INTRODUCTION

A vertical lightness gradient from a dark back to a light belly is the most common body coloration phenotype in the animal kingdom (Allen et al., 2012), as reported in various animal species, including mammals, birds, fishes and insects (Thayer, 1896; Michaud et al., 1994; Caro, 2005; Kamilar, 2009; Kamilar and Bradley, 2011).

While the house mouse (Mus musculus), widely distributed in Eurasia, is known to have substantial coat color variation between and within local populations, in both primary and secondary distribution areas, including the Japanese archipelago, the evolutionary history of the color variation is poorly understood. To address the ventral fur color variation, we quantified the lightness of museum skin specimens, and found that the southern subspecies, M. m. castaneus (CAS), has high and low lightness in dry and rainy geographic regions, respectively. The northern subspecies, M. m. musculus (MUS), has low and high levels of lightness in the high and middle latitudes of northern Eurasia, respectively. We examined sequence variation of the agouti signaling protein gene (Asip), which is known to be responsible for the ventral fur color. We performed phylogenetic analyses with 196 haplotype sequences of Asip (~180 kb) generated by phasing the whole-genome data of 98 wild mice reported previously. Network and phylogenetic tree construction revealed clustering of haplotypes representing the two subspecies, MUS and CAS. A number of subclusters with geographic affinities appeared within the subspecies clusters, in which the essential results were consistent with those reconstructed with whole mitochondrial genome data, indicating that the phased haplotype genome sequences of the nuclear genome can be a useful tool for tracing the dispersal of geographical lineages. The results of phylogeographic analysis showed that CAS mice with darker ventral fur possessed similar Asip haplotypes across the geographic distribution, suggesting that these haplotypes are major causes of the historical introduction of Asip haplotypes for darker ventral fur in mice from northern India to the peripheral areas, including the Japanese archipelago. Similarly, MUS in East Asia, which has a white abdomen, formed an Asip haplogroup with that from northern Iran, also with a white abdomen.

Key words: agouti signaling protein gene, Mus musculus, ventral fur color variation, nuclear genome sequence, inter-subspecies hybridization
marked evolutionary feature of the countershading pattern is that of "repeated evolution," in which similar traits appear when animals face similar environmental conditions recurrently (Barrett and Schluter, 2008; Gompel and Prud’homme, 2009; Jones et al., 2012). In mammals, more than 150 genes are related to coat coloration (Bennett and Lamoreux, 2003). The agouti signaling protein (Asip) and melanocortin 1 receptor (Mc1r) genes are major determinants of red-yellow (pheomelanin) and black-brown (eumelanin) pigments, which are synthesized in melanocytes (Klungland et al., 1995; Morgan et al., 1999; Hoekstra and Nachman, 2003; Fontanesi et al., 2006, 2010; Kingsley et al., 2009; Mallarino et al., 2016; Suzuki et al., 2020). Switching between the pigment types is regulated by Mc1r, expressed on the cell surface of melanocytes. The production of eumelanin is suppressed by the presence of the antagonist agouti, the product of Asip, the antagonist/reverse agonist of Mc1r, which initiates the production of pheomelanin. In the house mouse Mus musculus, Asip harbors two exon 1-associated promoter regions with different functions related to the ventral coat color and agouti pattern of individual hairs with black and yellow bands (Vrieling et al., 1994). Both genes of Mc1r and Asip are considered to have important roles in the evolution of rapid changes in coat coloration (e.g., Vrieling et al., 1994; Chen et al., 1996; Hoekstra et al., 2004, 2005; Hoekstra, 2006; Iwasa et al., 2018; Sakuma et al., 2019). Geographical variation has also been investigated and it has been reported that natural selection is involved (Hoekstra et al., 2004, 2005; Kingsley et al., 2009; Mullen et al., 2009; Marín et al., 2018).

The color of the coat, in particular the ventral fur color, often differs depending on the geographic location in M. musculus (Marshall, 1998; Slábová and Frynta, 2007; Lai et al., 2008) and has been used as a diagnostic marker in the mouse (Schwarz and Schwarz, 1943; Marshall, 1998). The hair color of this species, which is currently divided into three subspecies groups, i.e., M. m. castaneus (CAS: southern Asia, Southeast Asia and southern China), M. m. domesticus (DOM: western Asia and western Europe) and M. m. musculus (MUS: northern Eurasia) (e.g., Suzuki et al., 2013; Takada et al., 2013), varies both within and between the subspecies (Takada et al., 2021). An understanding of the genetic background of phenotypic variation would allow observations of coat color variation in wild populations to provide insights into phylogenetic background (Kambe et al., 2011; Suzuki, 2013; Kodama et al., 2015; Sasamori et al., 2017; Marín et al., 2018). There are two possible explanations for the hair color variation in the peripheral regions: the alleles of hair color in the source region may have been transmitted to the surrounding regions; or new alleles that arose in situ may have generated the diversity of hair color (Myoshu and Iwasa, 2018). However, to date, no genetic study has investigated the historical episodes of geographic variation of coat color in M. musculus.

To test these explanations, it would be useful to examine genetic variation in the hair color-related genes, understand their spatiotemporal dynamics, and construct an evolutionary scenario. Sequence variation in the coat color-related gene Mc1r (Nunome et al., 2010; Kambe et al., 2011; Kodama et al., 2015; Kuwayama et al., 2017) has been used for phylogenetic inference of rodents, including M. musculus. However, the Asip gene responsible for ventral coat color has not been analyzed in M. musculus. We recently explored the whole-genome sequences of 98 wild-derived mice mainly from Asia (Fujiwara et al., 2021). This allowed us to reconstruct the prehistoric distribution pathways associated with humans based on mitochondrial whole-genome sequences (Li et al., 2021). For example, these results suggest that CAS and MUS arrived in the Japanese archipelago about 3,000 years ago, and since this time they have experienced genetic admixture (Fujiwara et al., 2021; Li et al., 2021).

The present study was performed to assess the evolutionary background of the geographic variation in ventral coat coloration in M. musculus. First, we examined the variation in ventral coloration by spectrophotometry to determine the degree of variability among individuals. Next, we performed phylogeographic analyses with Asip, encompassing a portion of about 180 kb of the mouse genome, utilizing the mouse genome sequences that we reported previously (Fujiwara et al., 2021). This enabled us to establish the coat color variation in the homelands of M. musculus and also the spatiotemporal dynamics of coat color variation as the distribution of M. musculus expanded to follow humans.

RESULTS

Coat color measurements We conducted quantitative measurement of ventral fur color in terms of the CIE L*a*b*-model color parameters in 114 mouse skin specimens from 35 localities (Fig. 1A; Supplementary Table S1). While there were clear differences in the values of lightness (L*) within a locality, there was also a tendency for L* to have roughly similar values within a locality. The averaged L* values representing each of the localities ranged from 40–80 and were classified into two groups for convenience: "gray" (40 < L* < 65) and "white" (70 < L*). The gray group was further divided into two subgroups: dark gray (L* < 50) and light gray (50 < L* < 65). Exceptionally, 10 samples from Wuhan, China exhibited substantial variation in the L* values, from 43 to 57 (Supplementary Table S1). We plotted the groups on a geographic map, which revealed an apparent geographic affinity with latitudinal zonation: the high-, mid-, and low-latitude zones harbored the light gray, white and
Fig. 1. Geographic variation of ventral fur color of *Mus musculus*. Box plots show the lightness (L*) profile of the ventral fur in museum-stored skin specimens from 31 localities with median, first and third quartiles, and the minimum and maximum values (A). The L* value is the median for mice (n = 1–6) from the same locality. The L* values were divided into three groups corresponding to haplogroups, namely dark gray (40.8–48.0, brown), light gray (51.8–62.7, gray) and white (71.8–79.8, yellow), and plotted on a map (B). The ventral color group in Nepal was adapted from Kishimoto et al. (2021). Dotted lines on the map show the approximate borders of the three major subspecies: *M. m. castaneus* (CAS), *M. m. domesticus* (DOM), and *M. m. musculus* (MUS).

dark gray color groups, respectively (Fig. 1B). The white color group was present in the western range of CAS (Iran and Pakistan), the eastern range of MUS (China, Korea and Japan) and the eastern range of DOM (Iran). The light gray group was present in DOM and MUS from Eastern Europe, Russia and northern Japan. The dark gray group included mice from southern India, Sri Lanka, Vietnam and Indonesia.
Sequences of the Asip locus and network analysis

From 98 wild mouse genome sequences (Fujiwara et al., 2021; Supplementary Table S2), the 180-kb region from the 3’ end of Eif2s2 to the 3’ end of Asip (specifically, to the end of the exon 4 sequence) was selected (Fig. 2A). We generated a haplotype data set by phasing, resulting in 196 haplotype sequences with 6,671 segregating sites that were used in subsequent phylogenetic analyses.

A neighbor-net (NN) network constructed using the entire Asip gene sequence revealed an overall pattern of clusters representing the three major subspecies groups of CAS, DOM and MUS (Fig. 2B). The MUS network consists of many clusters, and for convenience we divided them into two groups (referred to here as MUS-I and MUS-II) based on geographic information (Fig. 2B). The NN network showed the presence of recombinant-like haplotypes that branched off at intermediate positions of the major subspecies clusters (Fig. 2B). We detected haplotypes integrated into different subspecies-specific clusters in the NN networks or that were present as isolated branches in the middle positions of the subspecies clusters. To identify recombination fragments between haplotypes representing each subspecies, we partitioned the 180-kb sequence into four segments (domains I–IV; ~45 kb each) (Fig. 2A). Domain II contained the ventral-specific promoter region. We constructed NN networks with the four partitioned domains (Fig. 2C). The DOM sequences were shown to be located close to the haplotype cluster of CAS in domain III and embedded within the CAS cluster in domain II (Fig. 2C), suggestive of introgression from CAS to DOM prior to the movement with humans in prehistory. Of the 196 haplotypes, 175 were judged to be non-recombinants having identical cluster formation between the four domains in the networks constructed. The remaining 21 haplotypes either showed different clustering patterns among the four networks or were placed at intermediate positions outside the subspecies-specific clusters, being designated as recombinants (Supplementary Table S2). For example, a recombinant haplotype of Tallinn (Estonia) had MUS-type for domains I–III and DOM-type for domain IV.

Many of the recombinant haplotypes were recovered from the peripheral regions, including the Japanese archipelago (Fig. 2C). Five haplotypes from Hokkaido (four from Sapporo and one from Takikawa) were shown to be similar, with the upper part being MUS-II and the remaining parts CAS. In the Busan (Korea) haplotype (MG0440), the upper half (domains I and II) was MUS-II, whereas the lower half (domains III and IV) involved CAS-derived segments. Recombinants were recovered from the homeland area in northern Iran. Two sequences from northern Iran, MG5135_1 from Now-shahr, and MG0417_1 from Mashhad, were shown to have DOM-derived segments in domain I. The segments of domains III and IV were located in the MUS cluster, while that of domain II was located close to but slightly outside the MUS cluster.

Phylogenetic tree analysis

We constructed a neighbor-joining (NJ) tree for the 175 haplotypes with the K2P substitution model and the option of deletion of all gaps, excluding the 21 predicted recombinant haplotypes (Fig. 3A). The NJ tree showed monophyletic clusters for DOM and MUS. In contrast, it showed a markedly divergent pattern for the CAS haplotypes, consistent with previous results of McIr analysis (Kodama et al., 2015).

More detailed branching patterns within subspecies clusters for CAS (Supplementary Fig. S1) and MUS (Supplementary Fig. S2) also appeared in the NN networks. In these phylogenetic analyses, three subclusters exhibited a bush-like clustering pattern with short terminal branches (termed haplogroups a, b and c) in the CAS cluster. The grouping of the three haplogroups exhibited a geographic structure (Fig. 3B). Haplogroup a was represented by haplotypes from southern India. Haplogroup b comprised broadly distributed haplotypes from the northern part of the Indian subcontinent, Nepal, Bangladesh, Yunnan (southeastern China), Sri Lanka and the Russian Far East, including Sakhalin. Haplogroup c was represented by haplotypes from areas including southern China, Vietnam, Indonesia, the Russian Far East and Japan.

Phylogenetic analysis divided MUS into MUS-II, consisting of haplotypes mainly from China, Korea and Japan, and to a lesser extent from Iran and Russia, and MUS-I, consisting of haplotypes from other regions (Fig. 4). The members of MUS-I were distributed over wide areas of the northern part of Eurasia from eastern Europe to the Russian Far East, including Sakhalin. The spatial pattern is essentially consistent with a previous nuclear gene analysis (Nunome et al., 2010), in which MAS was divided into two geographic groups, “northern” and “southern,” corresponding to MUS-I and MUS-II of Asip, respectively. It is notable that haplotypes from Ukraine (Donetsk), the Russian Far East (Birakan and Krementui) and northeast China (Mohe and Manzhouli) formed a cluster connected with short branches (marked by an asterisk in Fig. 4A). MUS-II had a branching pattern expanding from west to east across five regions in sequence: northern Iran (region 1), western China (region 2), northern China (region 3), the Korean Peninsula (region 4) and the Japanese archipelago (region 5) (Fig. 4; Supplementary Fig. S2). This was consistent with the results of a previous study of mitochondrial DNA (Li et al., 2021).

These results indicate close relationships between the source and peripheral areas in the Asip haplogroups, suggestive of the involvement of human-associated migration in determining types of ventral fur color (Fig. 5). In CAS,
Fig. 2. Phylogenetic inference with the genome sequence of the Asip locus (~180 kb). The gene region was obtained from whole-genome data of 98 wild mice (Fujiwara et al., 2021). Homology plot analysis of the gene region showing the positions of the inverted repeats and exon regions of Asip (A). Asip harbors the amino acid coding region, included in exons 2, 3 and 4 and two exon 1-associated promoter regions with different functions. The regulatory region known as the ventral-specific promoter region (shown below the plot) is located 100 kb upstream of the coding region and contains two exon 1-associated promoters, namely 1A and 1A', responsible for ventral coat color, and pseudo exon 1 (ps1A) (Vrieling et al., 1994; Chen et al., 1996). Convergent arrows mark the inverted repeats (see Chen et al., 1996 and Sakuma et al., 2019 for details). The other region contains two promoters, 1B and 1C, and is located 30 kb upstream of the coding region; it is related to the agouti pattern of individual hairs with black and yellow bands (Vrieling et al., 1994). Neighbor-net networks were constructed with the whole gene region (B) and partial segments partitioned into four domains (I, II, III and IV) (C). Subspecies-specific branches are shown in different colors of gray (CAS), red (DOM) and blue (MUS). Branches harboring predicted recombinant haplotypes are shown in black. Three haplogroups (a, b and c) representing recent rapid expansion events are marked with yellow curved lines.
Fig. 3. Neighbor-joining tree based on the genome sequence of the Asip locus (~180 kb), with labels for the three major subspecies groups, *M. m. musculus* (MUS), *M. m. castaneus* (CAS) and *M. m. domesticus* (DOM) (A). The 21 haplotypes predicted to be recombinants were excluded (see text for details). Three prominent clusters of geographic affinity consisting of haplotypes connected by short branches are marked (haplogroups *a*, *b* and *c*). Node scores (> 50%) are bootstrap values from 1,000 bootstrapped trees. The geographic localities of the haplotype members (square) of the three haplogroups *a* (purple), *b* (orange) and *c* (green) are shown in maps (B). Localities (circles) examined are shown with subspecies-specific mitochondrial DNA types: MUS (blue), CAS (yellow), DOM (red) and Nepalese lineage (gray) (for details, see Li et al., 2021).
Mouse agouti gene phylogeography

Fig. 4. Neighbor-joining tree based on the ~180-kb genome sequence of the Asip locus, with haplotypes assigned to be *M. m. musculus* (MUS) and *M. m. domesticus* (DOM) (A). The MUS cluster can be conveniently divided into a group called MUS-II, which consists of haplotypes mainly from China, Korea and Japan, and MUS-I, which includes other haplotypes, as observed in the NN network (Fig. 2B). Haplotypes of MUS-II consist of two groups: a five-member subcluster labeled ’MUS-II*’ and the other members shown by a blue collapsed triangle labeled ’MUS-II**’, the latter divided into five subgroups representing each of the geographic regions from northern Iran to the Japanese archipelago. Node scores (> 50%) are bootstrap values from 1,000 bootstrapped trees. A non-collapsed pattern for MUS-II** is shown (B). It shows stepwise diversification of the clusters from west to east, divided into five steps for convenience (1–5). The haplogroups of MUS-I and MUS-II are marked with unique symbols and plotted on the map (C). Localities (circles) examined are shown with subspecies-specific mitochondrial DNA types: MUS (blue), CAS (yellow), DOM (red), and the Nepalese lineage (gray) (for details, see Li et al., 2021).
Fig. 5. Possible prehistoric human-associated migration routes of the house mouse (*Mus musculus*) based on phylogeographical analysis of *Asip*, with lines connecting the collection sites for the three CAS haplogroups a (purple), b (orange) and c (green) as well as for the subcluster MUS-II (blue) (A). The black dotted line shows the hypothetical route that the MUS-I mouse took from Iran to Sakhalin, moving eastward through the Russian continental area. The five stepwise eastward movements (1–5) of the mitochondrial lineage of MUS-1 are shown (Li et al., 2021). The phylogenetic trees of mtDNA of the CAS-1 and MUS-1 lineages (for details, see Li et al., 2021) (B). Estimated divergence times for the five steps of the eastward movements are shown to the right of the MUS-1 tree. Branches that are thought to correspond to the three *Asip* haplogroups a, b and c with scattered regional diversity in CAS are indicated with their respective colors. The initial lineage divergence of CAS-1 to C1a and C1b is indicated.

the allele of *Asip*, which darkens the abdominal coat color (“dark gray”) somewhere in northern India, was transported to the neighboring eastern area. In MUS, the coat color type “white” was detected in the northeastern part of Iran and the *Asip* type in northern Iran and the easternmost peripheral region, the Japanese archipelago.

**DISCUSSION**

**Quantitative evaluation of regional differences in the lightness of ventral hair color** In this study, we examined the phylogenetic relationships of the *Asip* gene to better understand evolutionary implications of coat
color variation in the Eurasian house mouse, which has dispersed alongside humans since prehistory. The two subspecies groups of *M. musculus* that dispersed over a wide area of Asia exhibit considerable diversity in ventral coat color and form three distinct groups based on the lightness (L*) of the ventral fur color (Fig. 1): dark gray (L* < 50) and white (L* > 70) for CAS, and light gray (50 < L* < 65) and white (L* > 70) for MUS in their respective western and eastern distribution areas. DOM in western Europe tends to have light gray fur (50 < L* < 65) similar to that observed in the western region of MUS. However, individuals from Ahvaz in westernmost Iran characterized as having the DOM background are white (L* > 70). MUS distributed in the northern part of Eurasia is light gray (50 < L* < 65) at high latitudes, while that occurring at low latitudes, covering China, Korea and Japan, is white (L* > 70). In general, these results support those reported by Lai et al. (2008) indicating that there is a relationship between coat color and rainfall. Considering the historical migrations confirmed by mitochondrial DNA sequences (Li et al., 2021) and the *Asip* sequences (this study), the light and dark coat colors in the secondary distribution areas are likely to be shaped by the eastward movements of MUS and CAS, respectively, as discussed in detail later. Clear variability in coat color was observed in China and the Japanese archipelago (Supplementary Table S1; Fig. 1B). In Wuhan, for example, the majority of L* values indicated dark gray but included some light gray, which could be interpreted as a result of the inclusion of MUS in the CAS background.

**Use of the genome sequences of the *Asip* locus as a genealogical marker** In this study we used the phased genome sequences of the *Asip* locus (around 180 kb) as a genealogical marker to assess the evolutionary history of coat color variation in *Mus musculus*. As a result, we observed subspecies-specific clustering patterns in the majority of haplotypes (Figs. 2 and 3). In contrast, recombinant haplotypes between subspecies were also observed to some extent (21/196) (Supplementary Table S2). Many of these recombinants were detected in the marginal regions, such as the Russian Far East and the Japanese archipelago, probably as a consequence of secondary contacts between individuals from multiple migrations of different subspecies at certain ancient times. For example, recombinant haplotypes were recovered from China (CAS/MUS), Korea (CAS/MUS), Japan (CAS/MUS) and Primorye in Russia (CAS/DOM/MUS). The recombinant haplotypes in Japanese mice each appear to be the product of a single break point along the 180-kb segment, which is similar to what was inferred from the previous analysis of *Mc1r* and its surrounding gene regions (200 kb, Nunome et al., 2010). It is possible to infer the time when hybridization occurred based on the fragment size, number of generations of heterozygosity stage, and recombination rate (Stephens et al., 2001). In their analysis of the gene region containing *Mc1r*, Kuwayama et al. (2017) examined a 5-Mb segment containing *Mc1r* and observed that the length of CAS fragments incorporated is ~180 kb, from which they predicted that about 800 generations ago CAS and MUS hybridization started, assuming a recombination rate of 0.5 cM/Mb (Jensen-Seaman et al., 2004), and heterozygous state in every generation. The number of generations per year is uncertain, ranging from one to three, and heterozygosity with different subspecies of haplotypes did not occur in every generation, so there are many unknowns, but it is safe to assume that interspecific hybridization has occurred sometime in the past several thousand years in the Japanese archipelago.

A DOM-like fragment was detected in a MUS population in northern Iran (localities Nowshahr and Mashhad; Supplementary Table S2). We speculate that this introgressed DOM fragment was subsequently moved to the periphery. Our previous studies revealed the presence of short DOM elements embedded in the genome of mice of the MUS background from northern Japan (Nunome et al., 2010; Kuwayama et al., 2017; Isobe et al., 2018). The event of integration of the short DOM segments can be easily discriminated from very recent (e.g., a few decades ago) introgression events, judging from the large DOM fragments of 5 Mb and 3 Mb (Nunome et al., 2010; Kuwayama et al., 2017). To date, the issue of the short DOM fragments in the Japanese archipelago remains unresolved. Future genome-wide analysis of this issue focusing on intersubspecies recombination events may provide a plausible scenario for the presence of short DOM segments in wild Japanese mice.

Phylogenetic evaluation using recombinant haplotypes may be useful not only for relatively recent hybridization, but also for hybridization events that occurred in older times. In the NN network of domains II and III, the cluster of CAS haplotypes encompasses the cluster of DOM haplotypes (Fig. 2C). This indicates the possibility of genetic introgression between CAS and DOM in the past, prior to human-associated migration, as suggested in our previous study on nuclear gene sequences (Kodama et al., 2015). The analysis of nuclear genome sequences can provide useful information to unravel the complex history of interactions between different lineages from the distant past to the present.

Moreover, assessment of recombinant haplotypes provides information on which regions of the gene are correlated with a phenotype of interest. Domain II contains a promoter region specific for ventral coat color (Fig. 2A; Vrielings et al., 1994). Individuals in Hokkaido with haplotypes incorporating a CAS-derived fragment in this region have darker abdominal coloration. This suggests
that the type of this segment is an important determinant of the type of ventral coat color. Further studies are required to identify the mutation responsible at the DNA sequence level.

Do nuclear gene sequences provide temporal information in phylogenetic analysis within species, as mtDNA does? The branch length of nuclear gene sequences is known to be greatly influenced by the size of the effective population (Hudson, 1990). For example, the fact that CAS clusters are connected to each other by long branches, as seen in the NN network (Fig. 2) and NJ tree (Fig. 3), may indicate greater genetic diversity and historically larger population sizes. Therefore, in estimating the age of branching and dispersal, it will be important to carefully match the results of mtDNA analysis.

**Dispersal of Asip lineages of CAS in association with humans in prehistoric times in southeastern Asia** The ventral coat coloration of “dark gray” occurred in a wide area, including southern India, Southeast Asia and southern China (Fig. 1). The southern foothills of the Himalayas, northern India (here represented by mice from Delhi) and Nepal are likely part of the homeland of CAS, and the candidate source area of the allele responsible for the darker color phenotype. According to Marshall (1998), the dark gray mouse, *Mus castaneus,* occurs in the eastern part of the region of northern India, while mice in the western part of northern India, represented by the locality Delhi in this study, have a white abdomen. In this study, however, the four haplotypes from Delhi exhibited trends of high nucleotide divergence from one another, and two were integrated into haplogroup b (Fig. 3A, Supplementary Fig. S1), which may have coincided with prehistoric human migration and represent the dark gray fur color. Mice from Nepal are known to harbor highly polymorphic CAS-related mtDNA haplotypes (Prager et al., 1998) and are known to be a distinct lineage (NEP) that differs from CAS, MUS, DOM and *M. m. gentilulus* from Yemen (Duplantier et al., 2002; Sakuma et al., 2016), indicating the presence of NEP haplotypes from around half a million years ago (Li et al., 2021). However, *M. musculus* from central Nepal (Kishimoto et al., 2021) is shown to have dark-colored ventral fur, with L* values ranging from 63 to 50, corresponding to the class of “light gray” and not “dark gray,” suggesting that mice from central Nepal may not be responsible for the migration of dark gray fur that is now distributed in the southern part of Asia. Taking these observations together, it is possible to conclude that CAS mice expanded their range eastward with humans in prehistoric times from the vicinity of northern India.

The haplotypes for southern India, Bangladesh, Southeast Asia, southern China and Japan, which are part of the current CAS territory, form three bush-like clusters: haplogroups a, b and c (Fig. 3A). The presence of clusters connected by short branches indicates that these haplogroups experienced bottlenecks, which may have been the result of founder effects due to migration. By focusing on these clusters, it will be possible to understand the early spatial dynamics that accompanied the migration of prehistoric people. The temporal aspect of the migration event can be understood by examining the spatial dynamics of Asip (this study) and mtDNA (Li et al., 2021), and then referring to the divergence age obtained by mtDNA analysis.

Our previous study on whole mtDNA sequences demonstrated the haplotype diversification of CAS within the last 8,000 years in these geographic areas, including southern India (Li et al., 2021). Together with the mitogenome analysis, it is possible to use the Asip data to infer the spatiotemporal dynamics of the human-associated prehistoric movements of CAS (Fig. 5). In the mtDNA data, an earlier prominent divergence event in the CAS lineage (CAS-1; see Li et al., 2021) resulted in the emergence of haplogroups including the haplotype from central and southern India (C1a2; see Li et al., 2021). This implies that mice carrying C1a2 of the mtDNA lineage and haplogroup a of Asip moved from the northern part of the Indian subcontinent to the southeastern part of the Indian subcontinent with humans approximately 3,500 years ago (Fig. 5B), which might have been related to the generation of haplogroup a.

In the NJ tree, the cluster comprising haplogroups b and c is monophyletic with a bootstrap value of 90% (Fig. 3). Accounting the geographic origins of the haplotypes, this cluster is assumed to represent C1a1 as described in the analysis of the mitogenome (Li et al., 2021). Diversification of cluster C1a1 seems to have occurred 5,600–4,300 years ago based on the mitogenome analysis (Li et al., 2021). While haplogroup b is weakly monophyletic, haplogroup c is monophyletic with a high bootstrap value of 98%. Based on the geographic distribution of these haplogroups shown in Fig. 3B, we infer that haplogroups b and c originated from northern India and southern China, respectively.

Given the high nucleotide diversity of members of haplogroup b from the west, an eastward migration from the Indian subcontinent to southwestern China (Yunnan Province) was inferred. However, members of haplogroup c are rare in southern China, where haplogroup c predominates, except in Yunnan, and are sporadically distributed in northeast Asia, including Khabarovsk and Poronaysk (central Sakhalin). Therefore, haplogroup b was probably introduced into both northern and southern China at some time in the ancient past. Assuming that it was linked to the divergence of the Indian and southern Chinese mtDNA lineage (C1a1) into two separate lineages (Li et al., 2021), the geographical development of haplogroup b is estimated to have started about 5,600 years ago. It may then have been replaced in the north.
by MUS and in the south by CAS haplogroup c until the present.

Asip haplogroup c is considered to be associated with distribution events covering a wide area of Southeast Asia, southern China and East Asia, including migration events, probably from Vietnam, to three sites in Indonesia (Bali and two sites on Java, i.e., Lembang and Bogor); each locality has a specific Asip haplotype (Fig. 5). This may be because different haplotypes were introduced and fixed locally, or different haplotypes were fixed in each location via the introduction of polymorphic haplotypes. The Asip data for haplogroup c are consistent with mitogenome analyses indicating a dispersal event of CAS 3,500–2,500 years ago in southern and southeastern China and the Japanese archipelago (Li et al., 2021). Interestingly, our data of Asip suggest that Taiwan and Okinawa were involved in the over-sea migration. Two mtDNA phylogenetic groups were probably introduced into the Japanese archipelago (Li et al., 2021). Whether polymorphic linkages were introduced from a single source, or whether there were two source areas, such as the Yangtze River and the Pearl River, should be clarified by analyzing other nuclear genes in the future. In either case, the introduction of M. musculus from southern China is confirmed as an historical event. Given the already widespread distribution of CAS-related sequences from Hokkaido to Okinawa, it is conceivable that human activity related to this event substantially affected the Japanese archipelago. These results are useful for understanding prehistoric human migration to and through the Japanese archipelago, and our findings strengthen archaeological and human genomic knowledge related to pre- and post-Yayoi period migration events (Leipe et al., 2020; Jinam et al., 2021).

Eastward spread of northern and southern lineages of MUS shaped the geographic variation of coat color As for the MUS subspecies, the majority of its current distribution range, except for the predicted source area of northern Iran, is in areas where it has expanded its distribution by accompanying humans (Li et al., 2021). Our present phylogenetic analysis divided MUS into two geographical groups, MUS-I and MUS-II (Fig. 4, Supplementary Fig. S2), in accordance with our previous analyses of nuclear gene sequences (Nunome et al., 2010; Kodama et al., 2015). MUS-I and MUS-II, the northern and southern groups, were shown to have the light gray and white ventral fur color, respectively (Fig. 1), indicating that the two groups have separate histories in their long-distance movements.

The MUS-I group consists of a number of clusters, which we divided into three clusters for convenience in order to understand their geographical distribution (Fig. 4); each cluster is widely distributed from the Caspian Sea area to the southern fringe of Russia’s border with China, Primorye and Sakhalin. This suggests an eastward migration, but due to the lack of mtDNA analysis, it is unclear when the species began to expand; the expansion is assumed to have started about 15,000 years ago, the same as MUS-II (Li et al., 2021). Overall, the diversity of MUS-I was higher than that of MUS-II (Fig. 4A), suggesting that a high degree of diversity was maintained during the migration without an efficient bottleneck effect. However, MUS-I has a cluster of short branches (marked with an asterisk in Fig. 4A) that extends from Ukraine to the Russian Far East and northeast China (represented by Manzhouli), suggesting a relatively recent expansion and hence possible multiple trans-Eurasian migration events in MUS-I.

MUS-II exists as a sequential, nested cluster in China (regions 2–3), Korea (region 4) and Japan (region 5) (Fig. 4C), providing a possible pathway for the movements of MUS in northern Asia related to prehistoric humans (Fig. 5). This supports the hypothesis of an eastward migration associated with the development of human agriculture over the past 15,000 years, as suggested by the results of mitogenome analysis (Li et al., 2021). Interestingly, the Asip data indicate that a haplotype (MG0417_2) from Mashhad in northern Iran, where mice have a white abdomen (Fig. 1), is integrated into the same group, MUS-II (Fig. 4B, Supplementary Fig. S2). It is conceivable that MUS with lighter abdominal hair color originated from northern Iran (Fig. 5). However, it is unclear whether the haplotype that causes the white abdomen was selected by chance during the widespread dispersal of humans, or whether it became dominant as a result of natural selection during migration through arid regions, such as the Taklamakan Desert. In any case, it is possible that the haplotype originating in northern Iran was introduced to western and northern China, the Korean Peninsula and the Japanese archipelago as a result of eastward expansion, suggesting that historical factors played the major role in establishing the white belly coat color of MUS in East Asia.

The last process of the MUS-II expansion was the transfer from the Korean peninsula to the Japanese archipelago. The cluster of MUS in the Japanese archipelago subdivides into two clusters with short branch lengths (5a and 5b in Supplementary Fig. S2). These “bush-like” clusters connected by short outer branches in the nuclear gene lineage may indicate divergence associated with bottlenecks caused by founder events. The previous mtDNA analysis showed that the cluster in the Japanese archipelago has experienced rapid expansion events (Li et al., 2021). We should see if nuclear gene sequences can be used to analyze population dynamics in the future. In the Japanese archipelago, MUS appears to be mixed with CAS from southern China and is associated with darker abdominal coat coloration. In fact, in mice from northern Kyushu (Fukuoka), the ventral fur
was black (Fig. 1) and one mouse examined retained the CAS haplotype (Supplementary Table S2). While CAS haplotypes are found in various areas of the Japanese archipelago, CAS is rare in the Korean Peninsula. We found CAS fragments in recombinant haplotypes in one mouse (MG0444) from Busan, Korea, but it remains to be determined whether the haplotypes originated independently in the Korean Peninsula or were recently introduced from Japan to Korea.

**Conclusion** In this study, we focused on the polymorphism of ventral coat color in wild mice and the genomic sequence diversity of the Asip gene, which is known to be responsible for the polymorphism. The results confirmed the geographic diversity within the three subspecies groups; the ventral coat color consisted of white and various levels of gray, conveniently represented by two forms: light gray and dark gray. Alignment of the Asip sequences revealed subspecies-specific haplotypes and recombinant haplotypes among them. It is conceivable that the Asip sequence can be used to trace the history of human-associated migrations, which may have begun as early as 15,000 years ago based on previous mitogenome sequence analysis (Li et al., 2021). In the most recently colonized areas of Asia, we observed the presence of Asip sequences for CAS in the case of dark gray ventral fur and for MUS in the case of white ventral fur. In the Japanese archipelago, the Asip haplotype of MUS probably arrived around 3,000 years ago (Li et al., 2021) from western and northern China and the Korean Peninsula to Kyushu, and expanded from Kyushu to Hokkaido. The CAS Asip haplotype is thought to have come from southern China to coastal areas of the Russian Far East and the Japanese archipelago shortly before MUS entered Japan. This study on Asip suggested that the dark gray and white fur types in the Asian periphery originated from somewhere in northern India and northern Iran, respectively. Further detailed studies are required to analyze the genetic changes in Asip that have led to the changes in coat color.

**MATERIALS AND METHODS**

**Measurement of coat color variation** An extensive set of wild-derived mouse specimens collected from areas across Eurasia has been preserved in the National Institute of Genetics (NIG) in Japan (now stored at the National Museum of Nature and Science, Tsukuba, Japan). In total, 114 mouse skin specimens with NIG codes from 35 localities were used in this study (Supplementary Table S1). We examined individual differences in the ventral coat color of all the mice as described previously (Sakuma et al., 2016). A Minolta CM-700D spectrophotometer (Minolta Camera Co., Osaka, Japan) with a 6-mm diameter window was used for all measurements. Three measurements were taken from random locations on the ventral side. A colorimetry model (CIE L*a*b* model) was employed to decompose the initial color (RGB) image into three gray levels: (L*), lightness from black to white [0, 100]; (a*), component encoding the initial color from green to red [-60, 60]; and (b*), component encoding the initial color from blue to yellow [-60, 60]. Skin specimens were available for some of the mice with which we compared Asip sequences, and we measured the ventral fur color.

**Phylogenetic analyses of the Asip genome sequences** We used the whole genome sequences of 98 wild mice (Supplementary Table S2), which we had previously generated, aligned to the reference mouse genome GRCm38 (mm10) and deposited in the GenBank/EMBL/DDBJ database under the project name PRJDB11027 (see Fujiwara et al. (2021) for further details). Haplotype inference was performed with SHAPEIT4 (Delaneau et al., 2019) using the recombination map reported by Morgan et al. (2017). We added prefixes (“1_” or “2_”) to the individual names in order to designate haplotypes. The datasets were trimmed to vcf files corresponding to the region of chr2:154,871–155,051,012, which includes the Asip region. The vcf files were converted to fasta files using the bcftools-1.10.02 consensus command. Missing variants were not skipped but were instead coded as “N.” The 196 sequences were aligned using MAFFT, an online multiple sequence alignment tool (Katoh et al., 2019). The 196 haplotype Asip sequences were used to construct a NN network in SPLITS TREE, version 4.11.3 (Huson and Bryant, 2006). The number of segregating sites was assessed using MEGA X (Kumar et al., 2018). After excluding predicted recombinant haplotypes, a NJ tree was constructed with MEGA X. We performed 1,000 bootstrap replicates to assess the robustness at each node.

We are grateful to Kuniya Abe, Hidetoshi Ikeda, Nobumoto Miyashita, Mitsuo Nunome, Daiki Usuda, Hiromichi Yonekawa, Kazuo Moriwaki and Kimiyuki Tsuchiya for their support in conducting this study. We thank two anonymous reviewers and Ian Smith for their comments, which helped us to improve the manuscript. This study was supported by JSPS KAKENHI Grant Number JP18H05508.

**REFERENCES**

Allen, W. L., Baddeley, R., Cuthill, I. C., and Scott-Samuel, N. E. (2012) A quantitative test of the predicted relationship between countershading and lighting environment. Am. Nat. 180, 762–776.

Barrett, R. D. H., and Schluter, D. (2008) Adaptation from standing genetic variation. Trends Ecol. Evol. 23, 38–44.

Bennett, D. C., and Lamoreux, M. L. (2003) The color loci of mice—a genetic century. Pigment Cell Res. 16, 333–344.

Caro, T. (2005) The adaptive significance of coloration in mammals. Bioscience 55, 125–136.
Chen, Y., Duhl, D. M. J., and Barsh, G. S. (1996) Opposite orientations of an inverted duplication and allelic variation at the mouse agouti locus. Genetics 144, 265–277.

Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L., and Dermitzakis, E. T. (2019) Accurate, scalable and integrative haplotype estimation. Nat. Commun. 10, 5436.

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

Fontanesi, L., Forestier, L., Alain, D., Scotti, E., Beretti, F., Deretz-Picoulet, S., Pecchioli, E., Vernesi, C., Robinson, T. J., Malaney, J. L., et al. (2010) Characterization of the rabbit agouti signaling protein (ASIP) gene: transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat color. Genomics 95, 166–175.

Fontanesi, L., Tazzoli, M., Beretti, F., and Russo, V. (2006) Mutations in the melanocortin 1 receptor (MC1R) gene are associated with coat colors in the domestic rabbit (Oryctolagus cuniculus). Anim. Genet. 37, 489–493.

Fujisawa, Y., Kawai, Y., Morigaki, K., Takada, T., Shiroishi, T., Saitou, N., Suzuki, H., and Osada, N. (2021) Insights into Mus musculus subspecies population structure across Eurasia revealed by whole-genome sequence analysis. BioRxiv. doi: https://doi.org/10.1101/2021.02.05.429881

Gompel, N., and Prud’homme, B. (2009) The causes of repeated genetic evolution. Dev. Biol. 332, 36–47.

Hoekstra, H. E. (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity 97, 222–234.

Hoekstra, H. E., Drumm, K. E., and Nachman, M. W. (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and natural genes. Evolution 56, 1239–1341.

Hoekstra, H. E., Krenz, J. G., and Nachman, M. W. (2005) Local adaptation in the rock pocket mouse (Chaetodipus intermedius): natural selection and phylogenetic history of populations. Heredity 94, 217–228.

Hoekstra, H. E., and Nachman, M. W. (2003) Different genes underlie adaptive melanism in different populations of rock pocket mice. Mol. Ecol. 12, 1185–1194.

Hudson, R. R. (1990) Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology 7, 1–44.

Huson, D. H., and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

Iwase, M., Nonume, M., Katakura, K., and Suzuki, H. (2018) Evolutionary dynamics of copy number and meiotic recombination in murine 5S rDNA: possible involvement of natural selection. J. Mol. Evol. 86, 312–323.

Iwasa, M. A., Kawamura, S., Myoshu, H., and Suzuki, T. A. (2018) Molecular analyses of the agouti allele in the Japanese house mouse identify a novel variant of the agouti gene. Genome 61, 195–200.

Jensen-Seaman, M. I., Furey, T. S., Puiseux, B. A., Lu, Y., Roskin, K. M., Chen, C.-F., Thomas, M. A., Haussler, D., and Jacob, H. J. (2004) Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. 14, 528–538.

Jinnai, T., Kawai, Y., Kamatani, Y., Sonoda, S., Makisumi, K., Sameshima, H., Tokunaga, K., and Saitou, N. (2021) Genome-wide SNP data of Izumo and Makurazaki populations support inner-dual structure model for origin of Yamato people. J. Hum. Genet. 66, 681–687.

Kamilar, J. M. (2009) Interspecific variation in primate counter-shading: effects of activity pattern, body mass, and phylogeny. Int. J. Primatol. 30, 877–891.

Kamilar, J. M., and Bradley, B. J. (2011) Countershading is related to positional behavior in primates. J. Zool. 283, 227–233.

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 20, 1160–1166.

Kingsley, E. P., Manseau, M., Wiley, C. D., and Hoekstra, H. E. (2009) Melanism in peromyscus is caused by independent mutation in agouti. PLoS One 4, e6435.

Kishimoto, M., Kato, M., and Suzuki, H. (2021) Morphological and molecular recharacterization of the rodent genus Mus from Nepal based on museum specimens. Mammal Study 46, 297–308.

Klungland, H., Våge, D. I., Gomez-Rayla, R., Adalbinsson, S., and Lien, S. (1995) The role of melanocyte-stimulating hormone (LISH) receptor in bovine coat color determination. Mamm. Genome 6, 636–639.

Kodama, S., Nonume, M., Morigaki, K., and Suzuki, H. (2015) Ancient onset of geographical divergence, interpopulation genetic exchange, and natural selection on the Mc1r coat color gene in the house mouse (Mus musculus). Biol. J. Linn. Soc. 114, 778–794.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.

Kuwamura, T., Nonume, M., Kinoshita, G., Abe, K., and Suzuki, H. (2017) Heterogeneous genetic make-up of Japanese house mice (Mus musculus) created by multiple independent introductions and spatio-temporally diverse hybridization processes. Biol. J. Linn. Soc. 122, 661–674.

Lai, Y.-C., Shiroishi, T., Morigaki, K., Motokawa, M., and Yu, H.-T. (2008) Variation of coat color in house mice throughout Asia. J. Zool. 274, 270–276.

Leipe, C., Long, T., Wagner, M., Goslar, R., and Tarasov, P. E. (2020) The spread of rice to Japan: Insights from Bayesian analysis of direct radiocarbon dates and population dynamics in East Asia. Quat. Sci. Rev. 244, 106507.

Li, Y., Fujisawa, K., Osada, N., Kawai, Y., Takada, T., Kryukov, A. P., Abe, K., Yonekawa, H., Shiroishi, T., Morigaki, K., et al. (2021) House mouse Mus musculus dispersal in East Eurasia inferred from 98 newly determined complete mitochondrial genome sequences. Heredity 126, 132–147.

Mallarino, R., Linden, T. A., Linnen, C. R., and Hoekstra, H. E. (2016) The role of isofoms in the evolution of cryptic coloration in Peromyscus mice. Mol. Ecol. 26, 245–258.

Marín, J. C., Rivera, R., Varas, V., Cortés, J., Agapito, A., Chero, A., Chávez, A., Johnson, W. E., and Orozco-terWengel, P. (2018) Genetic variation in coat colour genes MC1R and ASIP provides insights into domestication and management of south American cameldas. Front. Genet. 9, 457.

Marshall, J. T. (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.
Davison, M. T., and Woychik, R. P. (1994) Differential expression of a new dominant agouti allele (A\textsuperscript{cmon}) is correlated with methylation state and is influenced by parental lineage. Genes Dev. 8, 1463–1472.

Morgan, A. P., Gatti, D. M., Najarian, M. L., Keane, T. M., Galante, R. J., Pack, A. I., Mott, R., Churchill, G. A., and de Villena, F. P.-M. (2017) Structural variation shapes the landscape of recombination in mouse. Genetics 206, 603–619.

Morgan, H. D., Sutherland, H. G. E., Martin, D. I. K., and Mullen, L. M., Vignieri, S. N., Gore, J. A., and Hoekstra, H. E. (2019) Multiple conserved elements structuring inverted repeats in the mouse. Mol. Biol. Evol. 36, 222–230.

Sakuma, Y., Ranorosoa, M. C., Kinoshita, G., Shimoji, H., Tsuchiya, K., Ohdachi, S. D., Arai, S., Tanaka, C., Ramino, H., and Suzuki, H. (2016) Variation in the coat-color-controlling genes, McIr and Asip, in the house mouse Mus musculus from Madagascar. Mammal Study 41, 131–140.

Sasamori, S., Wiewel, A. S., Thomson, V. A., Kobayashi, M., Nakata, K., and Suzuki, H. (2017) Potential causative mutation for melanism in rats identified in the agouti signalling protein gene (Asip) of the Rattus rattus species complex on Okinawa Island, Japan. Zool. Sci. 34, 513–522.

Schwarz, E., and Schwarz, H. K. (1943) The wild and commensal stocks of the house mouse, Mus musculus Linnaeus. J. Mammal. 24, 59–72.

Slabová, M., and Frynta, D. (2007) Morphometric variation in nearly unstudied populations of the most studied mammal; the non-commensal house mouse (Mus musculus domesticus) in the Near East and Northern Africa. Zool. Anz. 246, 91–101.

Stephens, M., Smith, N. J., and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–989.

Suzuki, H. (2013) Evolutionary and phylogeographic views on McIr and Asip variation in mammals. Genes Genet. Syst. 88, 155–164.

Suzuki, H., Kinoshita, G., Tsunoi, T., Noju, M., and Araki, K. (2020) Mouse hair significantly lightened through replacement of the cysteine residue in the N-terminal domain of McIr using the CRISPR/Cas9 system. J. Hered. 111, 640–645.

Takeda, T., Ebata, T., Noguchi, H., Keane, T. M., Adams, D. J., Narita, T., Shin-I, T., Fujisawa, H., Toyoda, A., Abe, K., et al. (2013) The ancestor of extant Japanese fancy mice contributed to the mosaic genomes of classical inbred strains. Genome Res. 23, 1329–1338.

Thayer, A. H. (1896) The law which underlies protective coloration. Auk 13, 124–129.

Vrielings, H., Duhl, D. M., Millar, S. E., Miller, K. A., and Barsh, G. S. (1994) Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. Proc. Natl. Acad. Sci. USA 91, 5667–5671.