygous, with predominantly MUS segments in each generation. Assuming a recombination rate of 0.52 cM/Mb, elapsed generation times underlying the long DOM fragments of 2–5 Mb are estimated to be 70–28 and 23–10 years, respectively, assuming generation times of one and three per year. These considerations suggest that stowaway introduction of DOM mice and the introgression of DOM elements are ongoing.

The main aim of this study was to assess historical hybridisation events between CAS and MUS in the Japanese Islands. On the basis of the mtDNA study mentioned above, the efficient demographic expansion of CAS mice in Japan occurred approximately 4,000 years ago, in association with the rapid expansion of mice in the coastal area of the Yangtze River and their geographic expansion to Japan, colonising from Kyushu through Honshu and Hokkaido, and ultimately to Sakhalin. Subsequently, MUS mice entered Japan via the Korean Peninsula approximately 2,000 years ago. Our nuclear DNA analyses showed that the majority of the genome of mice from Kyushu and western and central Honshu consists of MUS (Table 3), implying that the introduction of MUS was effective, resulting in the replacement of CAS with MUS in this geographic region of the habitat. This conclusion is consistent with the skeletal morphology of the Japanese people, indicating that there are marked influences from the continent on the human populations of western Japan, whereas the genetic continuity of the Jomon people is apparent in eastern Japan (Hanihara, 1991).

In northern Japan, in contrast, it is evident from the results of our nuclear DNA analyses (Table 3) that the two subspecies lineages CAS and MUS are subject to ongoing genetic hybridisation, contrary to our initial prediction from the resultant mtDNA, in which the northern and southern parts of Hokkaido are now inhabited exclusively by CAS and MUS, respectively (Fig. 1A; Terashima et al., 2006). In northern Japan, the 1 Mb tract analysis disclosed that the CAS segments are short, i.e.

170 kb in length on average (Table 3). Accounting for the shortened CAS segment (170 kb on average), a predicted rodent recombination rate of 0.52 cM/Mb for rodents (Jensen-Seaman et al., 2004), the elapsed generation time following the beginning of hybridisation between CAS and MUS is estimated to be 815 generations.

Contrary to our initial expectation, we observed long CAS fragments in central Honshu (Atsugi) and southern Kyushu (Kagoshima), of 500 kb (R21) and 400 kb (R24), respectively, which were similar to those obtained from the reference CAS individuals from Kunming and Taiwan (Table 3). Assuming the backcrossing mode (heterozygous every generation), the elapsed time after introduction is estimated to be 280 (one generation/year) or 90 (three generations/year) years ago. These hybridisation events were clearly relatively recent compared to their equivalents in northern Japan.

Overall, our study illustrates several interesting features of the genetic architecture of Japanese mice. The genetic components of the predominant MUS lineage are less polymorphic in wild mice; however, the genetic background is highly heterogeneous among geographic localities due to several reasons, including different admixing states with the CAS components and the influence of sporadic introductions from overseas in the modern and contemporary ages.

A third nuclear genetic component of Japanese wild mice

In the current study (Table 3), we found unique allele sequences in *Cdt1* (allelic type *u1*) and *Cbfa2t3* (*u1*) in mice from southern Hokkaido, Onuma (locality 30 in Fig. 1) and Otaru (locality 25). The sequences differ from haplotypes assigned to MUS, DOM and CAS. Notably *u1* of *Cbfa2t3* was recovered from Nepal. These results confirm that other CAS sublineage(s), differing from those now present in South China and Southeast Asia (e.g., Indonesia, Myanmar and Bangladesh) are found in mice of northern Japan as a minor component. In our previous study (Nunome et al., 2010), we detected short segments of “source-unknown CAS” in mice from northern Japan, sequences that differed from those commonly occurring in the region, where CAS expanded through prehistorical human movement (Kodama et al., 2015). Additionally, we detected short DOM components in Japanese mice from northern Honshu and Hokkaido (Nunome et al., 2010; Kodama et al., 2015). It is possible that colonisation of mice occurred in the northern part of Japan (Hokkaido and north Honshu) from some unknown region of the CAS homeland in which admixture of CAS and DOM occurred

prior to the dispersal event (see Kodama et al., 2015).

Conclusion

Our approach, of addressing haplotype structure with intermittent markers at various intervals, such as 20 kb, 200 kb and 1 Mb, is useful to infer the phylogeographic history of organisms with past and present gene introgression. Using analysis for linkage disequilibrium in the introduced fragments, together with analysis of mtDNA sequences of non-recombination traits, we improved performance in assessing the evolutionary history of species with complex secondary contact processes. From previous and current studies (Nunome et al., 2010; Kodama et al., 2015), it has been clarified that three distinct lineages, namely those from South China, the Korean Peninsula and somewhere in the CAS homeland.

Acknowledgements

We would like to thank Kazuo Moriwaki, Toshihiko Shiroishi, Kimiyuki Tsuchiya and Hiromichi Yonekawa for providing valuable comments on an early version of this manuscript. We wish to express our appreciation to Sang-Hoon Han, Naoto Hanzawa, Hidetoshi Ikeda, Mei-Lei Jin, Alexey P. Kryukov, Miwako Kusayama, Yoshifumi Matsushima, Pavel Munclinger, Robert Palmer, Peter Vogel and numerous other mouse collectors for their help in supplying the valuable samples used in this study. We thank three anonymous reviewers for their comments that helped improve the manuscript. This study was conducted with the support of a grant-in-aid for Scientific Research (C) to HS (No. 15K07177) from the Japan Society for the Promotion of Science (JSPS).

REFERENCES

Auffray JC, Marshall JT, Thaler L, Bonhomme F. 1991. Focus on the nomenclature of European species of *Mus*. *Mouse Genome* **88**: 7–8.

Bonhomme F, Catalan J, Britton-Davidian J, Chapman VM, Moriwaki K, Nevo E et al. 1984. Biochemical diversity and evolution in the genus *Mus*. *Biochemical Genetics* **22:** 275–303.

Bonhomme F, Miyashita N, Boursot P, Catalan J, Moriwaki K. 1989. Genetical variation and polyphyletic origin in Japanese *Mus musculus*. *Heredity* **63:** 299–308.

Bonhomme F, Rivals E, Orth A, Grant GR, Jeffreys AJ, Bois PR. 2007. Species-wide distribution of highly polymorphic minisatellite markers suggests past and present genetic exchanges among house mouse subspecies. *Genome Biology* **8:** R80.

Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F. 1993. The evolution of house mice. *Annual Review of Ecology and Systematics* **24:** 119–152.

Britton J, Thaler L. 1978. Evidence for the presence of two sympatric species of mice (genus *Mus*) in southern France based on biochemical genetics. *Biochemical Genetics* **16:** 213–225.

Daly M, Rioux J, Schaffner S, Hudson T, Lander E. 2001. High-resolution haplotype structure in the human genome. *Nature Genetics* **29:** 229–232.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147:** 915–925.

Fuller DQ, Sato YI, Castillo C, Qin L, Weisskopf AR, Kingwell-Banham EJ et al. 2010. Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences* **2:** 115–131.

Fuller DQ, Denham T, Arroyo-Kalin M, Lucas L, Stevens CJ, Qin L, Allaby RG, Purugganan MD. 2014. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proceedings of the National Academy of Sciences* **111**: 6147–6152.

Gabriel SI, Stevens MI, Mathias MdL, Searle JB. 2011. Of mice and ‘vonvicts’:

Origin of the Australian house mouse, *Mus musculus*. *PLoS ONE* **6:** e28622.

Hanazaki K, Tomozawa M, Kinoshita G, Suzuki Y, Yamamoto M, Irino T. 2017. Estimation of the evolutionary rates of mitochondrial DNA in two Japanese wood mouse species based on calibrations with Quaternary environmental changes. *Zoological Science* **34:** 201–210.

Hanihara K. 1991. Dual structure model for the population history of the Japanese. *Nichibunken Japan Review*, pp.1–33.

Herman JS, Searle JB. 2011. Post-glacial partitioning of mitochondrial genetic variation in the field vole. *Proceedings of the Royal Society of London Series B, Biological Sciences* **278:** 3601–3607.

Ho SYW, Larson G. 2006. Molecular clocks: when times are a-changin'. *Trends in Genetics* **22:** 79–83.

Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490:** 497–501.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23:** 254–267.

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen C-F et al. 2004. Comparative recombination rates in the rat, mouse, and human genomes. *Genome Research* **14:** 528–538.

Jing M, Yu HT, Bi X, Lai YC, Jiang W, Huang L. 2014. Phylogeography of Chinese house mice (*Mus musculus musculus*/*castaneus*): distribution, routes of colonization and geographic regions of hybridization. *Molecular Ecology* **23:** 4387–4405.

Jinam TA, Kanzawa-Kiriyama H, Saitou N. 2015. Human genetic diversity in the Japanese Archipelago: dual structure and beyond. *Genes and Genetic Systems* **90:** 147–152.

Larson G, Piperno DR, Allaby RG, Puruggana MD, Andersson Leif, Arroyo-Kalin M et al. 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences* **111:** 6139–6146.

Karn R, Orth A, Bonhomme F, Boursot P. 2002. The complex history of a gene proposed to participate in a sexual isolation mechanism in house mice. *Molecular Biology and Evolution* **19:** 462–471.

Kodama S, Nunome M, Moriwaki K, Suzuki H. 2015. Ancient onset of geographical divergence, interpopulation genetic exchange, and natural selection on the Mc1r coat-colour gene in the house mouse (*Mus musculus*). *Biological Journal of the Linnean Society* **114:** 778–794.

Koopman WJM, Li YH, Coart E, De Weg EV, Vosman B, Roldan-Ruiz I et al. 2007. Linked vs. unlinked markers: multilocus microsatellite haplotype-sharing as a tool to estimate gene flow and introgression. *Molecular Ecology* **16:** 243–256.

Khush GS. 1997. Origin dispersal cultivation and variation of rice. *Plant Molecular Biology* **35:** 25–34.

Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA. 2006. Phylogeography of Asian wild rice Oryza rufipogon reveals multiple independent domestications of cultivated rice Oryza sativa. *Proceedings of Natural Academy of Science USA* **103:** 9578–9583.

Martinsen GD, Whitham TG, Turek RJ, Keim P. 2001. Hybrid populations selectively filter gene introgression between species. *Evolution* **55:** 1325–1335.

Minezawa M, Moriwaki K, Kondo K. 1979. Geographical distribution of Hbb^p allele in the Japanese wild mouse, *Mus musculus molossinus*. *The Japanese Journal of Genetics* **54:** 165–173.

Nakagome S, Sato T, Ishida H, Hanihara T, Yamaguchi T, Kimura R et al. 2015. Model-based verification of hypotheses on the origin of modern Japanese revisited by Bayesian inference based on genome-wide SNP data. *Molecular Biology and Evolution* **32:** 1533–1543.

Nunome M, Ishimori C, Aplin KP, Tsuchiya K, Yonekawa H, Moriwaki K, Suzuki H. 2010. Detection of recombinant haplotypes in wild mice (Mus *musculus)* provides new insights into the origin of Japanese mice. *Molecular Ecology* **19:** 2474–2489.

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.

Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9:** 552–569.

Rogers AR. 1995. Genetic evidence for a Pleistocene population explosion. *Evolution* **49:** 608–615.

Sage RD. 1981. Wild mice. In: Forester HL, Small JD, Fox JG (eds) *The Mouse in Biomedical Research*. Academic Press: New York. Vol 1, pp 40–90.

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

Silva F, Stevens CJ, Weisskopf A, Castillo C, Qin L, Bevan A, Fuller DQ. 2015. Modelling the geographical origin of rice cultivation in Asia using the Rice Archaeological Database. *PloS ONE* **1:** e0137024.

Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M et al. 1998. Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. *American Journal of Human Genetics* **62:** 1507–1515.

Stephens M, Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73:** 1162–1169.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68:** 978–989.

Suzuki Y, Tomozawa M, Koizumi Y, Tsuchiya K, Suzuki H. 2015. Estimating the molecular evolutionary rates of mitochondrial genes referring to Quaternary ice age events with inferred population expansions and dispersals in Japanese *Apodemus*. *BMC Evolutionary Biology* **15:** 1.

Suzuki H, Sato JJ, Tsuchiya K, Luo J, Zhang YP, Wang YX et al. 2003. Molecular phylogeny of wood mice (*Apodemus,* Muridae) in East Asia. *Biological Journal of the Linnean Society* **80:** 469–481.

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

Suzuki H, Nunome M, Kinoshita G, Aplin KP, Vogel P, Kryukov AP 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.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585–595.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28:** 2731–2739.

Terashima M, Furusawa S, Hanzawa N, Tsuchiya K, Suyanto A, Moriwaki K *et al.* 2006. Phylogeographic origin of Hokkaido house mice (*Mus musculus*) as indicated by genetic markers with maternal, paternal and biparental inheritance. *Heredity* **96:** 128–138.

Tsuda K, Tsuchiya K, Aoki H, Iizuka S, Shimamura H, Suzuki S, et al. 2007. Risk of accidental invasion and expansion of allochthonous mice in Tokyo metropolitan coastal areas in Japan. *Genes & Genetics Systems* **82:** 421–428.

Yonekawa H, Moriwaki K, Gotoh O, Miyashita N, Matsushima N, Shi LM et al. 1988. Hybrid origin of Japanese mice "*Mus musculus molossinus*": evidence from restriction analysis of mitochondrial DNA. *Molecular Biology and Evolution* **5:** 63–78.

Zhang J, Lu H, Gu W, Wu N, Zhou K, Hu YY et al. 2012. Early mixed farming of millet and rice 7800 years ago in the Middle Yellow River region, China. *PLoS ONE* **7:** e52146.

Zheng YF, Sun GP, Qin L, Li C, Wu X, Chen X. 2009. Rice fields and modes of rice cultivation between 5000 and 2500 BC in east China. *Journal of Archaeological Science* **36:** 2609–2616.

FIGURE LEGENDS

Figure 1.

Assessment of genetic variation using mitochondrial gene sequences (4,225 bp). A, collection localities and mitochondrial genotypes in Eurasia of *Mus musculus* samples, representing three major mitochondrial subspecies lineages of *Mus musculus musculus* (MUS), *M. m. castaneus* (CAS) and that from Nepal (NEP). Detailed locality names and sample codes are listed in Table 1. Names of sublineages of MUS and CAS are in Suzuki et al. (2013). Localities only examined for nuclear DNA analysis are marked with an asterisk. B, a neighbour net (NN) network of the phylogroup CAS-1. C, a NN network of the phylogroup MUS. D, mismatch distribution of three clusters of CAS-1, CAS-1a and MUS-1c. The bar indicates observed frequency, and a line denotes the expected frequency under the sudden expansion model. SSD, sum of squared deviations; *r*, Harpending's raggedness index.

Figure 2.

Positions of analysed regions (open triangle) on chromosome 8 (A) and network trees with resultant allelic sequences of the 13 genes of *Zcchc14*, *Cdt1*, *Cbfa2t3*, *Acsf3*, *Ankrd11*, *Cdk10*, *Fanca*, *Tcf25*, *Dbndd1*, *Rhou*, *Pgbd5*, *Sipa1l2* and *Irf2bp2* (B). Network trees were constructed in each data set with the NN method. Assessment for the subspecies groups of MUS (blue circle), CAS (yellow circle) and DOM (red circle) were tentatively done based on the geographic origins of alleles. Alleles unassigned to subspecies group are shown with grey circles for the network trees of *Cdt1*, *Cbfa2t3*, *Fanca*, *Tcf25* and *Sipa1l2*. In *Cdt1* and *Cbfa2t3*, unassigned allelic and CAS sequences observed in mice from Japan are marked with black and white arrows, respectively. The bars indicate the genetic distances. Detailed information of allelic arrangements in haplotypes is described in Table 3.

Table 1. List of samples used in this study and mtDNA phylogroups

	Locality	DNA code	mtDNA
1	Canada: Pegion	PGN2/Ms	-
2	Germany: Weidesgrun	MB8/HS3958	-
3	France: Montpellier	BFM/2Ms	-
4	Czech Republic: Lhotka	PWK	-
5	Bulgaria: Toshevo	BLG2/Ms	-
6	Ukraine: Donetsk	MG3065	MUS-1
7	Kazakhstan: Aktyubinsk	HS1464	MUS-2
8	Uzbekistan: Tashkent	HS1338	MUS-2
9	Russia: Gorno-Altaysk	HS3605	MUS-2
10	Russia: Irkutsk	HS3608	MUS-2
11	Russia: Amurskii	MG3041	MUS-2
		MG3042	MUS-2
12	Russia: Yuzhno-Sakhalinsk	MG3046	CAS-1b
		MG3047	CAS-1a
13	Russia: Vladivostok	MG3023	CAS-1b
		MG3077	CAS-1b
14	Russia: Kraskino	HS1413	MUS-1
15	Russia: Khasan	HS1411	CAS-1b
16	Hokkaido, Japan: Nayoro	HS2326	CAS-1a
17	Hokkaido, Japan: Asahikawa	HS1947	CAS-1a
18	Hokkaido, Japan: Fukagawa	HS2324	CAS-1a
19	Hokkaido, Japan: Kitami	HS4977	CAS-1a
20	Hokkaido, Japan: Takikawa	HS2445	CAS-1a
21	Hokkaido, Japan: Kushiro	HS1946	CAS-1a
		HS2402	CAS-1a
		HS2422	CAS-1a
22	Hokkaido, Japan: Obihiro	HS2781	CAS-1a
23	Hokkaido, Japan: Naganuma	HS2446	CAS-1a
24	Hokkaido, Japan: Tobetsu	HS4976	CAS-1a
25	Hokkaido, Japan: Otaru	HS2340	CAS-1b
26	Hokkaido, Japan: Kyowa	HS2779	MUS-1c
		HS2780	MUS-1c
		HS4975	MUS-1c
27	Hokkaido, Japan: Date	HS2451	MUS-1c
28	Hokkaido, Japan: Setana	HS4271	MUS-1c
29	Hokkaido, Japan: Okushiri I.	HS298	MUS-1c
30	Hokkaido, Japan: Onuma	HS394	MUS-1c
31	Hokkaido, Japan: Hakodate	HS2323	MUS-1c
32	Honshu, Japan: Otsuchi	HS2454	CAS-1a
33	Honshu, Japan: Sakata	HS2457	-
34	Honshu, Japan: Sakekawa	HS2461	-
35	Honshu, Japan: Tendo	HS2458	-
36	Honshu, Japan: Sendai	HS2456	CAS-1a
37	Honshu, Japan: Aizuwakamatsu	MG488	CAS-1a
		MG489	CAS-1a
38	Honshu, Japan: Atsugi	HS3839	MUS-1c
		HS3840	MUS-1c
39	Honshu, Japan: Mishima	MSM/Ms	-
40	Honshu, Japan: Misasa	HS4097	MUS-1c
41	Kyushu, Japan: Miyazaki	HS2472	-
42	Kyushu, Japan: Kagoshima	HS3603	CAS-1a
43	Korea: Hwacheon-gun	HS4234	MUS-1c
		HS4235	MUS-1c
44	Korea: Baengnyeong I.	MG682	MUS-1
45	Korea: Gangwa Island	HS4238	MUS-1c
46	Korea: Busan	HS3540	-
47	China: Laiyang	MG963	MUS-2
48	China: Jinan	MG928	MUS-1
49	China: Manzhouli	MG863	CAS-1b
50	China: Manasi	MG597	MUS-2
51	China: Shanghai	MG438	CAS-1b
52	China: Ningbo	MG786	CAS-1b
53	China: Guilin	MG501	CAS-1a
54	China: Guanzhou	MG503	CAS-1b
		MG504	CAS-1b
55	China: Kunming	HS506	CAS-1a
		HS507	CAS-1a
		HS508	CAS-1a
		HS512	CAS-1a
56	China: Hutiaoxia	MG925	CAS-1a
		MG926	CAS-1a
57	China: Lijiang	MG916	CAS-1a
		MG917	CAS-1a
58	China: Dali	MG787	CAS-1a
		MG788	CAS-1a
59	Taiwan: Taitung	HS2400	-
60	Nepal: Kathmandu	HS1467	-
61	India: Delhi	HI186	CAS-1b
		HI187	-
62	India: Bhubaneswer	HI302	CAS-1b
63	Sri Lanka: Peradeniya	HI481	CAS-1b
64	Bangladesh: Comilla Distri	HS2925	CAS-1b
		HS3357	CAS-1b
		HS3689	CAS-1b
		HS3701	CAS-1b
65	Myanmar: Lashio	HS3721	CAS-1b
66	Myanmar: Mount Popa	HS3720	CAS-1b
67	Philippines: Caterman	HI196	CAS-1b
68	Indonesia: Bogor	HI111	CAS-1b
69	Indonesia: Lembang	HI134	CAS-1b
70	Indonesia: Bali Island	HI116	CAS-1b
71	Indonesia: Flores Island	HS3736	CAS-1b

Samples used for the nuclear gene analyeses are underlined.

Table 2. Standard genetic information of concatenated mitochondrial DNA haplotypes (*Nd2* to *Co1* and *Cytb*; 4,225 bp) and estimation of population expansion times (year before present) with three options of the evolutionary rates (μ). *N* (sample size), *S* (number of sites with substitutions), *h* (haplotype number), *Hd* (haplotype diversity), π (nucleotide diversity), Tajima's *D*, Fu's *Fs*, and τ were calculated using ARLEQUIN 3.5 (Excoffier & Lischer, 2010).

Haplotype									Estimated expansion time with μ (substitutions/site/myr)		
group	*N*	*S*	*h*	*Hd*	π (%)	Tajima's *D*	Fu's *Fs*	τ	0.03	0.11	0.16	0.39
CAS-1	52	75	36	0.968	0.1295	-2.347**	-25.347**	6.0781	23977	**6539**	4495	1844
CAS-1a	28	24	16	0.892	0.0747	-1.754*	-8.312**	3.6875	14546	**3967**	2727	1119
CAS-1b	24	53	20	0.974	0.1391	-2.278**	-11.795**	6.2090	24493	**6680**	4592	1884
MUS	26	73	17	0.951	0.1746	-0.353	-0.172	-				
MUS-1c	14	8	8	0.890	0.0463	-0.8388	-3.4226**	1.9921	7858	**2143**	1473	604
MUS-2	8	32	5	0.786	0.0943	-1.254	2.244	-				

* Significant at $P < 0.05$

** Significant at $P < 0.01$

The options of the evolutionary rates of the three slower ones and the heighest one were referred to Suzuki et al. (2015) and Herman & Searle (2011), respectively.

Supplemental Table S1. List of primers used in this study

Gene and primer code	Sequence (Reference)	Posi. 3' end*	Size (bp) Exon	Intron	Total	Cycle condition
Mitochondrial DNA						
Nd2 (NADH dehydrogenase 2) to *Co1* (cytochrome c oxidase I)			-	-	3,085	
Nd2-Co1_F1 (first primer)	TCTCCGTGCTACCTAAACACC	3834				95 °C (30 sec), 50 °C (30 sec), and 72 °C (60 sec)
Nd2-Co1_R1 (first primer)	GGAAGGCCTCCTAGGGATAG	4658				
Nd2-Co1_F2 (second primer)	TCCTTACAACCCATCCCTCA	4514				95 °C (30 sec), 50 °C (30 sec), and 72 °C (60 sec)
Nd2-Co1_R2 (second primer)	GAGGGTTCCGATATCTTTGTGA	5360				
Nd2-Co1_F3 (third primer)	GCAATTCGACATGAATATCACC	5249				95 °C (30 sec), 50 °C (30 sec), and 72 °C (90 sec)
Nd2-Co1_R3 (third primer)	TGAAGCAAAGGCCTCTCAAA	6724				
Nd2-Co1_F4 (fourth primer)	GAGCCCACCACATATTCACA	6209				95 °C (30 sec), 50 °C (30 sec), and 72 °C (60 sec)
Nd2-Co1_R4 (fourth primer)	TGGAATGGGTAGGCCATATAA	7009				
Cytb (cytochrome b)			-	-	1,040	
Cytb#247 (upper half)	GACATGAAAAATCATCGTTG (Suzuki *et al*., 2004)	14121				95 °C (30 sec), 50 °C (30 sec), and 72 °C (30 sec)
Cytb#956 (upper half)	GATTGTATAGTAGGGATGAAATGG (Suzuki *et al*., 2004)	14799				
Cytb#955 (lower half)	CCTATCAGCCATCCCATATATTGG (Suzuki *et al*., 2004)	14614				95 °C (30 sec), 50 °C (30 sec), and 72 °C (30 sec)
Cytb#248 (lower half)	GTTTACAAGACCAGAGTAAT (Suzuki *et al*., 2004)	15306				
Nuclear DNA						
Zcchc14 (zinc finger, CCHC domain containing 14)						
Zcchc14_F	ACCATGGGAGCAAAGAAGAA	121606904	5	543	548	95 °C (30 sec), 57 °C (30 sec), and 72 °C (30 sec)
Zcchc14_R	GGCACCGTCTCTCTGACTTC	121606314				
Cdt1 (chromatin licensing and DNA replication factor 1)						
Cdt1_F	ACAGAGAAGCTCACCACTGC	122571386	169 (1-16, 283-435)	343 (17-282, 435-512)	512	95 °C (30 sec), 57 °C (30 sec), and 72 °C (30 sec)
Cdt1_R	AGCTGCTTCTGGACCTCCTT	122571946				
Cbfa2t3 (core-binding factor, runt domain, alpha subunit 2, translocated to, 3)						
Cbfa2t3_F	GGAACTACCCTTCCCAGAGG	122697845	0	508	508	95 °C (30 sec), 57 °C (30 sec), and 72 °C (30 sec)
Cbfa2t3_R	GTGAACCCAGCTTACGGTGT	122697287				
Acsf3 (acyl-CoA synthetase family member 3)						
Acsf3_F	CCGTGTTCAAGGATGCTAGG	122813023	75	431	506	95 °C (30 sec), 57 °C (30 sec), and 72 °C (30 sec)
Acsf3_R	GCCCCATGTCATATCAGGAA	122813578				
Ankrd11 (ankyrin repeat domain 11)						
Ankrd11_F	CATTTCCTCTCCGACGTGAC	123042375	0	615	615	95 °C (30 sec), 59 °C (30 sec), and 72 °C (30 sec)
Ankrd11_R	GAGCCCTTCCTTAGCCTCTC	123041718				
Cdk10 (cyclin-dependent kinase 10)						
Cdk10_F	CCTGCACAGGAACTTCATCA	123228389	0	414	414	95 °C (30 sec), 60 °C (30 sec), and 72 °C (30 sec)
Cdk10_R	TCATGAGCAAGTTGGACACC	123228857				
Fanca (Fanconi anaemia, complementation group A)						
Fanca_F	GCAGACCGGTGTTCCAGACGCT (Nunome et al., 2010)	123318545	145 (1–48, 352–448)	303 (49–351)	448	95 °C (30 sec), 57 °C (30 sec), and 72 °C (30 sec)
Fanca_R	CTCAGCCAGGACAACTTCCTCT (Nunome et al., 2010)	123319319				
Tcf25 (transcription factor 25 (basic helix-loop-helix))						
Tcf25_F1	TCCAGACAAGCCCCTATCATGT (Nunome et al., 2010)	123390807	0	613	613	95 °C (30 sec), 57 °C (30 sec), and 72 °C (30 sec)
Tcf25_R1	TCCATGCTGTACAGGGCTCTCT (Nunome et al., 2010)	123391581				
Dbndd1 (Dysbindin (dystrobrevin binding protein 1) domain containing 1)						
Dbndd1_F1	AATACCAGCACCAGGGTTCCTG (Nunome *et al*., 2010)	123509861	0	550	550	95 °C (30 sec), 57 °C (30 sec), and 72 °C (30 sec)
Dbndd1_F2	TGGCATCCCAATACCAGCACCAG** (Nunome *et al*., 2010)	123509869				
Dbndd1_R	TCAGCTCAGTGAGGTCCAGGAG (Nunome *et al*., 2010)	123509170				
Rhou (ras homolog gene family, member U)						
Rhou_F	CTACGGCCTTCGACAACTTC	123654212	0	461	461	95 °C (30 sec), 60 °C (30 sec), and 72 °C (30 sec)
Rhou_R	TAGAGACTGGCCACGGAGAC	123654757				
Pgbd5 (piggyBac transposable element derived 5)						
Pgbd5_F	GATCCTGTGGGTTCCCTTTT	124372501	0	507	507	95 °C (30 sec), 57 °C (30 sec), and 72 °C (30 sec)
Pgbd5_R	GATTTCCCCACTCCTCCTCT	124371945				
Sipa1l2 (signal-induced proliferation-associated 1 like 2)						
Sipa1l2_F	CACCAGTGGCAAAGAGTTCA	125422592	17	569	586	95 °C (30 sec), 64 °C (30 sec), and 72 °C (30 sec)
Sipa1l2_R	CTTTCCCAGTCAGTGTGGAG	125421950				
Irf2bp2 (interferon regulatory factor 2 binding protein 2)						
Irf2bp2_F	AGGCAGGTTGTTGGGTTTC	126592473	0	431	431	95 °C (30 sec), 57 °C (30 sec), and 72 °C (30 sec)
Irf2bp2_R	CTTTTCCTTGCTGTCCTTGC	126591828				
Sequence primes***						
Acs_seq_F_1A	CCTGCTACTAGGGA					
Acs_seq_F_1C	CCTGCTACTAGGGC					
Acs_seq_R_1A	ACAGCATGCAGGAA					
Acs_seq_R_1T	ACAGCATGCAGGAT					

*The positions of the 3' ends of primers were designed referring Ensemble Mouse Genome Database (http://www.ensembl.org/).

**Specific to MUS

***The sequence primes were used to determine haplotype sequences in diplotypes that had more than two indel sites (i.e., HS506, HS507, HI187).

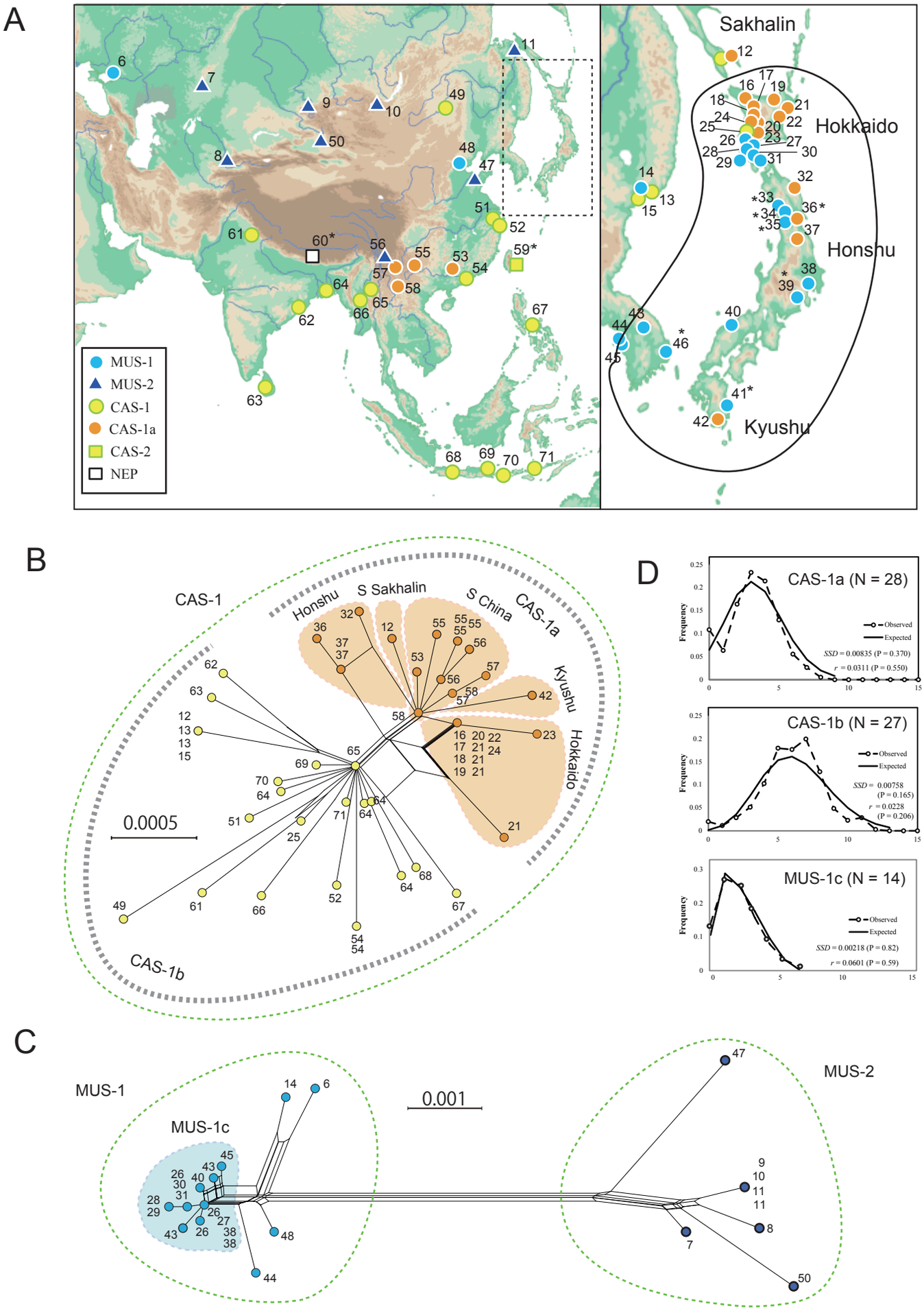
Table 3. List of samples used in this study and their allelic types and chromosomal constructs (haplotypes)

Locality*		DNA code**	Allelic assignment***												Predicted non-MUS length (kb)
			Zcc	*Cdt*							*Rho*	*Pgb*	*Sip*	*Irf*	
				Cdt	*Cbf*	*Acs*	*Ank*	*Cdk*	*Fan*	*Tcf*	*Dbn*	*Rho*				
Noth America, Europe															
1.	Pegion	PGN2/Ms	*d1*	*d1*	*d1*	*d1*	*d1*	*d1*	*d2*	*d1*	*d1*	*d1*	*u1*	*u1*	*d1*	
2.	Weidesgrun	MB9	*d2*	*d2*	*d1*	*d1*	*d1§*	*d1*	*d3§*	*d1*	*d1*	*d2§*	*d2§*	*d1*	*d2§*	
			d2	*d3*	*d2*	*d1*	*d2§*	*d1*	*d8§*	*d1*	*d1*	*d1§*	*d1§*	*d1*	*d1§*	
3.	Montpellier	BFM/2Ms	*d2*	*d2*	*d3*	*d2*	*d3*	*d2*	*d3*	*d1*	*d3*	*d1*	*u1*	*u1*	*d1*	
4.	Lhotka	PWK	*m1*	*m1*	*m1*	*m1*	*m1*	*m1*	*m1*	*m4*	*m4*	*m1*	*m1*	*u2*	*m1*	
5.	Toshevo	BLG2/Ms	*m1*	*m2*	*m1*	*m2*	*m1*	*m1*	*m1*	*m1*	*m3*	*m1*	*m2*	*m1*	*m1*	
Japan: Hokkaido															
16.	Nayoro	HS2326	*m1*	*m3*	*m1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	
			m1	*m3*	*m1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	
17.	Asahikawa	HS1947	*m1*	*m3§*	*m1§*	*m3*	*m1§*	*m1*	*m1*	*m1§*	*m10§*	*m1*	*m3§*	*u2*	*m1*	
			m1	*c1§*	*c1§*	*m3*	*c3§*	*m1*	*m1*	*c1§*	*c2§*	*m1*	*m4§*	*u2*	*m1*	100, 200, 200
21.	Kushiro	HS1946	*d2§*	*d3§*	*d3§*	*d2§*	*d1§*	*d2§*	*d6*	*d1*	*d1*	*d1*	*u1*	*u1*	*m1*	**5,000**
			m1§	*m3§*	*m1§*	*m3§*	*m1§*	*m1§*	*m1*	*m1*	*m10*	*m1*	*m3*	*m2*	*m1*	
		HS2422	*m1*	*m3*	*m1§*	*m1§*	*m2§*	*m1*	*d9*	*d1*	*d1*	*d1*	*u1*	*u1*	*m1*	**3,000**
			m1	*m3*	*c1§*	*c1§*	*m1§*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	200
22.	Obihiro	HS2781	*m1§*	*m3*	*c1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3§*	*m2§*	*m1*	100
			d2§	*m3*	*c1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m4§*	*u2§*	*m1*	100
25.	Otaru	HS2340	*c1*	*m3§*	*m1*	*m3*	*c3*	*m1*	*m1*	*m1*	*c2*	*c1*	*m3*	*u2*	*m1*	100, 200
			m1	*c1§*	*c1*	*m3*	*c3*	*m1*	*m1*	*m1*	*c2*	*c1*	*m3*	*u2*	*m1*	100, 200,200
26.	Kyowa	HS2779	*m1*	*d4§*	*m1*	*d2*	*d4*	*d2*	*d1*	*d1*	*d2*	*d1*	*u1*	*m2§*	*m1*	100, **2,000**
			m1	*m3*	*m1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2§*	*m1*	
		HS2780	*m1*	*m3*	*m1*	*c1*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	100
			m1	*m3*	*m1*	*c1*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*m2*	*m1*	100
		HS4975	*m1*	*d4§*	*m1*	*c3*	*d4*	*d2*	*d3*	*d1*	*d2*	*d1*	*u1*	*u3§*	*m1*	100, 100 **2,000**
			m1	*m3*	*m1*	*m1*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2§*	*m1*	
27.	Date	HS2451	*m1*	*m3*	*c1*	*c1*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	200
			m1	*m3*	*c1*	*c1*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*m2*	*m1*	200
29.	Okushiri Island	HS298	*m1*	*m3*	*m1*	*c1*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	100
			m1	*m3*	*m1*	*c1*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	100
30.	Onuma	HS394	*m1*	*m3§*	*c1§*	*m3*	*c3§*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*m2§*	*m1*	100, 100
			m1	*u1§*	*u1§*	*m3*	*m1§*	*m1*	*m1*	*c1*	*c2*	*m1*	*u1*	*u2§*	*m1*	200
31.	Hakodate	HS2323	*m1*	*m3*	*m1*	*m3*	*m1*	*m1*	*m1*	*c1*	*c2*	*c1*	*m3*	*m2§*	*m1*	300
			m1	*u1*	*u1*	*m3*	*m1*	*m1*	*m1*	*c1*	*c2*	*c1*	*m3*	*u3§*	*m1*	300
Japan: Honshu															
32.	Otsuchi	HS2454	*m1*	*m3*		*m3*		*m1*	*m1*	*m1*	*m10*	*m1*	*m3*		*m1*	
			m1	*m3*		*m3*		*m1*	*m1*	*m1*	*m10*	*m1*	*m3*		*m1*	
33.	Sakata	HS2457	*m1*	*c4§*	*c1*	*c1§*	*c3*	*m1*	*m1*	*c1§*	*m10*	*m1*	*m3*	*u1§*	*m1*	100, 400
			c1	*m3§*	*m1*	*m3§*	*c3*	*m1*	*m1*	*m1§*	*m10*	*m1*	*m3*	*u2§*	*m1*	100
34.	Sakekawa	HS2461	*m1*	*m3*	*m1§*	*c1*		*m1*	*m1*	*c1*	*c2*	*c1*	*u1*			100, 300
			m1	*m3*	*c1§*	*c1*		*m1*	*m1*	*c1*	*c2*	*c1*	*m3*			200, 300
35.	Tendo	HS2458	*m1*	*m3§*	*m1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	
			m1	*c4§*	*m1*	*m3*	*m1*	*m1*	*m1*	*c1*	*c2*	*c1*	*m3*	*m2*	*m1*	100, 300
36.	Sendai	HS2456	*m1*	*m3*	*m1*	*m3*			*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	
			m1	*m3*	*m1*	*m3*			*m1*	*m1*	*m10*	*m1*	*m3*	*m2*	*m1*	
37.	Aizuwakamatsu	MG488	*m1*	*m3*	*m1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	
			m1	*m3*	*m1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	
		MG489	*m1*	*m3*	*m1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	
			m1	*m3*	*m1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*u2*	*m1*	
38.	Atsugi	HS3839	*m1*	*d3§*	*d3§*	*d2§*	*d1§*	*d2§*	*d1§*	*d1§*	*d1§*	*d1§*	*u1§*	*d1§*	*d1§*	**5,000**
			m1	*m3§*	*m1§*	*m3§*	*m2§*	*m1§*	*m1§*	*m1§*	*m10§*	*m1§*	*m3§*	*m2§*	*m1§*	
		HS3840	*m1*	*m3*	*m1*	*m3*	*m1*	*c1*	*c1*	*c1*	*c1*	*c1*	*m3*	*m2*	*m1*	500
			m1	*m3*	*m1*	*m3*	*m1*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*m2*	*m1*	
39.	Mishima	MSM/Ms	*m1*	*m3*	*m1*	*m3*	*m1*	*m1*	*m1*	*m4*	*m10*	*m1*	*m3*	*m2*	*m1*	
Japan: Kyushu															
41.	Miyazaki	HS2472	*m1*	*m3*	*m1*	*m3*	*m1§*	*m1*	*m1*	*m1*	*m10*	*m1*	*m3*	*m3§*	*m1*	
			m1	*m3*	*m1*	*m3*	*m2§*	*m1*	*m1*	*m1*	*c2*	*c1*	*m3*	*m2§*	*m1*	200
42.	Kagoshima	HS3603	*m1*	*c1§*	*m1*	*m3*	*m2*	*m1*	*m1*	*m1*	*m10*	*m1*	*m5*	*u1*	*m1*	100
			m1	*m3§*	*m1*	*m3*	*m2*	*m1*	*c1*	*c1*	*c1*	*c1*	*c1*	*u1*	*m1*	400
Asia (outside Japan)															
46.	Busan	HS3540	*m1§*	*m3*	*m1*	*m3*	*m1§*	*m1*	*m1*	*m1§*	*m9*	*m2*	*m3§*	*m2*	*m1*	
			c1§	*m3*	*m1*	*m3*	*m2§*	*m1*	*m1*	*m2§*	*m9*	*m2*	*m5§*	*m2*	*m1*	
55.	Kunming	HS506	*c1*	*m3*	*m1*	*c1*	*c2*	*m1*	*m1*	*c1*	*c1*	*c1*	*u1*	*u1*	*m1*	
			c1	*c1*	*m1*	*c5*	*c4*	*c1*	*c1*	*c1*	*c1*	*c1*	*c2*	*u1*	*d1*	
		HS507	*c1*	*c1*	*m1*	*c4$*	*c4*	*c1*	*m1*	*m1*	*m9*	*c1*	*u1*	*u1*	*c1*	
			c1	*c1*	*m1*	*c5$*	*c4*	*c1*	*m1*	*c1*	*c1*	*c1*	*c2*	*u1*	*m1*	
59.	Taitung	HS2400	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*u1*	*c1*	
			c1	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*c1*	*u2*	*c1*	
60.	Kathmandu	HS1467	*c1*	*c1*	*u1*	*c3§*	*c3§*	*c3§*	*u1*	*u1*	*c1*	*c1*	*c4§*	*u1*	*c3*	
			c1	*c1*	*u1*	*c1§*	*c2§*	*c2§*	*u1*	*u1*	*c1*	*c1*	*c2§*	*u4*	*c3*	
61.	Delhi	HI187	*m2*	*c2*	*c2*	*c4$*	*c1*	*d2*	*c1*	*c3$*	*c1*	*c1*	*c2*	*u1*	*c1*	
			c1	*c3*	*c2*	*c6$*	*c1*	*c2*	*u2*	*c4$*	*c1*	*c1*	*c3*	*c1*	*c2*	
71.	Flores Island	HS3736	*c1*	*c1*	*c1*	*c2§*	*m3§*	*c1*	*c1*	*c2§*	*c1*	*c1*	*c2§*	*u1*	*c1*	
			c1	*c1*	*c1*	*c1§*	*m1§*	*c1*	*c1*	*c1§*	*c1*	*c1*	*u1§*	*u1*	*c1*	

*See Table 1 for the detail of locality and serial number. **Individuals derived from laboratory strains are underlined.

***Gene markers were represented with initial three letters. Abbrebiation for each of subspecies groups: *c*, *castaneus* ; *d*, *domesticus* ; *m*, *musculus* ; *u*, unknown. The names of alleles in *Fanca*, *Tcf25*, and *Dbndd1* were referred to the previous study (Nunome et al., 2010).

§Alleles with uncertain phase in haplotype estimation


A
Sakhalin
Hokkaido
Honshu
Kyushu
MUS-1
MUS-2
CAS-1
CAS-1a
CAS-2
NEP
B
CAS-1
Honshu
S Sakhalin
S China
CAS-1a
Kyushu
Hokkaido
CAS-1b
0.0005
D
CAS-1a (N = 28)
CAS-1b (N = 27)
MUS-1c (N = 14)
Frequency
Observed
Expected
SSD = 0.00835 (P = 0.370)
r = 0.0311 (P = 0.550)
SSD = 0.00758 (P = 0.165)
r = 0.0228 (P = 0.206)
SSD = 0.00218 (P = 0.82)
r = 0.0601 (P = 0.59)
C
MUS-1
MUS-1c
MUS-2
0.001

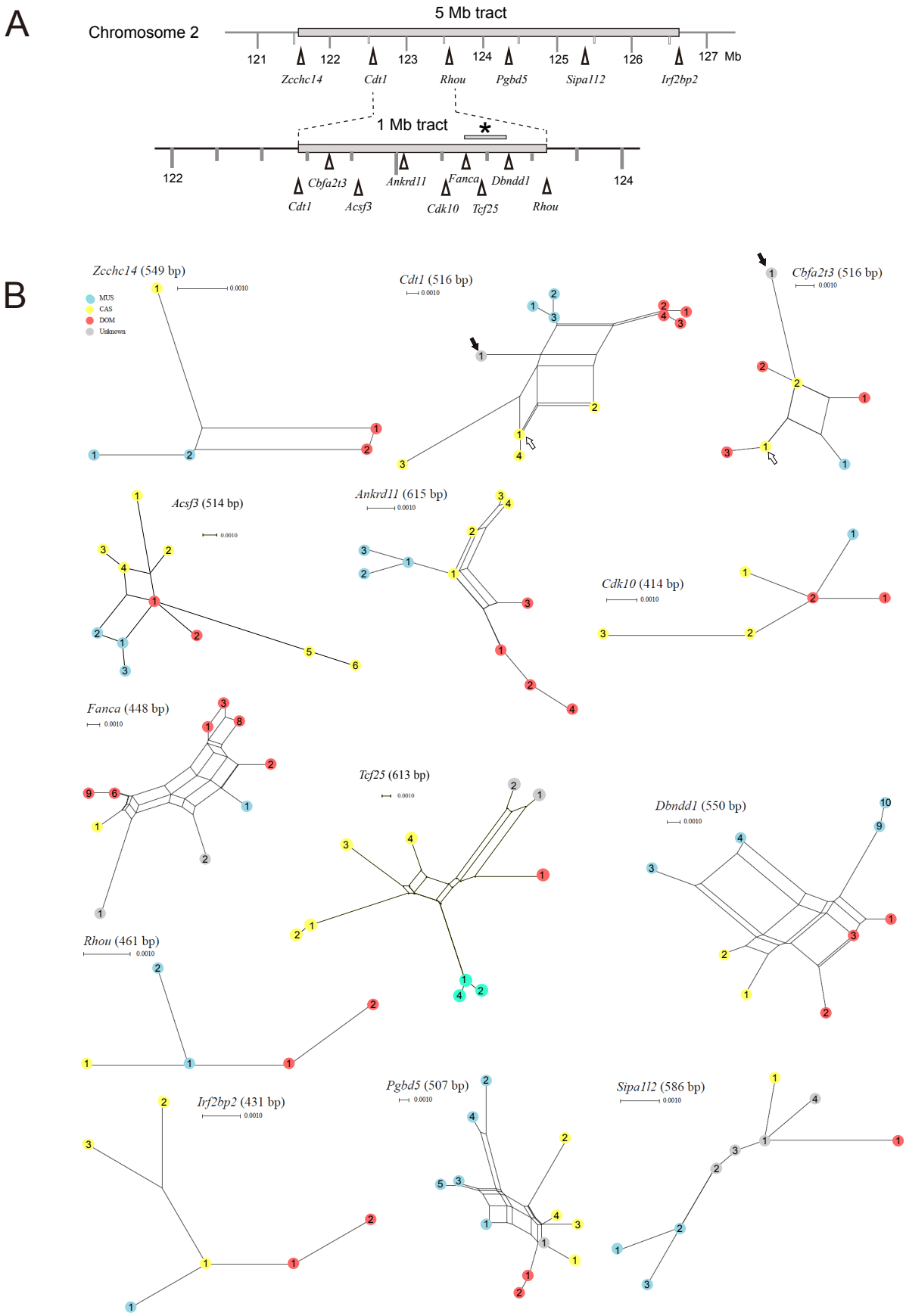

A
Chromosome 2
5 Mb tract
121
122
123
124
125
126
127 Mb
Zcchc14
Cdt1
Rhou
Pgbd5
Sipa1l2
Irf2bp2
1 Mb tract
122
124
Cbfa2t3
Ankrd11
Fanca
Dbndd1
Cdt1
Acsf3
Cdk10
Tcf25
Rhou
B
Zcchc14 (549 bp)
MUS
CAS
DOM
Unknown
Cdt1 (516 bp)
Cbfa2t3 (516 bp)
Acsf3 (514 bp)
Ankrd11 (615 bp)
Cdk10 (414 bp)
Fanca (448 bp)
Tcf25 (613 bp)
Dbndd1 (550 bp)
Rhou (461 bp)
Irf2bp2 (431 bp)
Pgbd5 (507 bp)
Sipa1l2 (586 bp)

Figure 2

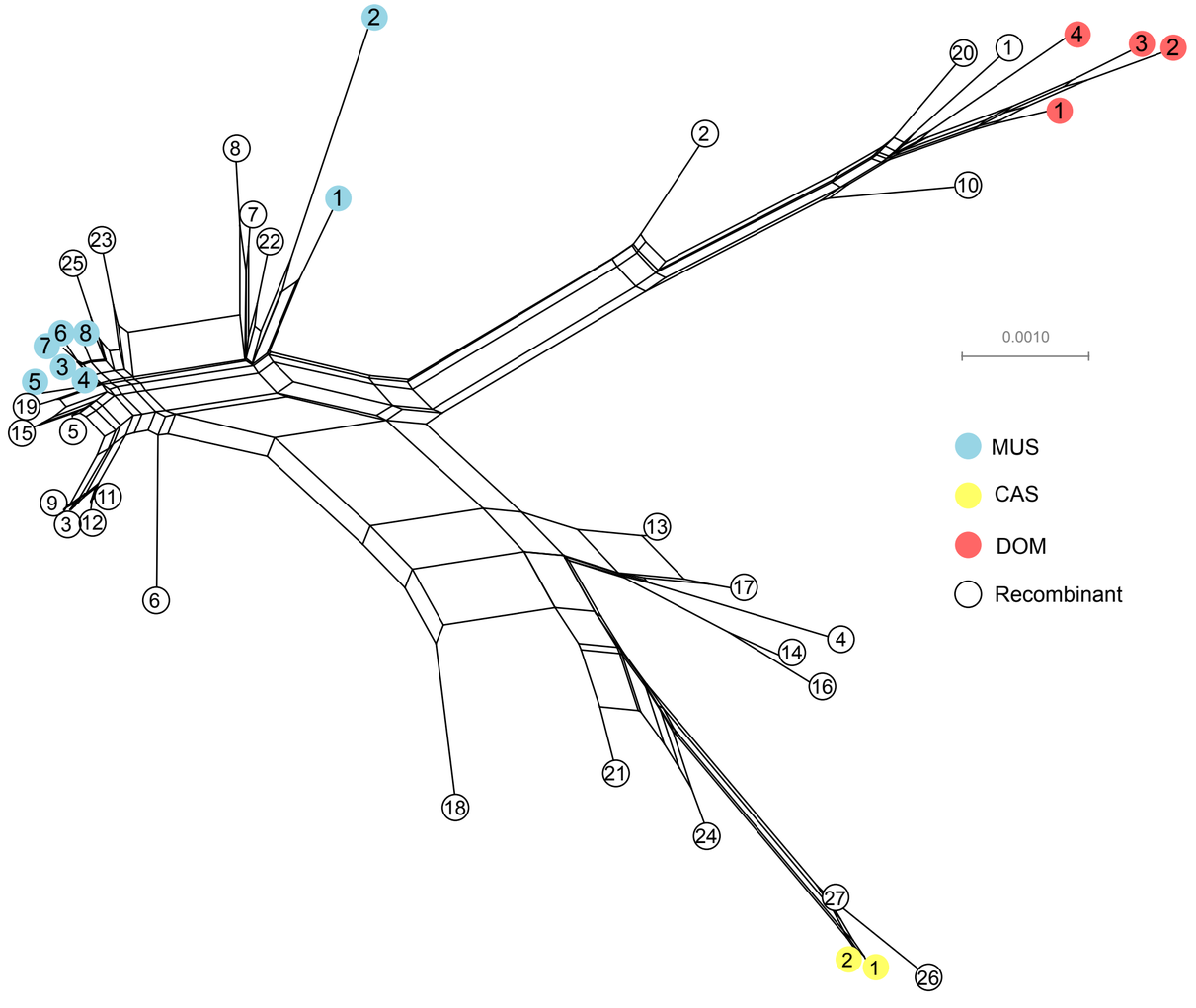
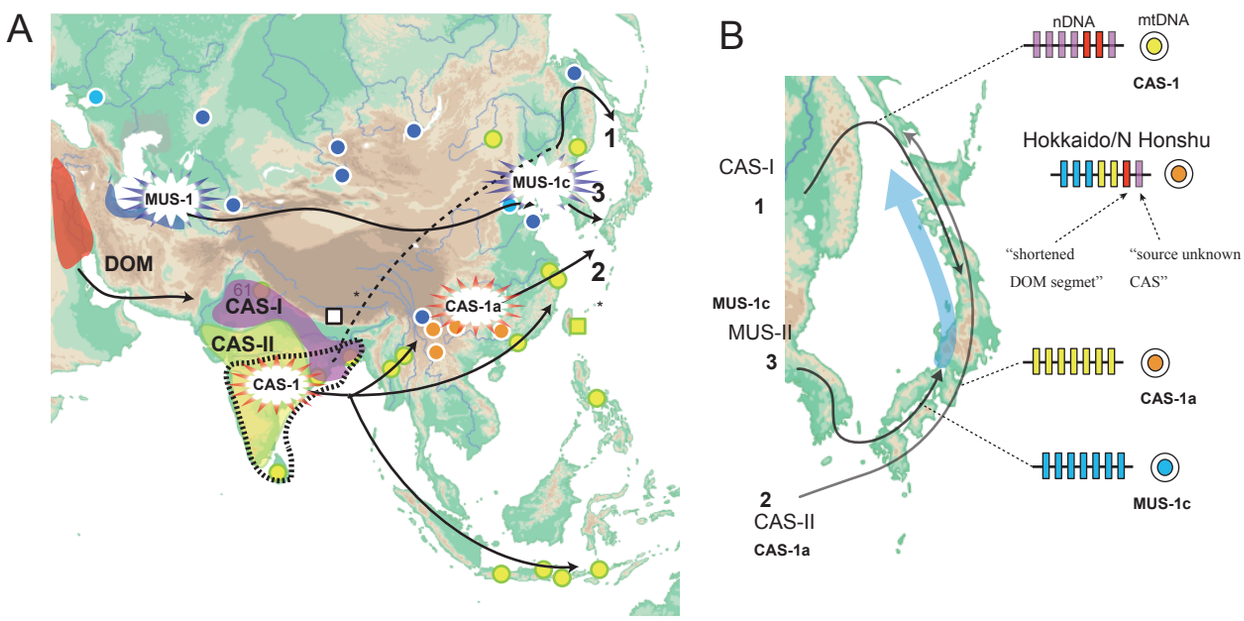


Figure 3



Kuwayama et al., Fig. 4