

Detection of recombinant haplotypes in wild mice (*Mus musculus*) provides new insights into the origin of Japanese mice

MITSUO NUNOME,* CHIKAKO ISHIMORI,* KEN P. APLIN,† KIMIYUKI TSUCHIYA,‡
HIROMICHI YONEKAWA,§ KAZUO MORIWAKI¶ and HITOSHI SUZUKI*

*Laboratory of Ecology and Genetics, Graduate School of Environmental Earth Science, Hokkaido University, Kita-ku, Sapporo 060-0810, Japan, †Australian National Wildlife Collection, CSIRO Division of Sustainable Ecosystems, GPO Box 284, Canberra, ACT 2601, Australia, ‡Laboratory of Bioresources, Ooyo Seibutsu Co., Ltd, Tokyo 107-0062, Japan, §Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113-8613, Japan, ¶RIKEN, Bioresource Center, Tsukuba, Ibaraki 305-0074, Japan

Abstract

Japanese house mice (*Mus musculus molossinus*) are thought to be a hybrid lineage derived from two prehistoric immigrants, the subspecies *M. m. musculus* of northern Eurasia and *M. m. castaneus* of South Asia. Mice of the western European subspecies *M. m. domesticus* have been detected in Japanese ports and airports only. We examined haplotype structuring of a 200 kb stretch on chromosome 8 for 59 mice from throughout Eurasia, determining short segments (≈ 370 –600 bp) of eight nuclear genes (*Fanca*, *Spire2*, *Tcf25*, *Mc1r*, *Tubb3*, *Def8*, *Afg3l1* and *Dnadd1*) which are intermittently arranged in this order. Where possible we identified the subspecies origin for individual gene alleles and then designated haplotypes for concatenated alleles. We recovered 11 haplotypes among 19 Japanese mice examined, identified either as ‘intact’ haplotypes derived from the subspecies *musculus* (57.9%), *domesticus* (7.9%), and *castaneus* (2.6%), or as ‘recombinant’ haplotypes (31.6%). We also detected recombinant haplotypes unique to Sakhalin. The complex nature of the recombinant haplotypes suggests ancient introduction of all three subspecies components into the peripheral part of Eurasia or complicated genomic admixture before the movement from source areas. ‘Intact’ *domesticus* and *castaneus* haplotypes in other Japanese wild mice imply ongoing stowaway introductions. The method has general utility for assessing the history of genetic admixture and for disclosing ongoing genetic contamination.

Keywords: intron sequences, Japanese wild mice, *Mus musculus*, phylogeography, recombination

Received 10 June 2009; revision received 28 March 2010; accepted 30 March 2010

Introduction

Sequences of mitochondrial DNA (mtDNA) and Y-linked genes are commonly used in phylogeographic studies, including those addressing populational hybridization, because they have no recombination events and thus retain the genealogical information of parental populations (Avice 2000). In contrast, nuclear gene sequences are generally thought to be unsuitable

or at least problematic for such usage, mainly due to the possible occurrence of meiotic recombination, which stands to complicate genealogical inference (Hare 2001; Zhang & Hewitt 2003). However, nuclear gene sequences are of growing popularity for phylogeographic studies of both animals (e.g. Gómez-Zurita & Vogler 2006; Taylor & Hellberg 2006) and plants (e.g. Bartish *et al.* 2006), particularly in groups for which microsatellite markers have not been developed. Meanwhile, the Human Haplotype Map (HapMap) Project has provided valuable insights into geographic distribution of genomic variation and provided convincing

Correspondence: Hitoshi Suzuki, Fax: +81-11-706-2279; E-mail: htsuzuki@ees.hokudai.ac.jp

demonstration of how inference of haplotype structure can reveal concealed demographic histories (Sabeti *et al.* 2006). Additionally, Koopman *et al.* (2007) demonstrated how analysis of recombination might be used to infer introgression within European apple populations. Here, we explore the possibility that patterns of recombination might shed new light on the history of divergent evolutionary lineages in secondary genetic contact, using as our example the house mouse, *Mus musculus*. In so doing, we also hope to further understanding of various evolutionary issues related to the origin and genetic structure of Japanese mice, including which subspecies groups have been involved in their genetic composition and how original haplotype structure has been modified by meiotic recombination through multiple generations.

Mus musculus is an ideal subject animal to explore the utility of recombination events for phylogeographic inference, partly because of the detailed knowledge of its genomic organization (e.g. Sage *et al.* 1993) and partly because it harbours several known hybrid populations (e.g. Gündüz *et al.* 2000, 2001; Guénet & Bonhomme 2003; Bonhomme *et al.* 2007). Moreover, recent genome-wide analyses with SNP markers have revealed that wild-derived inbred strains representing lineages of diverse geographic origin are genetically well-differentiated, with pervasive contrasts in haplotype structure (Abe *et al.* 2004; Frazer *et al.* 2007).

Mus musculus was once classified into more than 10 classical subspecies based on morphological traits, behaviour and geographical distribution (Schwarz & Schwarz 1943). Molecular studies of *M. musculus* favour a simpler classification into four 'subspecies groups', namely a *domesticus* subspecies group (DOM) from western Europe and North Africa (also southern Africa, Australia and the Americas, all as historical introductions), a *musculus* subspecies group (MUS) from the northern part of Eurasia excluding western Europe, a *castaneus* subspecies group (CAS) from South Asia (India and Southeast Asia) and a *bactrianus* subspecies group (BAC) from Afghanistan and Nepal (e.g. Moriwaki *et al.* 1986, 1994; Boursot *et al.* 1993; Yonekawa *et al.* 1994; Awasthi *et al.* 1998). Mitochondrial cytochrome *b* gene sequences show reciprocal monophyly and substantial divergences among the four subspecies lineages (Yonekawa *et al.* 1994; Terashima *et al.* 2006). Three of the groups (DOM, MUS and CAS) are also well-differentiated on nuclear gene datasets (Awasthi *et al.* 1998). However, CAS and BAC lineages are not clearly differentiated on nuclear genes and the BAC lineage appears from mtDNA evidence to occupy a small geographic area only. In this study, we treat all mice from southern Eurasia as a single subspecies group, CAS. Although some taxonomists have tended

to treat the house mice as a complex of closely related species (e.g. Marshall 1998), we maintain the subspecies arrangement pending further clarification of the evolutionary history of this group.

The place of origin of *M. musculus* is highly contested, depending in large part on choice of molecular marker. Analysis of nuclear encoded proteins, as assayed by allozyme electrophoresis, favours an origin in northern India where allelic diversity is highest (Boursot *et al.* 1996; Din *et al.* 1996; Guénet & Bonhomme 2003), with centrifugal migrations to the west, north and east giving rise to three recognized subspecies. Lower allelic diversity in these peripheral populations is suggestive of genetic differentiation through drift. An alternative scenario, suggested from analyses of mitochondrial DNA and summarized most completely by Prager *et al.* (1998), posits a primary origin in south central Asia, within the current range of *M. m. domesticus*, followed by an eastward and northward expansion and allopatric differentiation of CAS and MUS lineages in South Asia and Europe, respectively. Subsequent dispersal of the CAS lineage through South and Southeast Asia is thought to predate spread of the MUS lineage through the northern part of Asia (Yonekawa *et al.* 1980, 1981, 1982, 1986). Under this scenario, the current DOM lineage is derived from an ancestral stock that dispersed into western Europe.

Mus musculus is not native to the Japanese Islands but arrived there during late prehistoric times. Japanese wild mice, sometimes distinguished as *M. musculus molossinus*, were domesticated to produce 'fancy mice' and these provided founding stock for some of the earliest laboratory strains. Partly for this reason, Japanese mice have been subjected to intensive study using various molecular markers (Minezawa *et al.* 1979, 1981; Suzuki *et al.* 1986; Yonekawa *et al.* 1986, 1988, 1994; Bonhomme *et al.* 1989; Sakai *et al.* 2005; Terashima *et al.* 2006). Much attention has focused on an unusual geographic pattern in mitochondrial DNA—populations in the southern part of Japan possess MUS group haplotypes, while those in the northern part of Japan (northern Honshu and Hokkaido) possess CAS haplotypes (Yonekawa *et al.* 1986; Terashima *et al.* 2006). Japanese mice are thus thought to be the product of hybridization between MUS and CAS (Yonekawa *et al.* 1986, 1988), although the historical context of introductions and hybridization of Japanese wild mice remains unclear (Terashima *et al.* 2006). In part, this is because the original genetic pattern among wild mice is everywhere confounded by an ongoing process of modern stowaway introductions (Bonhomme *et al.* 1989; Yonekawa *et al.* 1994; Marshall 1998; Tsuda *et al.* 2007).

The *Mc1r* gene, a target of an earlier study in which we surveyed the evolutionary trends of the coat

colour-related gene in *Mus* species (Shimada *et al.* 2009), is embedded in a gene-rich chromosomal region and this facilitates relatively high resolution haplotype survey using exon-based PCR primers. We examined sequence variations in the *Mc1r* gene and its neighbouring seven genes, lying in a 200-kb stretch of chromosome 8, for a selection of inbred strains and for wild mice collected at dispersed localities across Eurasia and at numerous localities in Japan. Analysis of individual gene trees provides useful information on dispersal events of *M. musculus* across the Eurasian continent and allows us to identify the origins of diverse genetic components in Japanese mice. Additionally, patterns of allelic combination of the eight genes into haplotypes support inferences regarding recombination events and the history of genetic interaction among the various immigrant lineages.

Materials and methods

Materials

We examined five inbred strains and 54 wild-captured mice from 52 localities throughout Eurasia (Table 1, Fig. 1), including 19 individuals from 19 localities in Japan. A subset of these mice was characterized genetically in a previous study (Terashima *et al.* 2006). Five strains of BFM/2Ms (BRC No. 00659), BLG2/Ms (No. 00653), MSM/Ms (No. 00209), PGN2/Ms (No. 00667) and PWK (No. 00213) were provided by RIKEN BRC with the support of National Bioresource Project of Ministry of Education, Culture, Sports, Science and Technology, Japan. Wild mice from Germany (MB9/HS3959) and Czech (HZ788/HS3963) were kindly provided by Dr. Pavel Munclinger (Czech Republic). Total DNA was extracted from tissues by the conventional phenol-chloroform method.

Experimental design and assumptions

As already explained, the pattern of mtDNA and allozyme diversity in *M. musculus* suggests a long history of divergent evolution among what are generally recognized as subspecies, combined with a secondary signal of peripheral range expansion and long-distance dispersals associated with human activities. Relative coalescence times for nDNA is approximately four times longer than for mtDNA (Birky *et al.* 1983) and it is possible, perhaps even likely, that for many nuclear genes the partitioning of genetic diversity among the subspecies would not have attained a condition of reciprocal monophyly. Nevertheless, some level of geographic partitioning of nuclear gene sequence diversity might be expected among the various subspecies. Furthermore,

this partitioning is more likely to be expressed in populations located within the inferred natural range of each subspecies and is more likely to be disrupted in areas where mice of different origin have come into secondary contact, thereby allowing hybridization and gene introgression. Based on these assumptions, we feel justified in making a tentative subspecies assignment of nuclear gene alleles where networks show clusters that correspond, at least in part, to the geographic areas occupied by the subspecies. Given the relatively small number of samples sequenced from any one region and the fact that many networks include nodes that are as yet undiscovered, it also seems reasonable to infer subspecies assignment for haplotypes that are embedded within one of these subspecies clusters but are known only from areas outside of the natural subspecies range. In contrast, alleles that either occur widely across the range of multiple subspecies or fall outside of any subspecies cluster, might be regarded as retained ancestral alleles or unassigned alleles, respectively. Sampling of more individuals and localities might well lead to changes in interpretation of unassigned alleles as missing parts of the network are populated.

Secondary contact among the genetically differentiated subspecies of mice will lead both to introgression of individual alleles and to recombination events involving alleles of divergent origin. While much emphasis has been placed on allelic introgression across hybrid zones in *M. musculus* (Tucker *et al.* 1992; Terashima *et al.* 2006; Bonhomme *et al.* 2007; Macholán *et al.* 2007), detection of recombination events between the subspecies of mice in secondary contact would not only provide robust evidence of ancient hybridization events but also permit rough estimation of the timing of such events. The probability P that a given haplotype did not change from its ancestor G generations ago is $P = (1-r)^G$, where r = recombination and mutation rate (the equation can be transformed to $G = -\ln(P)/r$) (Stephens *et al.* 1998; Koopman *et al.* 2007). We took into account that the overall rate of recombination in *M. musculus* is 0.52 cM/Mb (Jensen-Seaman *et al.* 2004), though recombination hotspots are present at certain intervals, such as 10–100 kb (Daly *et al.* 2001). Using these estimates and ignoring the mutation rate, the half-life of a haplotype ($P = 0.5$) of 20–200 kb (≈ 0.01 – 0.1 cM) can be estimated to be around 700–7000 generations. For the house mouse, with a lifespan of little more than 1 year and a likely history of introductions and local hybridization in Japan spanning 5–10 000 years, a 200-kb portion of chromosome 8 thus should be adequate to detect multiple recombinant events. Older recombination events that occurred during earlier phases of range expansion and dispersal of house mice might also be anticipated.

Table 1 List of samples used in this study and their allelic types and chromosomal constructs (haplotypes)

		Gene*											
		Type [†]	<i>Fanca</i>	<i>Spire2</i>	<i>Tcf25</i>	<i>Mc1r</i>	<i>Tubb3</i>	<i>Def8</i>	<i>Afg3l1</i>	<i>Dbn1dd1</i>			
Locality		DNA code [‡]	(448 bp)	(436 bp)	(613 bp)	(409 bp)	(400 bp)	(370 bp)	(510 bp)	(550 bp)	Haplotype		
1	Canada	Pegion	d	<u>PGN2/Ms</u>	<i>d2</i>	<i>d2</i>	<i>d1</i>	<i>u1</i>	<i>d1</i>	<i>d1</i>	<i>d1</i>	hD4	
2	Germany	Weidesgrun	d	MB8/HS3958	<i>d3</i>	<i>d3</i>	<i>d1</i>	<i>u1</i>	<i>d1</i>	<i>d1</i>	<i>d1</i>	hD6	
					<i>d8</i>	<i>d1</i>	<i>d1</i>	<i>u1</i>	<i>d1</i>	<i>d1</i>	<i>d1</i>	hD12	
3		Kubelhof	d	MB9	<i>d6</i>	<i>d1</i>	<i>d1</i>	<i>u1</i>	<i>d1</i>	<i>d2</i>	<i>d1</i>	hD10	
					<i>d6</i>	<i>d1</i>	<i>d1</i>	<i>u1</i>	<i>d1</i>	<i>d2</i>	<i>d6</i>	hD11	
4	France	Montpellier	d	<u>BFM/2Ms</u>	<i>d3</i>	<i>d2</i>	<i>d1</i>	<i>u1</i>	<i>d1</i>	<i>d1</i>	<i>d3</i>	hD5	
5	Czech Republic	Lhotka	m	<u>PWK</u>	<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>m1</i>	<i>m1</i>	<i>m18</i>	<i>m4</i>	hM21
6	Bulgaria	Toshevo	m	<u>BLG2/Ms</u>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m6</i>	<i>m1</i>	<i>m3</i>	<i>m4</i>	<i>m3</i>	hM13
7	Italy	Aosta	d	HS589	<i>d1</i>	<i>d1</i>	<i>d2</i>	<i>u1</i>	<i>d1</i>	<i>d1</i>	<i>d1</i> ^s	<i>d1</i>	hD1
					<i>d1</i>	<i>d1</i>	<i>d2</i>	<i>u1</i>	<i>d1</i>	<i>d1</i>	<i>d2</i> ^s	<i>c4</i>	hRe1
8	Czech Republic	Liborezy	m	HZ788	<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>u1</i>	<i>m1</i>	<i>m1</i>	<i>m14</i>	<i>m6</i>	hM26
					<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>u1</i>	<i>m1</i>	<i>m1</i>	<i>m14</i>	<i>m6</i>	hM26
9	Estonia	Tallinn	ssp	MG3044	<i>d6</i> ^s	<i>m1</i>	<i>m5</i> ^s	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m2</i>	<i>m2</i> ^s	hRe6
					<i>m1</i> ^s	<i>m1</i>	<i>m4</i> ^s	<i>u1</i>	<i>m1</i>	<i>m1</i>	<i>m14</i>	<i>m4</i> ^s	hM25
10	Ukraine	Donetsk	ssp	MG3065	<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>m1</i>	<i>m1</i>	<i>m7</i> ^s	<i>m7</i>	hM19
					<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>m1</i>	<i>m1</i>	<i>m23</i> ^s	<i>m4</i>	hM22
11	Uzbekistan	Toshkent	m	HS1338	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m17</i>	<i>m9</i>	hM7
					<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m17</i>	<i>m9</i>	hM7
12	Kazakhstan	Aktubinsk	m	HS1464	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m5</i>	<i>m4</i>	hM4
					<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m19</i>	<i>m4</i>	hM18
13		Balkhashi	m	HS3602	<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>m1</i>	<i>m1</i>	<i>m16</i>	<i>m8</i>	hM20
					<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>m1</i>	<i>m1</i>	<i>m16</i>	<i>m8</i>	hM20
14	Iran	Ahvaz	ssp	NIG934	<i>d4</i>	<i>d3</i>	<i>m4</i>	<i>u1</i>	<i>d1</i>	<i>d1</i>	<i>d1</i>		hRe2
					<i>d5</i>	<i>d5</i>	<i>m4</i>	<i>u1</i>	<i>m1</i>	<i>d1</i>	<i>d4</i>		hRe3
15		Nowshahr	ssp	NIG935	<i>m1</i>	<i>m1</i> ^s	<i>m1</i>	<i>m3</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	<i>m1</i>	hM11
					<i>m1</i>	<i>m2</i> ^s	<i>m6</i>	<i>m3</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	<i>m1</i>	hM3
16	Russia	Moscow	ssp	MG3055	<i>d3</i>	<i>d3</i>	<i>d3</i>	<i>u1</i> ^s	<i>d1</i>	<i>d1</i> ^s	<i>d1</i> ^s	<i>d1</i> ^s	hD7
					<i>d3</i>	<i>d3</i>	<i>d3</i>	<i>u3</i> ^s	<i>d1</i>	<i>d2</i> ^s	<i>d3</i> ^s	<i>d2</i> ^s	hD8
17		Astrakhan	ssp	HS3612	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m13</i>	<i>m2</i>	hM6
					<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m13</i>	<i>m2</i>	hM6
18		Gorno-Altaysk	ssp	HS3605	<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>m1</i>	<i>m2</i>	<i>m8</i>	<i>m5</i>	hM23
					<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>m1</i>	<i>m2</i>	<i>m8</i>	<i>m5</i>	hM23
19		Irkutsk	ssp	HS3608	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m2</i>	<i>m1</i>	<i>m1</i>	<i>m15</i>	<i>m5</i>	hM10
					<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>m1</i>	<i>m2</i>	<i>m8</i>	<i>m5</i>	hM23
20		Tomsk	ssp	HS3604	<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>m1</i>	<i>m2</i>	<i>m8</i>	<i>m5</i>	hM23
					<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>m1</i>	<i>m2</i>	<i>m8</i>	<i>m5</i>	hM23
21		Magadan	ssp	HS1845	<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>d2</i>	<i>d1</i> ^s	<i>d5</i> ^s	<i>d1</i> ^s	hRe11
					<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m2</i>	<i>d2</i>	<i>m1</i> ^s	<i>m6</i> ^s	<i>m4</i> ^s	hRe12
22		Khabarovsk	ssp	HS1461	<i>d6</i>	<i>m1</i>	<i>m1</i>	<i>m2</i>	<i>m1</i>	<i>m2</i> ^s	<i>m11</i>	<i>m5</i>	hRe5
					<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m5</i>	<i>m1</i>	<i>m1</i> ^s	<i>m8</i>	<i>m1</i>	hM12
23		Amurskii	ssp	HS1466	<i>d7</i>	<i>m1</i>	<i>m2</i> ^s	<i>m2</i>	<i>m1</i>	<i>m1</i>	<i>m21</i>	<i>m5</i> ^s	hRe8
					<i>m1</i>	<i>m1</i>	<i>m1</i> ^s	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m8</i>	<i>m4</i> ^s	hM5
24		Kraskino	ssp	HS1412	<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m4</i>	<i>m1</i>	<i>m1</i>	<i>m12</i>	<i>m4</i>	hM24
					<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m4</i>	<i>m1</i>	<i>m1</i>	<i>m12</i>	<i>m4</i>	hM24
25		Khasan	ssp	HS1335	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>u2</i> ^s	<i>m1</i>	<i>m2</i>	<i>m21</i>	<i>m4</i>	hM14
					<i>m1</i>	<i>m1</i>	<i>m3</i>	<i>u4</i> ^s	<i>m1</i>	<i>m1</i>	<i>m22</i>	<i>m4</i>	hM16
26		Okha	ssp	HS3606	<i>d7</i>	<i>d4</i>	<i>d1</i>	<i>u1</i>	<i>m1</i>	<i>d1</i>	<i>d7</i>	<i>c3</i>	hRe7
					<i>d7</i>	<i>d4</i>	<i>d1</i>	<i>u1</i>	<i>m1</i>	<i>d1</i>	<i>d7</i>	<i>c3</i>	hRe7
27		Okha	ssp	HS3845	<i>d7</i>	<i>d4</i>	<i>d1</i>	<i>u1</i>	<i>m1</i>	<i>d1</i>	<i>d7</i>	<i>c3</i>	hRe7
					<i>d7</i>	<i>d4</i>	<i>d1</i>	<i>u1</i>	<i>m1</i>	<i>d1</i>	<i>d7</i>	<i>c3</i>	hRe7
28		Tomari	ssp	HS945	<i>d6</i>	<i>d1</i>	<i>d2</i>	<i>m6</i> ^s	<i>d1</i>	<i>m1</i> ^s	<i>m1</i>	<i>m1</i>	hRe4
					<i>m1</i>	<i>d6</i>	<i>m1</i>	<i>u1</i> ^s	<i>d1</i>	<i>c1</i> ^s	<i>m20</i>	<i>m4</i>	hRe9

Table 1 (Continued)

Locality	Type [†]	DNA code [‡]	Gene*							Haplotype	
			<i>Fanca</i> (448 bp)	<i>Spire2</i> (436 bp)	<i>Tcf25</i> (613 bp)	<i>Mc1r</i> (409 bp)	<i>Tubb3</i> (400 bp)	<i>Def8</i> (370 bp)	<i>Afg3l1</i> (510 bp)		<i>Dbn1d1</i> (550 bp)
29 Japan: Hokkaido	Kushiro	ssp HS1946	<i>d6</i>	<i>d6</i>	<i>d1</i>	<i>u1</i>	<i>d1</i>	<i>d1</i>	<i>d2</i>	<i>d1</i>	hD9
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
30	Asahikawa	ssp HS1947	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>m1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>d1</i>	<i>d2</i>	<i>c2</i>	hRe15
31	Obihiro	ssp HS2781	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
32	Otaru	ssp HS2340	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>d1</i>	<i>d2</i>	<i>c2</i>	hRe10
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>d1</i>	<i>d2</i>	<i>c2</i>	hRe10
33	Naganuma	ssp HS2446	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
34	Date	ssp HS2451	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
35	Kyowa	ssp HS2779	<i>d1</i>	<i>d3</i>	<i>d1</i>	<i>u1</i>	<i>d1</i>	<i>d1</i>	<i>d1</i>	<i>d2</i>	hD2
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
36	Rankoshi	ssp HS3834	<i>m1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>d1</i>	<i>d1</i>	<i>c2</i>	hRe13
			<i>m1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>d1</i>	<i>d1</i>	<i>c2</i>	hRe13
37	Onuma	ssp HS394	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>d1</i>	<i>d2</i>	<i>c2</i> ^s	hRe10
			<i>m1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>d1</i>	<i>d2</i>	<i>m10</i> ^s	hRe14
38 Japan: Honshu	Sakekawa	ssp HS2461	<i>m1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>d1</i>	<i>d2</i>	<i>c2</i>	hRe15
			<i>m1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>d1</i>	<i>d2</i>	<i>c2</i>	hRe15
39	Tendo	ssp HS2458	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>m1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>d1</i>	<i>d2</i>	<i>c2</i>	hRe15
40	Otsuti	ssp HS2454	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
41	Sendai	ssp HS2456	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
42	Sakata	ssp HS2457	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>m1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>d1</i>	<i>d2</i>	<i>c2</i>	hRe15
43	Chiba	ssp HS2468	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
44	Atsugi	ssp HS3839	<i>d1</i>	<i>d3</i>	<i>d1</i>	<i>u1</i>	<i>d2</i>	<i>d1</i>	<i>d1</i>	<i>d1</i>	hD3
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
45	Mishima	ssp <u>MSM/Ms</u>	<i>m1</i>	<i>m1</i>	<i>m4</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM17
46 Japan: Kyushu	Miyazaki	ssp HS2472	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>c2</i>	hRe22
47	Kagoshima	m HS3603	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m10</i>	hM2
			<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	hC1
48 Korea	Gyeongju	m HS1368	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m9</i>	hM1
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m9</i>	hM1
49	Busan	m HS3540	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m9</i>	hM1
			<i>m1</i>	<i>m1</i>	<i>m2</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m9</i>	hM15
50 China	Beijing	ssp NIG5065	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m2</i>	<i>m3</i>	<i>m9</i>	hM9
			<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m2</i>	<i>m3</i>	<i>m9</i>	hM9
51	Kunming	c HS506	<i>m1</i> ^s	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>c2</i>	<i>c3</i>	<i>c1</i>	hRe16
			<i>c1</i> ^s	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	hC1
52	Kunming	c HS507	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m1</i>	<i>m24</i>	<i>m9</i>	hM8
			<i>m1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>c4</i>	<i>c1</i>	<i>c1</i>	hRe18
53 Taiwan	Taitung	c HS2400	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>m1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	hRe17
			<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	hC1
54 Nepal	Kathmandu	ssp HS1467	<i>u1</i>	<i>u1</i>	<i>u1</i>	<i>m1</i> ^s	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	hRe21
			<i>u1</i>	<i>u1</i>	<i>u1</i>	<i>u1</i> ^s	<i>c1</i>	<i>c1</i>	<i>c2</i>	<i>c1</i>	hC6

Table 1 (Continued)

Locality	Gene*										Haplotype
	Type [†]	DNA code [‡]	<i>Fanca</i>	<i>Spire2</i>	<i>Tcf25</i>	<i>Mc1r</i>	<i>Tubb3</i>	<i>Def8</i>	<i>Afg3l1</i>	<i>Dbndd1</i>	
			(448 bp)	(436 bp)	(613 bp)	(409 bp)	(400 bp)	(370 bp)	(510 bp)	(550 bp)	
55 Myanmar Lashio	c	HS3357	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	hC1
56 Bangladesh Comilla	c	HS3701	<i>u1</i> [§]	<i>c1</i>	<i>c1</i>	<i>u5</i> [§]	<i>c1</i>	<i>c5</i> [§]	<i>m9</i>	<i>u1</i>	hRe19
			<i>u2</i> [§]	<i>u1</i>	<i>u2</i>	<i>u1</i> [§]	<i>c1</i>	<i>c2</i> [§]	<i>c3</i>	<i>u1</i>	hC8
57 India Mizoram	c	HS3441	<i>c1</i> [§]	<i>c1</i>	<i>u1</i> [§]	<i>u1</i> [§]	<i>c1</i> [§]	<i>c1</i>	<i>c1</i>	<i>c1</i>	hC4
			<i>u1</i> [§]	<i>c1</i>	<i>u2</i> [§]	<i>m1</i> [§]	<i>m1</i> [§]	<i>c1</i>	<i>c1</i>	<i>c1</i>	hRe20
58 Indonesia Bogor	c	HS3736	<i>c1</i>	<i>c1</i>	<i>c2</i>	<i>u1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	hC3
			<i>c1</i>	<i>c1</i>	<i>c2</i>	<i>u1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>c1</i>	hC3
59 Philippines Los Banos	c	HS3882	<i>c1</i>	<i>c1</i>	<i>c1</i>	<i>u1</i>	<i>c1</i>	<i>c3</i>	<i>c3</i>	<i>c1</i>	hC2
			<i>u1</i>	<i>u2</i>	<i>u2</i>	<i>u1</i>	<i>c1</i>	<i>c3</i>	<i>c3</i>	<i>c1</i>	hC7

*Abbreviation for each of the subspecies groups: *c*, *castaneus*; *d*, *domesticus*; *m*, *musculus*; *u*; unknown.
[†]Subspecies type is that initially labelled when sample was obtained: *c*, *castaneus*; *d*, *domesticus*; *m*, *musculus*; *ssp*, unknown.
[‡]Individuals derived from laboratory strains are underlined.
[§]Alleles with uncertain phase in haplotype estimation.

Sequence analyses

We chose a gene-rich region including a color related gene, *Mc1r*, for which we previously have examined interspecific evolutionary patterns (Shimada *et al.* 2009). We amplified segments of *Fanca* (448 bp), *Spire2*

(436 bp), *Tcf25* (613 bp), *Mc1r* (409 bp), *Tubb3* (400 bp), *Def8* (370 bp), *Afg3l1* (510 bp) and *Dbndd1* (550 bp), which are distributed at 20–30 kb interval along a 200-kb portion on mouse chromosome 8 (Fig. 2a). The primers used in the PCR analyses (Table 2) were designed using Ensemble Mouse Genome Database

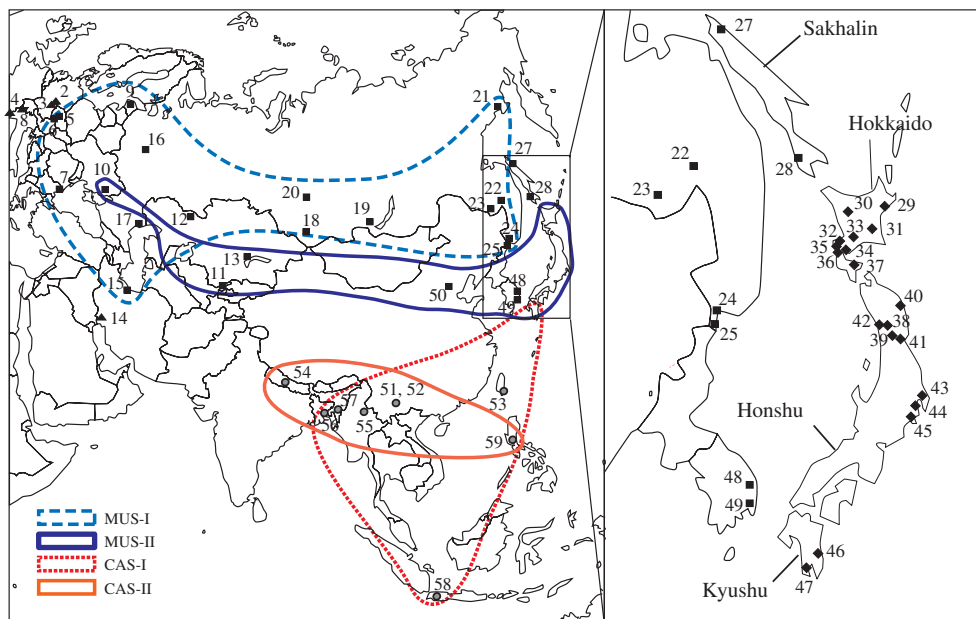


Fig. 1 Collection localities of *Mus musculus* used in this study. Locality numbers are as in Table 1. Based on previous studies, three major subspecies groups are represented: *M. m. musculus* (filled square), *M. m. domesticus* (filled triangle), and *M. m. castaneus* (gray circle). Because of taxonomic uncertainty regarding the status of Japanese wild mice, they are treated as unknown subspecies here (filled diamond). Geographic areas representing clades of MUS-I, MUS-II, CAS-I, and CAS-II inferred from eight nuclear gene sequences (see Fig. 3) are indicated.

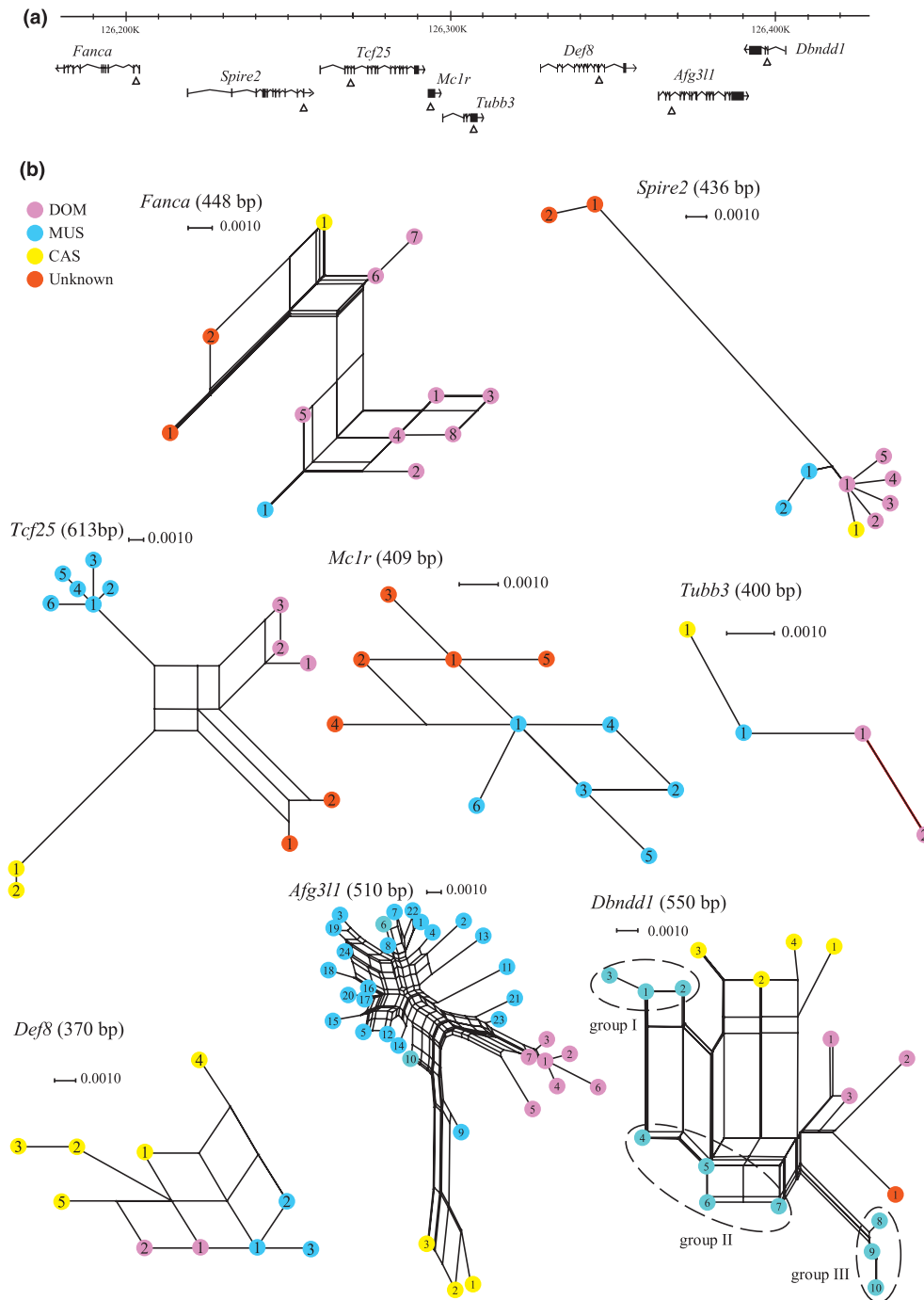


Fig. 2 Positions of analysed regions (open triangle) in eight genes of murine chromosome 8 (a) and network trees with resultant allelic sequences of the eight genes of *Fanca*, *Spire2*, *Tcf25*, *Mc1r*, *Tubb3*, *Def8*, *Afg3l1* and *Dbndd1* (b). Network trees were constructed in each data set with Neighbor-Net method. Assessment for the subspecies groups of *M. m. domesticus* (pink circle), *M. m. castaneus* (yellow circle) and *M. m. musculus* (blue circle) were tentatively done based on the geographic origins of alleles. Alleles unassigned to subspecies group are shown with orange circles for the network trees of *Fanca*, *Spire2*, *Tcf25*, *Mc1r* and *Dbndd1*.

(<http://www.ensembl.org/>). The amplifications were carried out for 35 cycles each consisting of 30 s at 96 °C for denaturation, 30 s at 57 °C for annealing and 30 s at 72 °C for extension. These reaction mixtures (20 µL)

contained 2.5 mM MgCl₂. AmpliTaq Gold DNA Polymerase (Applied Biosystems, ABI) was used in the PCR. Cycling of the first PCR was preceded by 10 min at 95 °C for activation of the polymerase. The

Table 2 List of primers used in this study

Name	Posi 3' end*	Exon	Intron	Total
Initial survey				
<i>Fanca</i> (Fanconi anaemia, complementation group A)				
Fanca_F: GCAGACCGGTGTTCCAGACGCT	126,204,635	145 (1–48, 352–448)	303 (49–351)	448
Fanca_R: CTCAGCCAGGACAACCTCCTCT	126,204,165			
<i>Spire2</i> (spire homolog 2 (Drosophila))				
Spire2_F: ACGCCAACACTGAGGAGAGATGCCT	126,254,879	61	375	436
Spire2_R: CTGGTGCAGTCACTGCAGACATCCT	126,255,324			
<i>Tcf25</i> (transcription factor 25 (basic helix-loop-helix))				
Tcf25_F: TCCAGACAAGCCCCTATCATGT	126,276,897	0	613	613
Tcf25_R: TCCATGCTGTACAGGGCTCTCT	126,277,671			
<i>Mc1r</i> (melanocortin 1 receptor)				
Mc1r_F: AGCAGCTGGACAACCTCATT	126,293,953	409	0	409
Mc1r_R: TGGCTGCGGAAAGCATAGATGA	126,294,480			
<i>Tubb3</i> (tubulin beta 3)				
Tubb3_F: TTACCTGCTCTTCTCTCCTCA	126,306,693	400	0	400
Tubb3_R: AAGCCGGGCATGAAGAAGT	126,307,204			
<i>Def8</i> (differentially expressed in FDCP 8)				
Def8_F: AACTAACACCTGCTGGGTCTTGG	126,345,603	211 (1–100, 260–370)	159 (100–259)	370
Def8_R: TAGCCGAGAGTGAGCTACAGG	126,346,027			
<i>Afg3l1</i> (ATPase family gene 3 like-1 (yeast))				
Afg3l1_F: ACTGACTTCATTTTCCCGGAGG	126,366,504	0	510	510
Afg3l1_R: TTGGAGGTACTGAATTCCTGG	126,367,208			
<i>Dnnd1</i> (Dysbindin (dystrobrevin binding protein 1) domain containing 1)				
Dnnd1_F1: AATACCAGCACCAGGGTTCCTG	126,395,951	0	550	550
Dnnd1_F2: TGGCATCCCAATACCAGCACCAG [†]	126,395,959			
Dnnd1_R: AGAGAGGAGACGCTGCTCAGAG	126,395,260			

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

[†]Specific to MUS.

double-stranded PCR product was purified by a 20% polyethylene glycol-2.5 M NaCl precipitation method and sequenced utilizing the PRISM Ready Reaction Dye Deoxy Terminator Cycle Sequencing Kit (ABI) and run on ABI 3100 Stretch automated sequencer. Both strands were sequenced using either the forward or reverse primer used for PCR.

The sequence fragments obtained with different primers were assembled by DNASIS (Hitachi, Japan) and sequence alignments were done by eye. Sequences with more than one polymorphic site were separated into alleles mainly by the parsimony method (Karn *et al.* 2002). This method, however, failed to obtain allele structures in *Dnnd1* and *Afg3l1*, in which sequences were moderately and highly heterogeneous, respectively. For the data of *Dnnd1*, we performed a mismatch sequence analysis, in which sequence primers (14-mer in length; data not shown) with one mismatch at the 3' end were used. For *Afg3l1*, we inferred allele sequences of heterozygous individuals using the software PHASE 2.1 (Stephens *et al.* 2001; Stephens & Donnelly 2003). All new sequences were deposited in the DDBJ/EMBL/GenBank DNA database with accession numbers AB504780 to AB504892.

Phylogeographic data analysis

The number of alleles, expected heterozygosity (*He*) and mean number of pairwise difference (π) were calculated using ARLEQUIN 3.1 (Excoffier *et al.* 2005). Minimum number of recombination events (*Rm*) was inferred with DNASP 5.00 (Librado & Rozas 2009). Network trees were constructed for each locus by the Neighbor-Net method, as implemented in SPLITS TREE (Huson & Bryant 2006). Kimura's two-parameter model, in which ts/tv ratio is set at 2, was employed for construction of each network tree.

As explained above, for each gene we assigned alleles to subspecies groups (*c*: CAS, *d*: DOM, *m*: MUS) on two criteria: (i) clustering patterns in the network trees; and (ii) geographic origins of the individual mice, with particular emphasis on mice from within the inferred natural range of each subspecies (Boursot *et al.* 1993; Moriwaki *et al.* 1994; Guénet & Bonhomme 2003). Under the second criterion, alleles exclusive to mice from peripheral parts of Eurasia, Primorye, Sakhalin and Japan were only assigned to a subspecies group if the haplotype is closely associated with an otherwise

assigned haplotype. To obtain a second perspective on subspecies assignments, we conducted a Bayesian clustering analysis with STRUCTURE 2.3 (Pritchard *et al.* 2000) for each of the eight genes. Markov Chain Monte Carlo simulations were performed with 10^6 generations after at 10^6 generation burn-in period. The number of 'cluster', K , was determined based on estimated probability.

Haplotype assessment from allelic combinations of the eight genes

We used PHASE 2.1 (Stephens *et al.* 2001; Stephens & Donnelly 2003) with default setting to determine unique allele arrangements (henceforth called haplotypes) with respect to the eight gene regions of *Fanca*, *Spire2*, *Tcf25*, *Mc1r*, *Tubb3*, *Def8*, *Afg3l1*, and *Dbn1d1* (Fig. 2b). A network tree was constructed from the concatenated eight-gene dataset by the Neighbor-Net method, as described above.

Results

Allele variation in the eight genes

The eight genes showed marked variation in the number of variable sites, ranging from 3 (*Tubb3*) to 41 (*Tcf25*) (Table 3). Higher than expected levels of heterozygosity was observed in *Afg3l1* (0.91) and *Dbn1d1* (0.90). Indels of a few bases were seen in *Fanca*, *Spire2*, *Tcf25* and *Afg3l1* and these were taken into consideration in the allelic designation. The total number of alleles, after resolving heterozygous sequences into individual alleles, was 12, 11, 13, 11, 4, 10, 34 and 18 alleles in *Fanca*, *Spire2*, *Tcf25*, *Mc1r*, *Tubb3*, *Def8*, *Afg3l1* and *Dbn1d1*, respectively. The minimum numbers of recombination events (R_m) estimated

Table 3 Standard genetic information of the eight loci examined. The number of alleles, expected heterozygosity (H_e) and mean number of pairwise difference (π) were calculated using ARLEQUIN 3.1 (Excoffier *et al.* 2000). Minimum number of recombination events (R_m) was inferred with DnaSP 5.00 (Librado & Rozas 2009)

Gene	No. of alleles	No. of variable sites	No. of indels	H_e	π	R_m
<i>Fanca</i>	12	14	4	0.56	4.35	2
<i>Spire2</i>	11	16	2	0.63	1.96	0
<i>Tcf25</i>	13	41	1	0.78	9.42	4
<i>Mc1r</i>	11	9	0	0.69	1.17	1
<i>Tubb3</i>	4	3	0	0.56	0.67	0
<i>Def8</i>	10	11	0	0.70	1.43	3
<i>Afg3l1</i>	34	30	9	0.91	8.01	7
<i>Dbn1d1</i>	18	24	0	0.90	7.06	4

by the four-gamete test, as implemented in DnaSP 5.0, ranged from 0 (*Spire2* and *Tubb3*) to 7 (*Afg3l1*).

The majority of network trees exhibited clustering patterns of apparent phylogeographic significance (Fig. 2b; Fig. S1, Supporting Information) and this allowed tentative subspecies assignment groups for the majority of alleles. This was further tested using a Bayesian approach as implemented in STRUCTURE (Pritchard *et al.* 2000). The STRUCTURE plots confirmed the presence of strong phylogeographic signal within our data sets (Fig. S2, Supporting Information). Alleles were enumerated with letters representing the subspecies groups, CAS (c), DOM (d) and MUS (m), followed by a number (e.g. $c1$, $c2$ etc).

A number of the network trees displayed four clusters rather than three. In each of the *Fanca*, *Spire2*, *Tcf25* and *Dbn1d1* networks, alleles found in the natural CAS territory are divided between two distinct clusters. These clusters were tentatively labelled as CAS and 'unassigned' (u), with CAS assigned to the cluster with higher allele frequencies (Table 1, Fig. 2B). In the *Mc1r* network, individuals within natural DOM and CAS territories shared the same allele. Although this was labelled as an unknown allele, $u1$, it might conceivably be a shared ancestral allele. Several alleles recorded only from Russia also showed a dispersed pattern and were labelled as 'unassigned'. In the *Dbn1d1* network, an allele designated as CAS was recovered from Italy ($c4$), within the DOM territory. In the *Dbn1d1* network, MUS alleles comprise three subgroups [Fig. 2b; I ($m1$ – $m3$), II ($m4$ – $m7$) and III ($m8$ – $m10$)]; these differed by 0.8–2.4% substitution per site on average and show apparent geographic affinity, with I and II from East Europe and Russia, and III from Uzbekistan, northern China, Korea and Japan. Similarly, in the *Dbn1d1* gene, there is a clear distinction between northern (western) ($m1$ – 7) and southern (eastern) alleles ($m8$ – 10). The separateness of these groups is also clearly seen in the network tree of concatenated haplotypes (Fig. 3).

In Japan, CAS alleles were recovered from Hokkaido, eastern Honshu and Kyushu and DOM alleles from Hokkaido and northern Honshu (Table 1). CAS alleles were observed in heterozygous state with MUS alleles in an individual from Kagoshima, Kyushu. Similarly, DOM alleles were recovered in all individuals from three localities (Kushiro and Kyowa on Hokkaido and Atsugi on Honshu). Even more chimeric haplotypes, with combinations of MUS, CAS and DOM alleles, were observed in some individuals from Hokkaido and also in a mouse from Miyazaki, Kyushu. Similar allelic mixing was also seen in several mice from Sakhalin (Okha and Tomari) and Primorye (e.g. Khabarovsk) but notably, only very rarely in other parts of the species range.

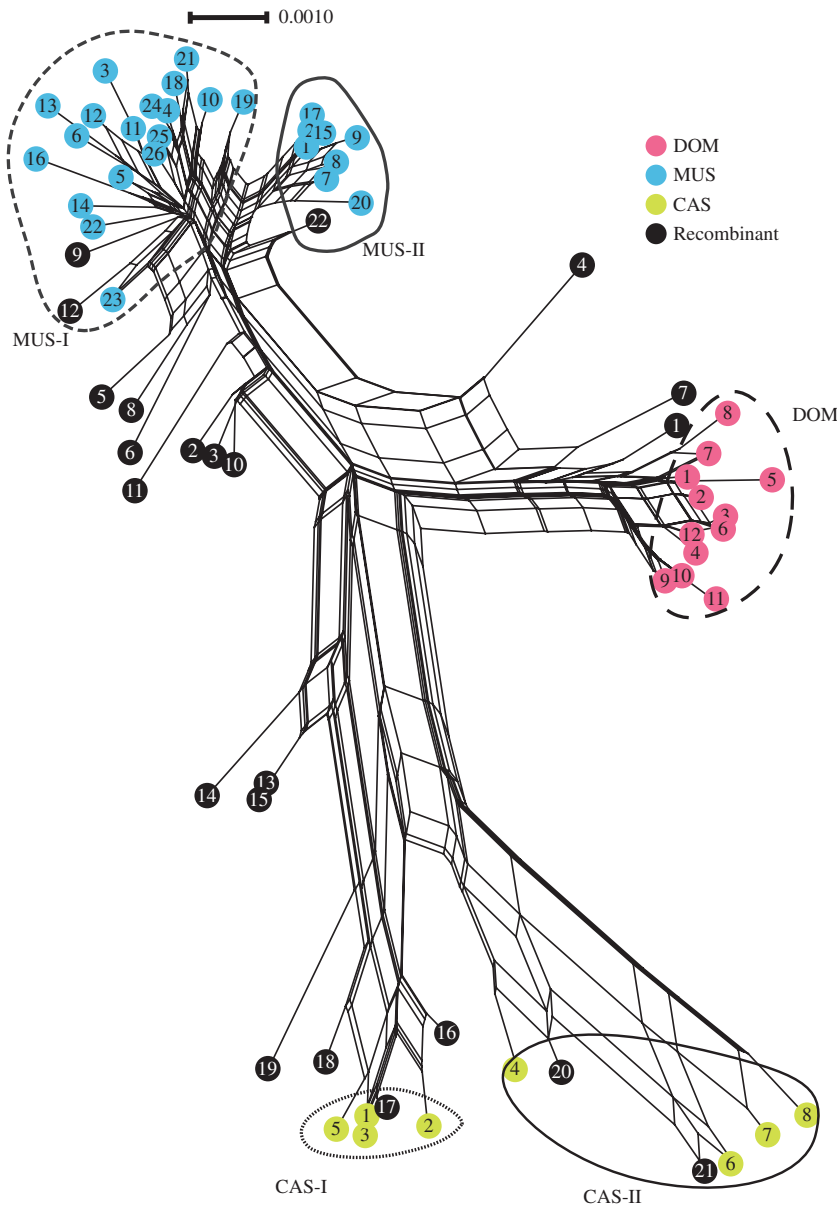


Fig. 3 Network tree derived from the concatenated dataset of eight nuclear genes. The three subspecies, *M. m. domesticus* (DOM), *M. m. castaneus* (CAS), *M. m. musculus* (MUS) and their putative recombinant haplotypes are shown with pink, yellow, blue and black circle respectively. Further subgroups seen in each of CAS (CAS-I, CAS-II) and MUS (MUS-I, MUS-II) are indicated with coloured lines. Detailed information of allelic arrangements in haplotypes is described in Table 1.

Haplotype assessment

Allelic combinations with respect to the eight gene regions of *Fanca*, *Sprie2*, *Tcf25*, *Mc1r*, *Tubb3*, *Def8*, *Afg3l1* and *Dnnd1* were assessed with a Bayesian method using PHASE (Table 1). We tentatively classified all identified haplotypes as either non-recombinant or recombinant, based on the subspecies assignment of individual allele components of each haplotype. We identified the following subspecies-specific non-recombinant haplotypes: CAS (hC1-hC8), DOM (hD1-hD12) and MUS (hM1-hM26). In addition, we denoted 22 different putative recombinant haplotypes (hRe1-hRe22), derived mainly from the peripheral areas of Japan,

Sakhalin, and Primorye, but also from more central areas such as Nepal and southern China (Kunming). Putative recombinant haplotypes were also identified from several localities within the CAS territory, namely Taiwan (hRe17), Bangladesh (hRe19) and India (hRe20).

A Neighbor-Net tree based on concatenated sequences for all eight genes (Fig. 3) separates the putative non-recombinant CAS, DOM, and MUS haplotypes into clear clusters. Additionally, the tree suggests the presence of two divergent lineages within the CAS group (Fig. 3; CAS-I and CAS-II), reflecting clustering within the gene trees of *Fanca*, *Sprie2* and *Tcf25*, all lying in the upper part of the 200-kb region examined. These sub-groups display overlapping geographic

distributions, though the number of sampled localities is small (Fig. 1). The MUS group also shows a two-part clustering (Fig. 3) but in this case the groups correspond to geographic regions (Fig. 1), namely 'northern' (MUS-I) and 'southern' (MUS-II) subgroups, representing mice from eastern Europe, Siberia and Primorye and those from Uzbekistan, northern China, Korea and Japan, respectively. In contrast, the DOM group forms a tight-knit cluster on the network tree, with much lesser internal divergence than within either CAS or MUS. Many of the putative recombinant haplotypes occupy intermediate positions on the network tree (e.g. hRe2–6, hRe10 and hRe13–15). However, others fall within or close to one or other of the subspecies clusters (e.g. hRe17 within CAS-I; hRe20–21 within CAS-II). This contrasting location clearly reflects the dominant composition of particular haplotypes (e.g. hRe17 has seven CAS alleles and one MUS) but it also might shed light on the affinities of 'unassigned' alleles which could belong to poorly sampled parts of the genetic diversity for a subspecies or might themselves be a product of an intra-genic recombination event, thereby producing a truly chimeric allele.

As noted above, putative recombinant haplotypes were detected mainly in areas marginal to the main subspecies territories, most notably in the far eastern part of Eurasia; i.e. Magadan, Sakhalin and the northern part of Japan (Table 1). Two unique haplotypes from Magadan (hRe11 and hRe12) were both putative recombinants, in which putative MUS and DOM alleles are mingled. In Sakhalin, haplotypes combined DOM, MUS, CAS and 'unassigned' alleles but the mixing patterns were substantially different between the northern (Okha) and southern (Tomari) tips of the peninsula.

In Japan MUS haplotypes are common (hM2, hM17; $n = 22$ of 38, 57.9%) which is very similar to a haplotype detected in Korean mice (hM1). The remaining Japanese haplotypes are variously identified as DOM ($n = 3$, 7.9%), CAS ($n = 1$, 2.6%), and putative recombinants ($n = 12$, 31.6%). Seemingly non-recombinant DOM haplotypes were recovered from three mice from Kushiro (Hokkaido; hD9), Kyowa (Hokkaido; hD2) and Atsugi (central Honshu; hD3). Haplotype hC1, which occurs commonly in mice from Southeast Asia, was obtained from Kagoshima, at the southern tip of Kyushu. All putative recombinants detected in Japan were unique to that context and most involved MUS, DOM and CAS alleles. The most common, hRe15 ($n = 6$, 17%) with the composition of *Fanca*^{m1}, *Spire2*^{c1}, *Tcf25*^{c1}, *Mc1r*^{u1}, *Tubb3*^{c1}, *Def8*^{d1}, *Afg3l1*^{d2} and *Dbn1dd1*^{c2}, was recovered from a wide area of northern Japan (Hokkaido and northern Honshu). On Hokkaido, in addition to hRe15 (Asahikawa), three other putative recombinants were recovered (hRe10 in Otaru and Onuma; hRe13 in

Rankoshi; and hRe14 in Onuma). On Honshu, hRe15 was recovered from three of eight sampled localities (Sakekawa, Tendo and Sakata). Notably, on Kyushu, a mouse from Miyazaki possessed a unique putative recombinant (hRe22) which differs from the common MUS haplotype hM2 only in the insertion of the CAS element *Dbn1dd1*^{c2}.

Temporal assessment

To illustrate the potential use of the extended chromosome region as a source of temporal information on hybridization, we performed a 'ballpark' calculation for the time of secondary contact in the peripheral parts of Eurasia, using the patterns of the putative recombinant haplotypes. In Japan, the putative recombinant haplotypes hRe10, hRe13–15 and hRe22 each appear to be the product of a single break point along the 200-kb segment. As we mentioned in Materials and Methods, the half-life of a 200-kb segment is estimated at 700 generations. Considering other factors such as generations with homozygous CAS haplotypes, this clearly represents a minimum generational lapse since secondary contact—the exact duration might be considerably longer. At any rate, the unique occurrence in Japan of multiple putative recombinant haplotypes clearly points to a long history of secondary contact between MUS and CAS alleles rather than an admixture during modern times. Contrariwise, the occurrence in Japan of 'intact' DOM haplotypes, as in mice from Kyowa and Atsugi, suggests a time frame for arrival of DOM mice of a few hundreds of years and perhaps considerably less. Such estimates are admittedly crude but to obtain robust evidence for stowaway introduction in recent times, it might be necessary to examine multiple chromosomal regions with longer window sizes, such as several cM (see Koopman *et al.* 2007).

Discussion

Utility of the haplotype reconstruction approach using linked nuclear genes

Each of the eight gene regions (*Fanca*, *Spire2*, *Tcf25*, *Mc1r*, *Tubb3*, *Def8*, *Afg3l1* and *Dbn1dd1*) examined on chromosome 8 displays a strong geographic signal within the phyletic clustering. This it appears that coalescence for each gene is approaching the state of exclusive clades if not full reciprocal monophyly (Funk & Omland 2003). Significantly, the geographic clusters also are broadly consistent with the subspecies classification of *M. musculus* as defined previously on mtDNA and allozyme evidence; i.e. MUS in northern Eurasia (eastern Europe, Siberia, Primorye, northern China,

Korea and Japan); CAS in south Asia (India, Southeast Asia and southern China); and DOM in western Eurasia (western Europe and western part of Iran). This is further illustrated in the Neighbor-Net tree for the concatenated haplotypes (Fig. 3), where house mice obtained from within the core territory of each subspecies possess multi-gene haplotypes with unambiguous phylogenetic relationships. Our evidence thus joins the growing support for recognition of house mouse subspecies as genetically well-differentiated evolutionary lineages with long and largely separate histories (Abe *et al.* 2004; Frazer *et al.* 2007).

Our study is the first to identify possible recombinant haplotypes in wild house mice across broad geographic areas. The diversity and complex structure of many of the putative recombinant haplotypes is suggestive of ancient, *in situ* recombination events, as postulated previously by Bonhomme *et al.* (2007) from minisatellite data and the high values for the estimated minimum recombination events (Rm) within the gene regions (Table 3) also supports this view. Significantly, we found recombinant haplotypes to be most common in mice obtained from peripheral parts of Eurasia, including the Russian Far East (hRe5, 8, 11–12), Sakhalin (hRe4, 7, 9) and the Japanese Islands (hRe10, 13–15, 22). The strong phylogeographic pattern among recombinant haplotypes further highlights the potential of this approach for examining the history of inter-subspecies hybridization in *M. musculus* and provides effective confirmation of our contention that recombination histories can be inferred from multiple linked genes.

Genetic composition of wild house mice in potential source areas

Phylogenetic patterns of alleles in each gene (Fig. 2; Fig. S1, Supporting Information) and of the concatenated haplotypes (Fig. 3) support the traditional view that *M. musculus* comprises three well-differentiated phyletic lineages. However, gene networks for each of *Fanca*, *Spire2* and *Tcf25* indicate the possible presence of a fourth lineage, distributed in Southeast Asia and India, effectively within the wider CAS territory (Fig. 2). This finding is reminiscent of allozyme data (e.g. Din *et al.* 1996) which identified a fourth genetic component within *M. musculus*—the ‘*bactrianus* group’ from northern India—and with RFLP evidence for genetic complexity within the CAS group for the nuclear ribosomal RNA gene (Suzuki *et al.* 1986). These diverse yet convergent results highlight again the likely antiquity and intrinsic variability of subspecies groups among house mice and the need for further taxonomic assessment of *M. musculus*. In particular, much greater

attention is warranted for populations on the Indian subcontinent which might harbour more than one deeply divergent evolutionary lineage.

Our data also identify additional phylogeographic structure among mice of the MUS group. *Mc1r* alleles comprise two main subgroups which are distributed in the northern (eastern Europe, Siberia and Primorye: *m2–5*) and southern (Uzbekistan, northern China, Korea and Japan: *m1*) parts of the MUS territory, respectively (Table 1). The separateness of these groups is also clearly seen in the *Dnnd1* network and the network of concatenated haplotypes (Fig. 3). Based on these data, it is tempting to draw a division within MUS between a northern lineage found at high latitudes, from Bulgaria to Sakhalin (MUS-I) and a southern lineage, mostly found at lower latitudes, from Uzbekistan, through northern China to Korea and Japan (MUS-II). Other evidence of phylogeographic structure within MUS includes differences in C-banding patterns between MUS from eastern Europe and northern China/Japan (Moriwaki *et al.* 1986).

Although our study was not primarily focused on Europe, it also provides new insights into evolution of the DOM subspecies group. In *Dnnd1*, a mouse from Italy carries a CAS allele (*c4*) that is otherwise unparalleled among mice from western Europe. This implies that DOM mice experienced a concealed evolutionary event, culminating in the introgression of genetic components from CAS to DOM. The mouse in question was collected in an area known to harbour Robertsonian mutant mice (P. Vogel, Pers. comm.) and previous study of RFLP of nuclear ribosomal RNA indicated that Robertsonian rearrangements may be associated with inter-subspecies hybridization (Moriwaki *et al.* 1984; Suzuki & Kurihara 1994). Further analysis of wild mouse populations in western Europe is needed to understand the historical context and extent of genetic introgression from neighbouring subspecies groups.

Genetic composition and evolutionary history of Japanese wild mice

Our results confirm that MUS genetic components predominate in Japanese wild mice. At a general level, the MUS components of all Japanese wild mice, from Hokkaido to Kyushu, are referable to the ‘southern group’ within MUS. More specifically, particular allelic affinity is observed to geographically proximate Korean wild mice. Most tellingly, Japanese and Korean mice share the *m1* allele in the highly polymorphic locus *Afg3l1*, while Chinese mice possess different alleles for the MUS components (*m3* in Beijing and *m24* in Kunming). This finding is concordant with previous data for the Y-specific gene *Sry*, in which mice from Japan and Korea

were found to share the same variant sequence (Nagamine *et al.* 1994; Terashima *et al.* 2006).

The CAS and DOM components found in Japanese mice present more complex pictures. For CAS, the chromosome portion of *Spire2-Tcf25-Mc1r-Tubb3* in some Japanese mice possessed the *c1* allele found across a wide area of CAS territory, including South China, Southeast Asia and northeast India. In contrast, the CAS allele *Dbndd1^{c2}* is unique to Japan and differs at 0.4% substitution/site from the more dominant *Dbndd1^{c1}*. Interestingly, other minor CAS alleles were also recovered from areas peripheral to the main range of CAS, such as Sakhalin (*c3*) and Italy (*c4*). Whether the presence of these unusual CAS alleles in places such as Europe and Japan is due to early episodes of genetic interaction followed by drift cannot be answered without more intensive sampling within the core CAS territory, much of which remains poorly sampled due to lack of available wild mouse material. However, the fact that CAS alleles are inserted adjacent to DOM alleles in various recombinant haplotypes from Japan (hRe10, 13–15) and Sakhalin (hRe7, 9) hints at an extended interaction among MUS, CAS and DOM components, perhaps in the northeastern part of the range of *M. musculus*. Interestingly enough, we detected an 'intact' CAS haplotype (hC1) in one mouse (HS3603) from Kagoshima on Kyushu. The same haplotype was also recovered from China, Taiwan and Myanmar. It is presently unclear how and when it came to Kagoshima but a recent stowaway introduction seems likely.

Although DOM genetic components are generally not well represented in Japanese wild mice, various studies including analyses of both allozymes (Minezawa *et al.* 1979; Bonhomme *et al.* 1989) and mitochondrial DNA (Yonekawa *et al.* 1988) have detected DOM contributions in mice from northern Kyushu, westernmost Honshu and the south-easternmost islands, Ogasawara. Rather unexpectedly, we detected individual DOM alleles as well as 'intact' DOM haplotypes in mice from the northern part of Japan, including localities on both Hokkaido and Honshu. The presence of 'intact' DOM haplotypes (hD2, 3, 9) might be attributed to recent stowaway introductions, as has been suggested by earlier workers (Minezawa *et al.* 1979; Yonekawa *et al.* 1988, 1994; Bonhomme *et al.* 1989; Marshall 1998) and also postulated for black rats (*Rattus rattus*) collected from areas near ports and airports (Chinen *et al.* 2005; Tsuda *et al.* 2007). In contrast, several observations support the notion of prehistoric introduction of DOM genetic components to Japan, namely: (i) presence in Japan of several unique putative recombinant haplotypes carrying the DOM elements; (ii) conjunction of rare components of CAS and DOM in the same recombinant haplotype (hRe10, 13–15); (iii) appearance of

putative recombinants across a wide area of northern Japan; and (iv) the presence of putative recombinant haplotypes in the nearby region of Sakhalin (i.e. hRe7). Importantly, this evidence does not imply prehistoric movement of DOM mice between western Europe and East Asia. Rather, a more likely scenario is that DOM alleles entered Japan either as MUS/DOM recombinants from northern Eurasia or as CAS/DOM recombinants from southeast Asia, though sampling in the latter area has not yet detected appropriate source populations.

To more fully assess the status of introduced populations we may need to analyse a longer stretch of chromosome segments, such as 1–10 Mbp. Alternatively, this issue might be more effectively examined using microsatellites (Sakai *et al.* 2005) or SNP markers in which substantial variation is reported between representative strains of DOM and MUS (Abe *et al.* 2004).

Concluding remarks

Our results further demonstrate that *Mus musculus* has experienced a complex evolutionary history involving reticulate evolution among populations of diverse geographic origin and hint at a time frame that may predate the geographic expansion of mice as human commensals (Bonhomme *et al.* 1984; Boursot *et al.* 1993). For Japanese wild mice, the discovery of well-embedded DOM components on chromosome 8 introduces a new element into an already complicated and incompletely documented evolutionary history (Terashima *et al.* 2006). One possible scenario would have DOM alleles entering Japan in prehistoric times as part of recombinant CAS/DOM haplotypes in mice carrying CAS mtDNA. However, critical evidence in the form of ancient DOM elements in wild mice still resident within the CAS core area has not yet emerged, though large parts of this genetically diverse region remain poorly sampled. Under this scenario, MUS genetic components arrived in Japan more recently, probably via the Korean Peninsula. However, recombination frequencies indicate that this MUS component also arrived during prehistoric times, possibly several thousand years ago and perhaps also bringing with it some embedded DOM alleles. As discussed at length by Terashima *et al.* (2006), this scenario implies an intensive exclusion of CAS and DOM mtDNA lineages, especially in the western part of Japan, following the introduction of MUS-type mice. For nuclear elements, this seems unlikely unless interbreeding frequency and/or success was extremely low. However, the alternative scenario posited by Terashima *et al.* (2006), involving an original MUS-type wild mouse population, with later arrival of CAS-type mice in Japan, similarly fails to account for the evidence of prolonged interaction of CAS and DOM elements, unless this has

occurred in the Asian region, prior to migration. Both scenarios thus imply a complex evolutionary history, much of it significantly predating the period of human-associated long range movements. Our general finding is thus consistent with that of Bonhomme *et al.* (2007), based on analysis of minisatellite DNA.

Our scenario for the origin of Japanese mice shows several parallels with recent anthropological notions concerning the origin of the Japanese people. Firstly, our conclusion that the earliest immigrants mice were genetically heterogeneous, probably including components of both CAS and DOM lineages, shows parallels with evidence of high genetic diversity in the Paleolithic Japanese population, as denoted in the mitochondrial DNA data from ancient Jomon remains (e.g. Horai *et al.* 1996; Shinoda & Kanai 1999). Secondly, we identify the latest phase of dispersal of mice to Japan as occurring through the central route, with northern China as the primary source area and Korea as the final point of departure. This is in accord with the general notion that modern Japanese are most closely related to the Koreans (Horai *et al.* 1996; Omoto & Saitou 1997; Tanaka *et al.* 2004). These apparent similarities in the history of people and mice might be coincidental. However, they may also point to a far closer and longer association between the two species than currently credited, both species making their way eastward across the Eurasian continent during prehistoric times. Careful attention to the archaeological record is needed to explore the antiquity of the bond that unites mice and people.

Acknowledgements

We wish to express our appreciation to Y. Satta and T. Shiroishi for their valuable advice in this study. We also thank H. Igawa, G. Mise, P. Munclinger, N. Miyashita, N. Hanzawa, G. N. Chelomina, L. V. Frisman, S. Furusawa, A. Frost, I. Kartavtseva, V. P. Korablev, K. K. V. Korobisyna, S-H. Han, M. L. Jin, T. Hosoda, M. A. Iwasa, I. Munechika, N. Nakajima, E. Nevo, M. Pavlenko, M. Sakaizumi, B. Sakumoto, K. Sasaki, J. J. Sato, T. Shimada, M. H. Sinaga, A. Suyanto, S. Suzuki, M. Terashima, M. Tomozawa, S. P. Yasuda, K. Yokoyama and S. Wakana for cooperation in collecting the animals. We would like to thank Heiwa Nakajima Foundation for its generous support. This study was conducted as part of grants-in-aid for Scientific Research supported by the Japan Society for the Promotion of Science (JSPS).

References

Abe K, Noguchi H, Tagawa K *et al.* (2004) Contribution of Asian mouse subspecies *Mus musculus molossinus* to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SKIP analysis. *Genome Research*, **14**, 2439–2447.

- Avice JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA.
- Awasthi M, Bhat K, Anand R (1998) Genetic heterogeneity in the Indian *Mus musculus*. *Biochemical Genetics*, **36**, 245–258.
- Bartish IV, Kadereit JW, Comes HP (2006) Late Quaternary history of *Hippophaë rhamnoides* L. (Elaeagnaceae) inferred from chalcone synthase intron (Chsi) sequences and chloroplast DNA variation. *Molecular Ecology*, **15**, 4065–4083.
- Birky CW Jr, Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, **103**, 513–527.
- Bonhomme F, Catalan J, Britton-Davidian J *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 *et al.* (2007) Species-wide distribution of highly polymorphic minisatellite markers suggests past and present genetic exchanges among House Mouse subspecies. *Genome Biology* [online], **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.
- Boursot P, Din W, Anand R *et al.* (1996) Origin and radiation of the house mouse: mitochondrial DNA phylogeny. *Journal of Evolutionary Biology*, **9**, 391–415.
- Chinen A, Suzuki H, Aplin K, Tsuchiya K, Suzuki S (2005) Preliminary genetic characterization of two lineages of black rats (*Rattus rattus sensu lato*) in Japan, with evidence for introgression at several localities. *Genes and Genetic Systems*, **80**, 367–375.
- Daly M, Rioux J, Schaffner S, Hudson T, Lander E (2001) High-resolution haplotype structure in the human genome. *Nature Genetics*, **29**, 229–232.
- Din W, Anand R, Boursot P *et al.* (1996) Origin and radiation of the house mouse: clues from nuclear genes. *Journal of Evolutionary Biology*, **9**, 519–539.
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.
- Frazer K, Eskin E, Kang H *et al.* (2007) A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature*, **448**, 1050–1058.
- Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics*, **34**, 397–423.
- Gómez-Zurita J, Vogler AP (2006) Testing introgressive hybridization hypotheses using statistical network analysis of nuclear and cytoplasmic haplotypes in the leaf beetle *Timarcha goettingensis* species complex. *Journal of Molecular Evolution*, **62**, 421–433.
- Guénet J, Bonhomme F (2003) Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends in Genetics*, **19**, 24–31.
- Gündüz I, Tez C, Malicov V *et al.* (2000) Mitochondrial DNA and chromosomal studies of wild mice (*Mus*) from Turkey and Iran. *Heredity*, **84**, 458–467.

- Gündüz I, Lopez-Fuster M, Ventura J, Searle J (2001) Clinal analysis of a chromosomal hybrid zone in the house mouse. *Genetical Research*, **77**, 41–51.
- Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution*, **16**, 700–706.
- Horai S, Murayama K, Hayasaka K *et al.* (1996) mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *American Journal of Human Genetics*, **59**, 579–590.
- Huson D, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.
- Jensen-Seaman M, Furey T, Payseur B *et al.* (2004) Comparative recombination rates in the rat, mouse, and human genomes. *Genome Research*, **14**, 528–538.
- 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.
- Koopman WJM, Li Y, Coart E *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.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Macholán M, Munclinger P, Šugerková M *et al.* (2007) Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution*, **61**, 746–771.
- Marshall JT (1998) *Identification and scientific names of Eurasian house mice and their European allies, subgenus Mus Rodentia: Muridae*. Unpublished report, National Museum of Natural History, Washington.
- 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.
- Minezawa M, Moriwaki K, Kondo K (1981) Geographical survey of protein variations in wild populations of Japanese house mouse, *Mus musculus molossinus*. *The Japanese Journal of Genetics*, **56**, 27–39.
- Moriwaki K, Yonekawa H, Gotoh O *et al.* (1984) Implications of the genetic divergence between European wild mice with Robertsonian translocations from the viewpoint of mitochondrial DNA. *Genetical Research*, **43**, 277–287.
- Moriwaki K, Miyashita N, Suzuki H *et al.* (1986) Genetic features of major geographic isolates of *Mus musculus*. In: *Current Topics in Microbiology and Immunology 127, The Wild Mouse in Immunology* (eds Potter M, Nadeau JH, Cancro MP). pp. 255–267. Springer-Verlag, Berlin.
- Moriwaki K, Shiroishi T, Yonekawa H (1994) *Genetics in Wild Mice: Its Application to Biomedical Research*. Japan Scientific Societies Press, Tokyo.
- Nagamine C, Shiroishi T, Miyashita N *et al.* (1994) Distribution of the *molossinus* allele of *Sry*, the testis-determining gene, in wild mice. *Molecular Biology and Evolution*, **11**, 864–874.
- Omoto K, Saitou N (1997) Genetic origins of the Japanese: a partial support for the dual structure hypothesis. *American Journal of Physical Anthropology*, **102**, 437–446.
- Prager E, Orrego C, Sage R (1998) Genetic variation and phylogeography of central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics*, **150**, 835–861.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Sabeti PC, Schaffner SF, Fry B *et al.* (2006) Positive natural selection in the human lineage. *Science*, **312**, 1614–1620.
- Sage R, Atchley W, Capanna E (1993) House mice as models in systematic biology. *Systematic Biology*, **42**, 523–561.
- Sakai T, Kikkawa Y, Miura I *et al.* (2005) Origins of mouse inbred strains deduced from whole-genome scanning by polymorphic microsatellite loci. *Mammalian Genome*, **16**, 11–19.
- Schwarz E, Schwarz H (1943) The wild and commensal stocks of the house mouse, *Mus musculus* Linnaeus. *Journal of Mammalogy*, **24**, 59–72.
- Shimada T, Sato JJ, Aplin KP, Suzuki H (2009) Comparative analysis of evolutionary modes in *Mc1r* coat color gene in wild mice and mustelids. *Genes and Genetic Systems* (in press), **84**, 225–231.
- Shinoda K, Kanai S (1999) Intracemetery genetic analysis at the Nakazuma Jomon site in Japan by mitochondrial DNA sequencing. *Anthropological Science*, **107**, 129–140.
- 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 JC, Reich DE, Goldstein DB *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, Smith N, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.
- Suzuki H, Kurihara Y (1994) Genetic variation of ribosomal RNA in the house mouse, *Mus musculus*. In: *Genetics in Wild Mice: Its Application to Biomedical Research* (eds Moriwaki K, Shiroishi T, Yonekawa H). pp. 107–119. Japan Scientific Societies Press, Tokyo.
- Suzuki H, Miyashita N, Moriwaki K *et al.* (1986) Evolutionary implication of heterogeneity of the nontranscribed spacer region of ribosomal DNA repeating units in various subspecies of *Mus musculus*. *Molecular Biology and Evolution*, **3**, 126–137.
- Tanaka M, Cabrera V, González A *et al.* (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Research*, **14**, 1832–1850.
- Taylor MS, Hellberg ME (2006) Comparative phylogeography in a genus of coral reef fishes: biogeographic and genetic concordance in the Caribbean. *Molecular Ecology*, **15**, 695–707.
- Terashima M, Furusawa S, Hanzawa N *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 *et al.* (2007) Risk of accidental invasion and expansion of allochthonous mice in Tokyo metropolitan coastal areas in Japan. *Genes and Genetic Systems*, **82**, 421–428.
- Tucker PK, Sage RD, Warner J, Wilson AC, Eicher EM (1992) Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution*, **46**, 1146–1163.

- Yonekawa H, Moriwaki K, Gotoh O *et al.* (1980) Relationship between laboratory mice and the subspecies *Mus musculus domesticus* based on restriction endonuclease cleavage patterns of mitochondrial DNA. *Japanese Journal of Genetics*, **55**, 289–296.
- Yonekawa H, Moriwaki K, Gotoh O *et al.* (1981) Evolutionary relationships among five subspecies of *Mus musculus* based on restriction enzyme cleavage patterns of mitochondrial DNA. *Genetics*, **98**, 801–816.
- Yonekawa H, Moriwaki K, Gotoh O *et al.* (1982) Origins of laboratory mice deduced from restriction patterns of mitochondrial DNA. *Differentiation*, **22**, 222–226.
- Yonekawa H, Gotoh O, Tagashira Y *et al.* (1986) A hybrid origin of Japanese mice “*Mus musculus molossinus*”. *Current Topics in Microbiology and Immunology*, **127**, 62–67.
- Yonekawa H, Moriwaki K, Gotoh O *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.
- Yonekawa H, Takahama S, Gotoh O, Miyashita N, Moriwaki K (1994) Genetic diversity and geographic distribution of *Mus musculus* subspecies based on the polymorphism of mitochondrial DNA. In: *Genetics in Wild Mice* (eds Moriwaki K, Shiroishi T, Yonekawa H). pp. 25–40. Japan Scientific Society Press, Tokyo/ S. Karger, Basel.
- Zhang D, Hewitt G (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

Mitsuo Nunome has recently completed Doctor of Environmental Science from Hokkaido University. He is a postdoctoral fellow working on the adaptive evolution of coat colours in mammals. All authors share an interest in the natural evolutionary history of the house mouse, in particular of those from the Japanese Islands.

Supporting Information

Additional supporting information may be found in the online version of this article:

Fig. S1 Geographic distribution of alleles in each of the eight genes examined. Pairs of boxes represent combinations of alleles in each individual. Individuals from Japan are only shown for representative combinations of alleles that represent three geographic areas, Hokkaido, Honshu, and Kyushu. Single boxes indicate those of inbred strains. The colors and numbers of the boxes indicate putative subspecies groups and allele codes, respectively (see Table 1 for detail). The subspecies groups are coloured the same as in Figure 2 (pink: DOM, blue: MUS, yellow: CAS, orange: unassigned).

Fig. S2 Inferring population groups for alleles of each haploid set, estimated using STRUCTURE 2.3. The most likely number of clusters (K) was inferred for each of the eight loci, *Fanca* (F), *Spire2* (S), *Tcf25* (T), *Mc1r* (M), *Tubb3* (T), *Def8* (D), *Afg3l1* (A), and *Dbn1d1* (DB). Variable sites were regarded as characters to infer the number of clusters. Taken in combination with the results of the network analyses (Fig. 2), this analysis allowed us to infer subspecies affiliation for many individual alleles. Codes next to each bar indicate individual sample codes (see Table 1 for detail), gene names represented by one or two letters, and inferred subspecies sources (c: *M. m. castaneus*, d: *M. m. domesticus*, m: *M. m. musculus*, u: unknown).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.