*Pakistan J. Zool., vol. 49(4), pp 1175-1184, 2017.* **DOI**: http://dx.doi.org/10.17582/journal.pjz/2017.49.4.1175.1184

# Distribution of Subspecies of the House Mouse, *Mus musculus* (Rodentia: Muridae) in East China as Inferred from Mitochondrial D-loop Sequences

**Tianzhu Chao, Huiqiang Cai, Yuxun Zhou, Kai Li* and Junhua Xiao**

*College of Chemistry, Chemical Engineering and Biotechnology, Donghua University, Shanghai, China*

**Article Information**
**Received 22 November 2016**
**Revised 13 December 2016**
**Accepted 30 December 2016**
**Available online 12 June 2017**

**Authors' Contribution**
**TC, KL and JX conceived and designed the study. KL, YZ and JX collected samples. TC, HC and KL extracted mouse DNA, performed mtDNA sequencing and analyzed the data. TC, YZ and KL wrote the article.**

**Key words**
**House mouse, *Mus musculus*, Phylogenetics, Mitochondrial DNA, Control region.**

## ABSTRACT

The patterns of subspecies distribution, diversity and differentiation of the house mouse *Mus musculus* provide insights into past dispersal events in natural populations of this species. We investigated the molecular phylogenetics of Chinese house mice based on 466 DNA sequences from mitochondrial D-loop fragments. Our analyses revealed that *Mus musculus musculus* and *M. m. castaneus* are older colonized subspecies with broad distributions in East China, while *M. m. domestics* is a newly arrived subspecies currently restricted to Shanghai, indicating the elevated risk in this cosmopolitan cities. Phylogenetic analysis indicates that *M. m. musculus* migrated from northern China to eastern China and that *M. m. castaneus* invaded from southern China. The two subspecies form a hybrid zone in eastern China. There were three major well-differentiated clades within *M. m. castaneus*, which were designated as CAS1, CAS2 and CAS3 and confirmed previous internal sublineage differentiation of this subspecies. Of the three clades, only CAS1 was detected in all the samples from mainland China. *M. m. musculus* exhibited no broadly distributed clade, which might be explained by frequent human-assisted movements. Our study offers a newly detailed perspective on the subspecies distribution of *M. musculus* and its history of migration to eastern China.

## INTRODUCTION

Although the house mouse *Mus musculus* has been used as an important model for mammalian speciation (Geraldes *et al*., 2008), migration (Gabriel *et al*., 2010), introgression (Staubach *et al*., 2012), laboratory biology (Montero *et al*., 2013), and medicine (Guenet and Bonhomme, 2003; Flint and Eskin, 2012), its natural population structure within China is still incompletely described. The species as a whole probably originated in South or Southwest Asia approximately half a million years ago (Geraldes *et al*., 2008) and it is usually understood to contain at least three subspecies *Mus musculus domesticus*, *M. m. musculus*, and *M. m. castaneus* on account of their ability to cross yet maintain stable hybrid zones (Teeter *et al*., 2008).

As commensals of people, the global distribution of *M. musculus* shows many features suggestive of migration associated with human migration and trade (Gabriel *et al*., 2010). Dispersal of *M. m. domesticus* into Western Europe is relatively well dated to approximately 3000 years ago due to a well-studied archaeological record and intensive genetic surveys (Cucchi *et al*., 2006). By contrast, the timing and dispersal routes of *M. m. musculus* and *M. m. castaneus* are more complicated. It is likely that *M. m. musculus* reached Eastern Europe and East Asia via an initial northern immigration followed by eastward and westward dispersals, respectively (Suzuki *et al*., 2013; Kodama *et al*., 2015). One particular *M. m. castaneus* sublineage arrived to eastern Asia via a Southeast Asia road (Suzuki *et al*., 2013).

Currently, *M. m. domesticus* is indigenous to western Europe, while *M. m. musculus* is found in northern and eastern Europe and across northern Asia to the Pacific coast in China, Korea (Boursot *et al*., 1993) and Japan (Terashima *et al*., 2006). *M. m. castaneus* is present in western, southern, southeastern, and eastern Asia, including China and Japan (Terashima *et al*., 2006). Among the three subpopulations, *M. m. castaneus* has the highest genetic diversity and a larger effective population size (Phifer-Rixey *et al*., 2012). *M. m. domesticus* and *M. m. musculus* meet in a well-studied hybrid zone ranging across Denmark, the Netherlands, Germany, Czech Republic and Bulgaria (Jones *et al*., 2010; Wang *et al*.,

* Corresponding author: likai@dhu.edu.cn
0030-9923/2017/0004-1175 $ 9.00/0

2011), and *M. m. musculus* and *M. m. castaneus* meet in a poorly studied hybrid region in mainland China and Japan (Kawashima *et al*., 1991, 1995).

Previous studies of Chinese mice reported the general distribution pattern of the two subspecies with *M. m. musculus* distributed in northern China and *M. m. castaneus* in southern China (Bao *et al*., 1995, 1999). In 2002, high levels of genetic variation as determined by microsatellite loci analysis were found in the Taiwan and Fujian populations in southeast China (Yu and Peng, 2002). Subsequently, samples from Taiwan were also found to belong to *M. m. castaneus* based on mitochondrial DNA (Geraldes *et al*., 2008). More recently, the wider geographic relationships of the East Asian populations of *M. m. musculus* and *M. m. castaneus* were established (Suzuki *et al*., 2013). An interesting finding for *M. m. castaneus* was the presence of two sublineages on Taiwan compared with only one on mainland China. In another recent study, the presence of an extensive hybrid zone between *M. m. castaneus* and *M. m. musculus* that corresponds approximately with Yangtze River was confirmed to (Jing *et al*., 2014). Therefore, a high-density population of wild mice would provide more detailed information about this hybrid zone.

In our previous study in Shanghai (eastern China) (Fan *et al*., 2008), fluorescence-based conformation sensitive gel electrophoresis (F-CSGE) and DNA sequence analyses were performed to analyze the coding regions of mitochondrial DNA. Different types of novel mutations were recorded from 64 wild house mice. We undertook further geographic sampling of wild house mice in East China with a view to generating a distribution map of their mitochondrial diversity and a better understanding of their history of immigration and contact.

## MATERIALS AND METHODS

*Sampling locations*

We trapped 191 samples of house mice from eastern and other parts of China. The sampling localities are in five main districts: southwest China (SWC), southeast China (SEC), north China (NC), central China (CC) and east China (EC) (Fig. 1, Table I). All sampling sites were at least 10 km apart. Our field sampling was carried out in accordance with the recommendations of and approved by the Laboratory Animal Committee of Donghua University.

*mtDNA sequencing*

DNA was extracted from mouse tails using the phenol-chloroform method. The mtDNA control region and tRNA sequences were amplified using 5'-ATTACTCTGGTCTTGTAAACC-3' and 5'-TATAAGGCCAGGACCAAACCT-3' primers, which were synthesized by Sangon (Shanghai Sangon Biological Engineering and Technology and Service S. Ltd.). The optimized PCR mixture contained 1×PCR Buffer, 2.5 mM $MgCl_2$, 0.25 mM each dNTP, 0.1 µM forward and reverse primers, and 0.5 U hotstarTaq polymerase (Qiagen, Germany). The PCR cycling conditions were as follows: 15 min of denaturation at 94ºC, 35 cycles of 30 s at 94ºC, 60 s at 59ºC, and 1 min at 72ºC, and a final 5 min extension step at 72ºC. For each analysis, a negative control was included with the samples. The PCR products were detected by agarose gel electrophoresis. All sequences were determined in both directions on an ABI 3730 sequencer (ABI, USA) by Sangon (Shanghai Sangon Biological Engineering and Technology and Service S. Ltd.).

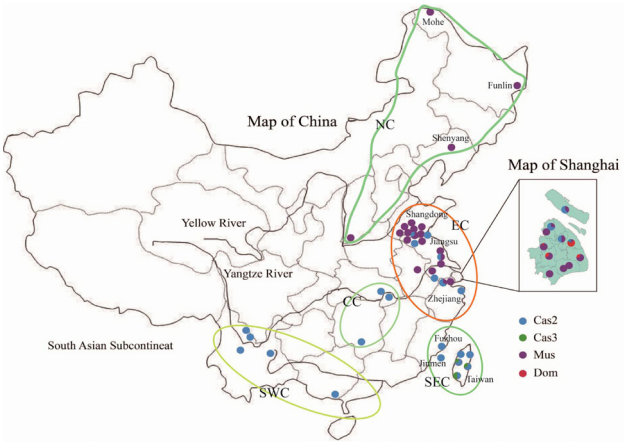


Fig. 1. Subspecies distributions of *Mus musculus* in China. Sampling locations are indicated. CAS1 and CAS2 are two main subclades of *M. m. castaneus* from Figure 2. Mus and Dom are used as abbreviations for *M. m. musculus* and *M. m. domestics*, respectively. The five main districts are southwest China (SWC), north China (NC), central China (CC), east China (EC) and southeast China (SEC).

A total of 191 new mtDNA sequences of house mouse origin are deposited in GenBank under accession numbers (KJ681118 - KJ681308). A further 275 sequences were retrieved from NCBI and *M. gentilulus* was used as the out-group. The combination of new sequences and those downloaded from GenBank produced a final data alignment of 466 sequences of 926 bp length.

*Sequence analysis*

Sequences were aligned with MAFFT 7. Fifty percent majority-rule consensus trees were generated using MrBayes 3.2 (Ronquist and Huelsenbeck, 2003), and the run conditions were 10 million generations of two independent runs of five chains.

**Table I.- Locations and numbers of samples used in this study.**

| Province | Locality | Latitude N | Longitude E | CAS1 | Mus | Dom |
|---|---|---|---|---|---|---|
| Helongjiang (NC) | Mohe | 52.25 | 121.07 | | 1 | |
| | Hulin | 45.75 | 133.97 | | 2 | |
| Liaoning (NC) | Shenyang | 41.48 | 123.23 | | 1 | |
| Henan (NC) | Sanmengxia | 34.47 | 111.12 | | 1 | |
| Shandong (EC) | Zaozhuang | 34.52 | 117.33 | 1 | 2 | |
| | Zaozhuang | 35.11 | 116.48 | | 1 | |
| | Zaozhuang | 34.7 | 117.23 | | 2 | |
| | Linyi | 35.1 | 118.01 | | 2 | |
| | Zaozhuang | 34.69 | 117.55 | | 2 | |
| | Zaozhuang | 34.77 | 117.58 | | 1 | |
| | Linyi | 34.86 | 118.1 | | 1 | |
| | Linyi | 35 | 118.37 | 1 | | |
| | Yancheng | 34.87 | 118.41 | | 1 | |
| Jiangsu (EC) | Xuzhou | 34.15 | 117.16 | | 2 | |
| | Xuzhou | 34.37 | 117.44 | 1 | | |
| | Xuzhou | 34.42 | 117.45 | | 1 | |
| | Taizhou | 32.45 | 119.92 | | 2 | |
| | Taizhou | 32.41 | 119.9 | 1 | 1 | |
| | Taizhou | 32.34 | 119.88 | 1 | | |
| | Nanjing | 31.56 | 119.1 | | 4 | |
| | Yixing | 31.07 | 119.31 | 2 | | |
| Shanghai (EC) | Qingpu | 31.1 | 121.1 | | 20 | 2 |
| | Pudong | 31.12 | 121.31 | 2 | 12 | 3 |
| | Yangpu | 31.27 | 121.52 | 1 | 1 | |
| | Songjiang | 31 | 121.24 | | 11 | 5 |
| | Chongming | 31.73 | 121.4 | 16 | 10 | |
| | Fengxian | 30.92 | 121.46 | | 4 | |
| | Fengxian | 30.97 | 121.55 | | 2 | |
| | Jinshan | 30.89 | 121.16 | | 28 | |
| | Nanhui | 31.05 | 121.76 | | 3 | 1 |
| | Jiading | 31.4 | 121.24 | 2 | 5 | |
| | Jiading | 31.23 | 121.15 | | 5 | |
| | Huangpu | 31.23 | 121.48 | 1 | | |
| Zhejiang (EC) | Jiaxing | 30.77 | 120.76 | | 1 | |
| | Huzhou | 31.02 | 119.9 | 4 | 1 | |
| | Ningbo | 30.15 | 120.18 | 1 | | |
| Anhui (EC) | Hefei | 31.51 | 117.17 | | 1 | |
| Hubei (CC) | Wuchan | 30.57 | 114.3 | 2 | | |
| | Xiaogan | 31.92 | 113.9 | 2 | | |
| Hunan (CC) | Yongzhou | 26.22 | 111.63 | 1 | | |
| Guangxi (SWC) | Daxin | 22.85 | 107.21 | 1 | | |
| Yunnan (SWC) | Lijiang | 26.83 | 100.47 | 1 | | |
| | Lijiang | 26.86 | 100.25 | 2 | | |
| | Kunming | 25.02 | 102.43 | 2 | | |
| | Dali | 25.43 | 100.1 | 4 | | |
| **Total** | | | | **49** | **131** | **11** |

CAS1 are the sub-clade of *Mus musculus castaneus* identified in the phylogenetic tree. Mus and Dom represent *M. m. musculus* and *M. m. domestics*, respectively.

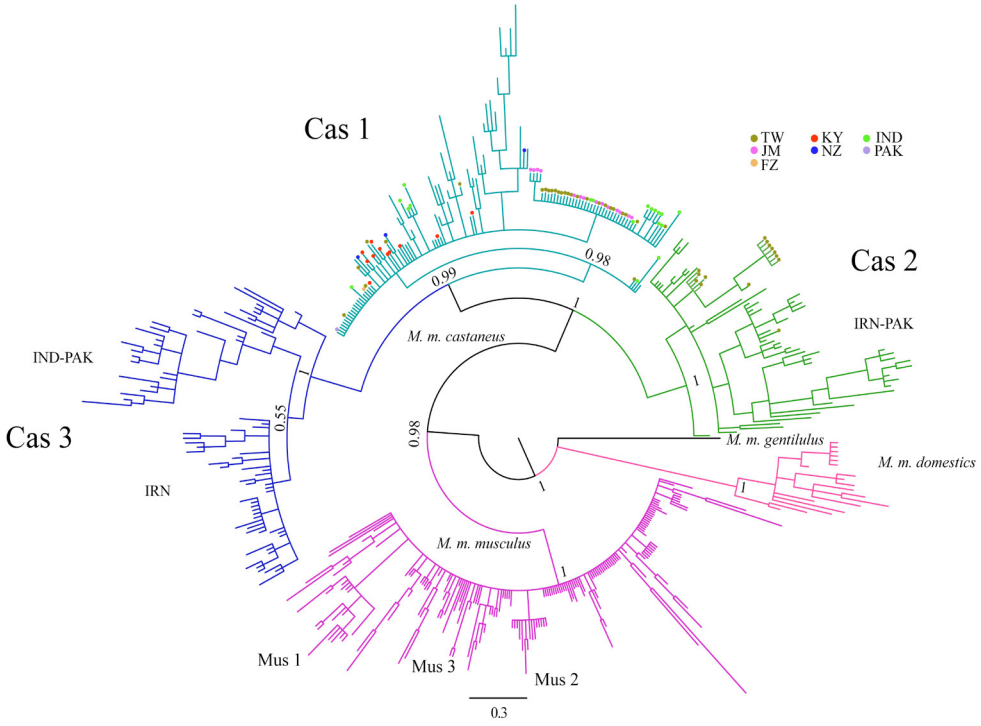


Fig. 2. MrBayes phylogeny with posterior probability values based on D-loop sequences from *Mus musculus*. The scale bar represents nucleotide substitutions per site. Clades and subclades are indicated. There are three main subclades in *M. m. castaneus*: CAS1, CAS2, and CAS3. All *M. m. castaneus* mice in mainland China are within the CAS1 clade (dark green branch), and there is no primary subclade for *M. m. musculus* (purple branch). Most samples from Taiwan (TW, the largest island in eastern China, which is close to Fuzhou, Fujian Province) are within CAS2, while samples from Fuzhou (FZ) and Jinmen (JM, Fujian Province) are in CAS1. KY, NZ, IND, and PAK are used as abbreviations for Kenya, New Zealand, India, and Pakistan, respectively.

The burn-in was determined from the convergence of the two runs using the potential scale reduction factor (PSRF) statistics. The analysis was performed using gamma-distributed rates across sites and variable transition and transversion rates as determined by JMODELTEST 0.1.1 (Posada, 2008). Polymorphic and divergence parameters was calculated with DnaSP version 5 (Librado and Rozas, 2009). Tajima's D and Fu and L's D neutrality tests, average number of nucleotide substitutions per site between populations (Dxy), and the net number of nucleotide substitutions per site between populations (Da) were also calculated with DnaSP. $F_{ST}$ between subspecies and between populations of a given subspecies were calculated using Arlequin 3.5 (Excoffier and Lischer, 2010). Neighbor-Net networks of *M. m. musculus* haplotypes and *M. m. castaneus* haplotypes of the CAS1 clade were generated with SplitsTree 4.14 software (Huson and Bryant, 2006).

## RESULTS

*Subspecies attribution of mtDNA haplotypes*

A Bayesian phylogenetic tree constructed from the D-loop alignment shows a clear division into three main lineages, with ample placement consistent with subspecies annotations in NCBI (Fig. 2). The newly sequenced samples included 49 with *M. m. castaneus* haplotypes, 131 with *M. m. musculus* haplotypes, and 11 with *M. m. domesticus* haplotypes (Table I).

*mtDNA lineage distributions*

Two main subspecies, *Mus m. musculus* and *M. m. castaneus*, coexist in EC, while *M. m. musculus* or *M. m. castaneus* alone were found in SWC, SEC, NC, and CC. All individuals from the four sites in NC to the north of the Yellow River were found to possess *M. m. musculus* mtDNA haplotype. By contrast, mice from CC, SWC and SEC all possessed the *M. m. castaneus* haplotypes. *M. m. musculus* and *M. m. castaneus* haplotypes co-occur over a large area in EC from Shandong Province south to Zhejiang Province (Fig. 1). All individuals with *M. m. domesticus* haplotypes were collected in Shanghai where they co-occur with *M. m. musculus* and *M. m. castaneus* haplotypes (Fig. 1, Table I).

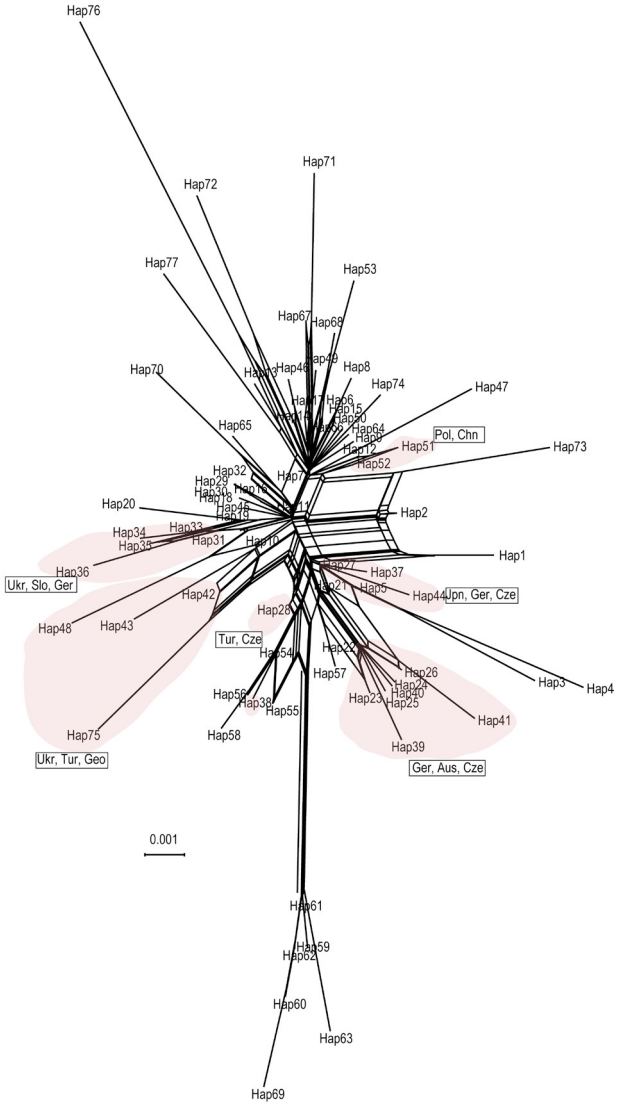


Fig. 3. NeighborNet tree of *M. m. musculus* haplotypes based on the mitochondrial DNA control region. Pol, Chn, Ukr, Slo, Ger, Geo, Tur, and Cze are used as abbreviations for Portugal, China, Ukraine, Slovenia, Germany, Georgia, Turkey, and Czech Republic, respectively. Haplotypes from the above regions are shaded. The remaining, unmarked haplotypes were all from China.

*mtDNA phylogeography*

There were 76 haplotypes detected within *M. m. musculus* (Fig. 3); however, no coincident geographic pattern was found. The *M. m. musculus* haplotypes cluster into a few small geographically based sub-clades; these are labeled on Figure 2 as Mus1, Mus2, and Mus3 with posterior probabilities of 1, 0.97, and 0.97, respectively. A series of 16 individuals from EC were grouped together within Mus1. The other two sub-clades (Mus2 and Mus3) are not represented in China but are reported from mice from Europe, central Asian, and Oceania. Mus2 consisted of the individuals from Germany, Czech, Austria, and New Zealand. Mus3 was found in two sequences from Georgia in central Europe.

The *M. m. castaneus* lineages encompassed three main clades, which are named CAS1, CAS2 and CAS3; each clade is well supported, with posterior probabilities of 0.99, 1 and 1, respectively. These three clades are almost equivalent to the haplogroups CAS-1, CAS-2, and CAS-3 that were described by Suzuki's group (Suzuki *et al*., 2013) or HG3, HG1, and HG2 described by Bonhomme's group (Rajabi-Maham *et al*., 2012). All samples from mainland China fall into CAS1.

CAS1 is the most widespread clade within *M. m. castaneus*. It is represented on at least three continents: Asia, Africa and Oceania, and there were 61 mtDNA haplotypes detected (Fig. 4). These haplotypes formed four sub-clades, CAS1A, CAS1B, CAS1C, and CAS1D. Indian samples were found in CAS1A, CAS1B, and CAS1C, and CAS1A contains samples only from India.

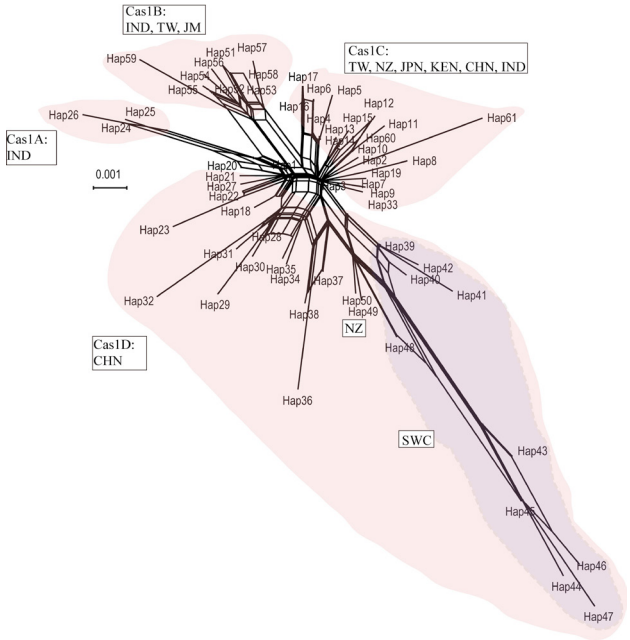


Fig. 4. NeighborNet tree of clade CAS1 haplotypes of *M. m. castaneus* based on the mitochondrial DNA control regions. Tip labels from the four major subgroups were defined as CAS1A, CAS1B, CAS1C, and CAS1D and are indicated by light shading. A geographical group for SWC (southwest China) exists within CAS1D. KY, NZ, IND, JPN, CHN, TW, JM and SWC are used as abbreviations for Kenya, New Zealand, India, Japan, mainland China, Taiwan Island (China), and Jinmen Island (China), respectively.

**Table II.- Polymorphisms within subspecies and the divergence between these subspecies and *Mus gentilulus*.**

| Populations | N | Polymorphic nucleotide sites | Haplotypes | Hd | л(%) | θ(%) | Tajima's D | Fu's Fs | Dxy | Da(%) |
|---|---|---|---|---|---|---|---|---|---|---|
| musChina | 131 | 66 | 52 | 0.9 | 0.614 | 1.717 | -2.03322 (p<0.05) | -3.58578 (p<0.02) | 2.906 | 2.599 |
| musWorld | 176 | 83 | 77 | 0.94 | 0.626 | 2.079 | -2.1805 (p<0.01) | -5.02493 (p<0.02) | 2.875 | 2.562 |
| casChina | 104 | 63 | 42 | 0.931 | 0.808 | 1.303 | -1.205 (p>0.1) | -1.51694 (p>0.05) | 2.919 | 2.515 |
| casMainland China | 50 | 36 | 30 | 0.904 | 0.589 | 1.119 | -1.59324 (p>0.1) | -2.17622 (p>0.05) | 3.23 | 2.936 |
| CAS1 | 131 | 52 | 61 | 0.936 | 0.516 | 1.368 | -1.94250 (p<0.05) | -2.93925 (p<0.05) | 2.904 | 2.646 |
| CAS2 | 64 | 50 | 42 | 0.976 | 0.813 | 1.437 | -1.45889 (p>0.1) | -1.85853 (p>0.1) | 2.915 | 2.508 |
| CAS3 | 75 | 55 | 59 | 0.99 | 0.906 | 1.52 | -1.33842 (p>0.1) | -3.29877 (p<0.05) | 3.14 | 2.687 |
| casWorld | 270 | 109 | 161 | 0.983 | 1.252 | 2.7 | -1.652740 (p>0.05) | -4.87136 (P<0.02) | 2.976 | 2.356 |

musChina and musworld, represent the sample sequences of *Mus musculus musculus* from China and elsewhere in the world, respectively; casChina, casMainlandChina, and casWorld represent the sequences of *M. m. castaneus* from China, mainland China, and elsewhere in the world, respectively. CAS1, CAS2, and CAS3 indicate the three main clades of *M. m. castaneus* shown in Figure 2.

CAS1B consisted of samples from India, Taiwan Island, and Jinmen Island. CAS1C was composed of mice in Africa from Kenya and in Oceania from New Zealand (Figs. 2, 4) together with those from Taiwan Island (China), Japan, Jinmen Island (China), and India (Fig. 4). Nearly all samples of *M. m. castaneus* from mainland China fall into CAS1D. Samples from the SEC region form a relatively independent subgroup, while nearly all the remaining samples of CAS1D came from the EC region.

CAS3 consists exclusively of haplotypes found in mice from Indian, Pakistan and Iran. CAS2 is best represented in mice from Iran and Pakistan but a few samples from Taiwan Island belong to this clade (Fig. 2). Suzuki's group (Suzuki *et al*., 2013) showed that house mice of this clade were also observed in Vietnam.

Within the *M. m. castaneus* subspecies worldwide, the CAS3 sub-clade is the most variable, followed by CAS2 and CAS1 (Table II). Although Chinese populations are restricted to only one of these clades (CAS3), these show more nucleotide diversity ($\pi$=0.808±0.082%) than Chinese *M. m. musculus* ($\pi$=0.614±0.054%) (Table II).

Tajima's D and Fu's Fs values for various subspecies and populations are shown in Table II. For both worldwide and Chinese samples *M. m. musculus* there are significantly negative for Tajima's D-values and for Fu's Fs. This indicates both global and regional population increases for this subspecies, either following period of positive selection or possibly the occurrence of bottlenecks. By contrast, Chinese populations of *M. m. castaneus* do not show significant negative values, either taken as a whole or on for mainland China only.

Pattern of genetic differentiation are compared between and within subspecies in Table III. In general, $F_{ST}$ values are higher for comparisons between subspecies than within subspecies. There was more differentiation between *M. m. musculus* and *M. m. castaneus* in China than between the two subspecies in the other locations worldwide. The highest levels of differentiation within subspecies were between the CAS1 and CAS3 sub-clades of *M. m. castaneus*, followed by CAS2 and CAS3, and CAS1 and CAS2.

**Table III.- Patterns of differentiation between different population or subspecies pairs.**

| | cas-mus[a] | CAS1-CAS2 | CAS3-CAS1 | CAS3-CAS2 | cas-mus (China)[b] |
|---|---|---|---|---|---|
| poly-morphism | 147 | 85 | 91 | 80 | 92 |
| Fixed differences | 0 | 0 | 2 | 1 | 4 |
| Shared mutation | 33 | 16 | 11 | 25 | 15 |
| Fst | 0.551 | 0.492 | 0.515 | 0.392 | 0.654 |
| Dxy (%) | 2.2 | 1.61 | 1.148 | 1.641 | 2.185 |
| Da (%) | 1.303 | 0.929 | 0.754 | 0.78 | 1.535 |

CAS1, CAS2, and CAS3 indicate the three main clades of *M. m. castaneus* shown in Figure 2. [a]cas-mus indicates all sequences from China and other countries. [b]cas-mus (China) indicates all sequences of *M. m. castaneus* and *M. m. musculus* from China.

*Diversity, differentiation, and divergence time*

As reported previously by Suzuki group (Suzuki *et al*., 2013), there is considerable variations in the level of polymorphisms among the various subspecies and populations of *M. musculus*. With our new samples

included, *M. m. castaneus* contains nearly twice the nucleotide diversity ($\pi$=1.252±0.34%) found in *M. m. musculus* ($\pi$=0.626±0.041%). In *M. m. castaneus*, more than 100 polymorphic sites contribute to 161 haplotypes ($H_d$=0.983±0.0034). By comparison, *M. m. musculus* has only 83 variable sites contributing to 77 haplotypes ($H_d$=0.940±0.013).

Within the *M. m. castaneus* subspecies worldwide, the CAS3 sub-clade is the most variable, followed by CAS2 and CAS1 (Table II). Although Chinese populations are restricted to only one of these clades (CAS3), these show more nucleotide diversity ($\pi$=0.808±0.082%) than Chinese *M. m. musculus* ($\pi$=0.614±0.054%) (Table II).

Tajima's D and Fu's Fs values for various subspecies and populations are shown in Table II. For both worldwide and Chinese samples *M. m. musculus* there are significantly negative for Tajima's D-values and for Fu's Fs. This indicates both global and regional population increases for this subspecies, either following period of positive selection or possibly the occurrence of bottlenecks. By contrast, Chinese populations of *M. m. castaneus* do not show significant negative values, either taken as a whole or on for mainland China only.

Pattern of genetic differentiation are compared between and within subspecies in Table III. In general, $F_{ST}$ values are higher for comparisons between subspecies than within subspecies. There was more differentiation between *M. m. musculus* and *M. m. castaneus* in China than between the two subspecies in the other locations worldwide. The highest levels of differentiation within subspecies were between the CAS1 and CAS3 sub-clades of *M. m. castaneus*, followed by CAS2 and CAS3, and CAS1 and CAS2.

## DISCUSSION

*Molecular markers for subspecies identification*

Our phylogenetic tree suggests that the control region of mitochondrial DNA is suitable for subspecies or taxa of lower rank (*i.e.*, population) for the availability of sequences in public database, although complete mitochondrial DNA is more desirable for distinguishing house mouse subspecies (Katouzian and Rajabi-Maham, 2013). The mutation rate per generation per site in the control region of *M. musculus* is approximately 0.41 million (Suzuki *et al*., 2004; Geraldes *et al*., 2008), which indicates a significant number of polymorphic sites suitable for identification. For instance, 147 variable sites were found between *M. m. musculus* and *M. m. castaneus*, although a fixed SNP has not been identified (Table III). All of our samples and downloaded sequences were precisely classified at the subspecies level. In terms of the descriptions of clades within individual subspecies, such as CAS1, CAS2, and CAS3, our results were remarkably consistent with those of previous studies (Rajabi-Maham *et al*., 2012; Suzuki *et al*., 2013).

Nuclear genes in autosomes are useful for studying mixed populations of wild mice from hybrid zones and could reveal additional details about gene flow (Yu and Peng, 2002). However, this method has produced complex and confusing results within *M. musculus*, as phylogenetic trees based on one gene may not be in agreement with trees based on other genes (Geraldes *et al*., 2008), even if those genes are nearby in the chromosome (Nunome *et al*., 2010). Therefore, the use of nuclear genes is important. Fine-scale phylogenetic analysis throughout the whole mouse genome might solve this problem and provide a large amount of phylogenetic history at the genomic scale of three closely related subspecies: *M. m. musculus*, *M. m. castaneus*, and *M. m. domesticus* (White *et al*., 2009).

*Biogeography and evolution*

Our study of mtDNA of Chinese house mice has shed new light on the regional mtDNA diversity and distributions of each of *M. m. musculus*, *M. m. castaneus*, and *M. m. domesticus*. For each subspecies, mtDNA diversity within China represents a restricted subset of that observed globally. This is consistent with current belief that each of the subspecies originated within a relatively small geographic area in southwest Asia, followed by dramatic episodes of range expansion, much of it achieved through human commensalism (Geraldes *et al*., 2008; Suzuki *et al*., 2013).

Although it seems reasonable to postulate that colonization of China by house mice occurred largely through their commensalism with humans, little is known regarding how and when this occurred for each of the subspecies. Though, there is still debate about precise timing of modern human settlement in mainland China, one point of general agreement is that Han populations were first established in South China, followed by a northward expansion (Shi *et al*., 2008). On these grounds, we might expect that the earliest introduction of house mice was into South China, and that these mice belonged to the *M. m. castaneus* lineage. A similar conclusion was reached in the previous mtDNA RFLP study (Yonekawa *et al*., 1988).

All individuals of *M. m. castaneus* from mainland China belong to CAS1 clade (Fig. 2). Within this clade, all individuals from mainland China are in CAS1D subgroups, except one individual in CAS1C (Fig. 4). Based on this evidence, we hypothesize that CAS1D house mice were the main *M. m. castaneus* mice introduced into China and dispersed to other part of this country. Moreover,

samples from Southwest Asia (Iran), South Asia (India and Pakistan), Southeast Asia (Thailand), East Asia (Japan, Taiwan), Africa (Kenya), and Oceania (New Zealand) were included within CAS1. A previous study also confirmed that southwestern Asia is the ancestral homeland of *M. m. castaneus* (Suzuki *et al*., 2013). Because this high regional diversity was not observed for CAS2 and CAS3, our data support the hypothesis that CAS1 *M. m. castaneus* is the primary clade that has expanded within Asia, and to Africa and Australia.

Within *M. m. musculus*, there is less differentiation and no strong phylogeographic patterning. This contrast with *M. m. castaneus* may reflect the contrasting topographies and histories of the main area of geographic distribution of the two subspecies. In the south of Eurasian continent, the area occupied by *M. m. castaneus*, mountain ranges and large river basins create a naturally fragmented landscape and this has been reflected in ancient division of civilization into multiple regional centers. In contrast, northern Eurasia is a landscape dominated by vast plateau and human history has been more strongly influenced by large scale migration at a population scale. At the global scale, *M. m. castaneus* harbors more mtDNA variation than *M. m. musculus* in this study and previous studies (Baines and Harr, 2007; Geraldes *et al*., 2008), both in terms of variable sites and haplotype number (Table II). In contrast, nucleotide diversity of mainland Chinese populations of *M. m. castaneus* is lee than that of Chinese *M. m. musculus*. We believe that this is explained by their contrasting opportunities for invasion and ongoing interchange with populations of the corresponding subspecies outside of China. Essentially, Chinese populations of *M. m. musculus* are more likely part of a larger regional population of this subspecies that stretches west into central Eurasia and beyond, whereas Chinese populations of *M. m. castaneus*, seem more isolated from other populations of this subspecies. In particular the last 1500 years has seen the repeated expansion of Proto-Mongal and Mongol peoples creating a cultural and biological realm stretching at times from the Caspian Sea to the Pacific coastline. The occurrence of *M. m. domesticus* in China is a more restricted and modern and modern phenomenon. To date, this classic marker of western European interaction is recorded exclusively from Shanghai, China's most cosmopolitan city with the most extensive network of global connections.

Human migration and geographic isolation have also determined the local populations of genetic diversity in each of *M. m. musculus* and *M. m. castaneus*. Jinmen Island is 4.3 km off the coast of Fujian Province of China (Fig. 1) and is the site of Jinmen City constructed in 1385. Today it is administered by Taiwan. Although we might expect to find commonality in *M. m. castaneus* sequences from Jinmen Island with the population in Fuzhou (capital city of Fujian Province) belonging to CAS1D, instead the Jinmen samples share CAS1B and CAS2C haplotypes with individuals from Taiwan (Figs. 2, 4). Thus the genetic pattern appears to mirror the political and human isolation between Jinmen (administered by Taiwan) and mainland China over the last 60years, since The Third Revolutionary Civil War in modern China. Similar instances of house mouse populations mirroring recent political or human immigration events are known from other parts of the world. For example, a single *M. m. musculus* sequence from New Zealand is identical to records of central Europe, consistent with a significant German immigration into New Zealand (Searle *et al*., 2009); and Japanese *M. m. musculus* sequences are very similar to those from Korean mice, again reflecting the recent cultural exchange between these regions (Nunome *et al*., 2010).

Our expanded geographic sampling of Chinese house mice has more precisely defined the distributions in China of *M. m. musculus* and *M. m. castaneus*. Importantly, our sampling has demonstrated a relatively large area of geographic overlap between the two mtDNA lineages in eastern China, including parts of Shangdong, Jiangsu and Zhejiang Provinces. No population of *M. m. castaneus* occurs north of the Yellow River. However, in the south several populations of *M. m. musculus* occur south of the Yangtze River. While these long and wide rivers are likely to pose significant barriers to the natural dispersal of a small terrestrial mammal such as *M. musculus*, a commensal mammal has the advantage of human-assisted dispersal across such barriers.

The interdigitating geographic distribution of the *M. m. musculus* and *M. m. castaneus* mtDNA types raises numerous questions regarding the biological interaction between the subspecies. Do the two subspecies overlap in range with limited exchange of genetic material? Alternatively, is the mosaic of mtDNA types indicative of a hybrid zone? If so, what is the nature of the genetic interaction that is occurring? Is it limited to introgression of mtDNA and/or other restricted genetic component? Or is it characterized by broad-scale exchange of genetic component? Further studies of the *musculus/castaneus* interaction in East China would represent a fascinating counterpoint to the detailed analyses available for the *Musculus/domesticus* hybrid zone in Europe (Teeter *et al*., 2008).

## ACKNOWLEDGMENTS

We thank Shimin Bao of Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences for his

kind help in sample collection. This work was supported by the Key Project of Science and Technology Commission of Shanghai Municipality (No. 13140900300), the National Science Foundation of China (No. 31171199), the Fundamental Research Funds for the Central Universities (No. 2232013A3-06), and the DHU Distinguished Young Professor Program (B201308).

*Statement of conflict of interest*

Authors have declared no conflict of interest.

## REFERENCE

Bao, S., Jin, M. and Shi, M., 1995. *Taxonomical characteristics of geneus* Mus *in Hainan, China*. China Forestry Publishing House, China, pp. 203-210 (in Chinese).

Bao, S., Zhao, G., Zhang, R., Gu, Z. and Xu, P., 1999. *Study on genetic differentiation of* Mus musculus *of China.* Zoology Society of China. Forestry Publishing House, China, pp. 987-994 (in Chinese).

Baines, J.F. and Harr, B., 2007. Reduced X-linked diversity in derived populations of house mice. *Genetics,* **175**: 1911-1921. https://doi.org/10.1534/genetics.106.069419

Boursot, P., Auffray, J.C., Brittondavidian, J. and Bonhomme, F., 1993. The evolution of house mice. *Annu. Rev. Ecol. System*., **24**: 119-152. https://doi.org/10.1146/annurev.es.24.110193.001003

Cucchi, T., Orth, A., Auffray, J.C., Renaud, S., Fabre, L., Catalan, J., Hadjisterkotis, E., Bonhomme, F. and Vigne, J.D., 2006. A new endemic species of the subgenus *Mus* (Rodentia, Mammalia) on the Island of Cyprus. *Zootaxa*, **1241**: 1-36.

Excoffier, L. and Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour*., **10**: 564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fan, Z.P., Zhu, W.S., Zhang, C., Zhou, Y.X., Li, K., Liang, Y.M., Xing, Z.H., Chen, G.Q., Bai, X. and Xiao, J.H., 2008. SNP discovery by F-CSGE in the coding region of mitochondrial DNA in wild house mice from Shanghai suburb. *Yi Chuan,* **30**: 475-482 (in Chinese). https://doi.org/10.3724/SP.J.1005.2008.00475

Flint, J. and Eskin, E., 2012. Genome-wide association studies in mice. *Nature Rev. Genet.,* **13**: 807-817. https://doi.org/10.1038/nrg3335

Gabriel, S.I., Johannesdottir, F., Jones, E.P. and Searle, J.B., 2010. Colonization, mouse-style. *BMC Biol.*, **8**: 131. https://doi.org/10.1186/1741-7007-8-131

Geraldes, A., Basset, P., Gibson, B., Smith, K.L., Harr, B., Yu, H.T., Bulatova, N., Ziv, Y. and Nachman, M.W., 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Mol. Ecol*., **17**: 5349-5363. https://doi.org/10.1111/j.1365-294X.2008.04005.x

Guenet, J.L. and Bonhomme, F., 2003. Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet.,* **19**: 24-31. https://doi.org/10.1016/S0168-9525(02)00007-0

Huson, D.H. and Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol., Evolut*., **23**: 254-267.

Jing, M., Yu, H.T., Bi, X., Lai, Y.C., Jiang, W. and Huang, L., 2014. Phylogeography of Chinese house mice (*Mus musculus musculus/castaneus*): distribution, routes of colonization and geographic regions of hybridization. *Mol. Ecol*., **23**: 4387-4405. https://doi.org/10.1111/mec.12873

Jones, E.P., Van Der Kooij, J., Solheim, R. and Searle, J.B., 2010. Norwegian house mice (*Mus musculus musculus/domesticus*): distributions, routes of colonization and patterns of hybridization. *Mol. Ecol*., **19**: 5252-5264. https://doi.org/10.1111/j.1365-294X.2010.04874.x

Katouzian, A.R. and Rajabi-Maham, H., 2013. Evaluation of effectiveness of some mitochondrial genes in biosystematics and phylogeographic studies of house mouse (*Mus musculus*) subspecies. *Progr. biol. Sci.,* **3**: 39-66.

Kawashima, T., Miyashita, N., Wang, C.H., He, X.Q., Jin, M.L., Wu, Z.G.. and Moriwaki, K., 1991. A new haplotype of the beta-globin gene-complex, Hbbw1, in Chinese wild mouse. *Jap. J. Genet*., **66**: 491-500. https://doi.org/10.1266/jjg.66.491

Kawashima, T., Miyashita, N., Tsuchiya, K., Li, H., Wang, F.S., Wang, C.H., Wu, X.L., Wang, C.Y., Jin, M.L., He, X.Q., Kryukov, A.P., Yakimenko, L.V., Frisman, L.V. and Moriwaki, K., 1995. Geographical-distribution of the Hbb Haplotypes in the *Mus Musculus* subspecies in Eastern Asia. *Jap. J. Genet*., **70**: 17-23. https://doi.org/10.1266/jjg.70.17

Kodama, S., Nunome, M., Moriwaki, K. and Suzuki, H., 2015. Ancient onset of geographical divergence, interpopulation genetic exchange, and natural selection on the Mc1r coat-colour gene in the house mouse (*Mus musculus*). *Biol. J. Linnean Soc*., **114**: 778-794. https://doi.org/10.1111/bij.12471

Librado, P. and Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**: 1451-1452. https://doi.

org/10.1093/bioinformatics/btp187

Montero, I., Teschke, M. and Tautz, D., 2013. Paternal imprinting of mating preferences between natural populations of house mice (*Mus musculus domesticus*). *Mol. Ecol.,* **22**: 2549-2562. https://doi.org/10.1111/mec.12271

Nunome, M., Ishimori, C., Aplin, K.P., Tsuchiya, K., Yonekawa, H., Moriwaki, K. and Suzuki, H., 2010. Detection of recombinant haplotypes in wild mice (*Mus musculus*) provides new insights into the origin of Japanese mice. *Mol. Ecol.,* **19**: 2474-2489. https://doi.org/10.1111/j.1365-294x.2010.04651.x

Phifer-Rixey, M., Bonhomme, F., Boursot, P., Churchill, G.A., Pialek, J., Tucker, P.K. and Nachman, M.W., 2012. Adaptive evolution and effective population size in wild house mice. *Mol. Biol. Evolut*., **29**: 2949-2955. https://doi.org/10.1093/molbev/mss105

Posada, D., 2008. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evolut*., **25**: 1253-1256. https://doi.org/10.1093/molbev/msn083

Rajabi-Maham, H., Orth, A., Siahsarvie, R., Boursot, P., Darvish, J. and Bonhomme, F., 2012. The south-eastern house mouse *Mus musculus castaneus* (Rodentia: Muridae) is a polytypic subspecies. *Biol. J. Linean Soc.*, **107**: 295-306. https://doi.org/10.1111/j.1095-8312.2012.01957.x

Ronquist, F. and Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**: 1572-1574. https://doi.org/10.1093/bioinformatics/btg180

Searle, J.B., Jamieson, P.M., Gunduz, I., Stevens, M.I., Jones, E.P., Gemmill, C.E.C. and King, C.M., 2009. The diverse origins of New Zealand house mice. *Proc. R. Soc. B-Biol. Sci*., **276**: 209-217. https://doi.org/10.1098/rspb.2008.0959

Shi, H., Zhong, H., Peng, Y., Dong, Y.L., Qi, X.B., Zhang, F., Liu, L.F., Tan, S.J., Ma, R.Z., Xiao, C.J., Wells, R.S., Jin L. and Su B., 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol*., **6**: 45. https://doi.org/10.1186/1741-7007-6-45

Staubach, F., Lorenc, A., Messer, P.W., Tang, K., Petrov, D.A. and Tautz, D., 2012. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *PLoS Genet.,* **8**: e1002891. https://doi.org/10.1371/journal.pgen.1002891

Suzuki, H., Shimada, T., Terashima, M., Tsuchiya, K. and Aplin, K., 2004. Temporal, spatial, and ecological modes of evolution of *Eurasian mus* based on mitochondrial and nuclear gene sequences. *Mol. Phylogen. Evolut*., **33**: 626-646. https://doi.org/10.1016/j.ympev.2004.08.003

Suzuki, H., Nunome, M., Kinoshita, G., Aplin, K.P., Vogel, P., Kryukov, A.P., Jin, M.L., Han, S.H., Maryanto, I., Tsuchiya, K., Ikeda, H., Shiroishi, T., Yonekawa, H. and Moriwaki, K., 2013. Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity,* **111**: 375-390. https://doi.org/10.1038/hdy.2013.60

Teeter, K.C., Payseur, B.A., Harris, L.W., Bakewell, M.A., Thibodeau, L.M., O'Brien, J.E., Krenz, J.G., Sans-Fuentes, M.A., Nachman, M.W. and Tucker, P.K., 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res.,* **18**: 67-76. https://doi.org/10.1101/gr.6757907

Terashima, M., Furusawa, S., Hanzawa, N., Tsuchiya, K., Suyanto, A., Moriwaki, K., Yonekawa, H. and Suzuki, H., 2006. Phylogeographic origin of Hokkaido house mice (*Mus musculus*) as indicated by genetic markers with maternal, paternal and biparental inheritance. *Heredity*, **96**: 128-138. https://doi.org/10.1038/sj.hdy.6800761

Wang, L.Y., Luzynski, K., Pool, J.E., Janousek, V., Dufkova, P., Vyskocilova, M.M., Teeter, K.C., Nachman, M.W., Munclinger, P., Macholan, M., Pialek, J. and Tucker, P.K., 2011. Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. *Mol. Ecol*., **20**: 2985-3000. https://doi.org/10.1111/j.1365-294X.2011.05148.x

White, M.A., Ane, C., Dewey, C.N., Larget, B.R. and Payseur, B.A., 2009. Fine-scale phylogenetic discordance across the house mouse genome. *PLoS Genet.*, **5**: e1000729. https://doi.org/10.1371/journal.pgen.1000729

Yonekawa, H., Moriwaki, K., Gotoh, O., Miyashita, N., Matsushima, Y., Shi, L.M., Cho, W.S., Zhen, X.L. and Tagashira, Y., 1988. Hybrid origin of Japanese mice *Mus musculus molossinus* - Evidence from restriction analysis of mitochondrial-DNA. *Mol. Biol. Evolut*., **5**: 63-78.

Yu, H.T. and Peng, Y.H., 2002. Population differentiation and gene flow revealed by microsatellite DNA markers in the house mouse (*Mus musculus castaneus*) in Taiwan. *Zool. Sci*., **19**: 475-483. https://doi.org/10.2108/zsj.19.475