A phylogeographic survey of the house mouse in Iran, taxonomic and karyotypic inference from MtDNA evidence

Farahnaz Molavi¹ *, Jamshid Darvish²,³, Farhang Haddad³, Maryam M. Matin³,⁴

¹Department of Biology, Faculty of Science, Islamic Azad University of Mashhad, Mashhad, Iran, ²Rodentology Research Group, Institute of Applied Zoology, Ferdowsi University of Mashhad, Mashhad, Iran, ³Department of Biology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran, ⁴Novel Diagnostics and Therapeutics Research Group, Institute of Biotechnology, Ferdowsi University of Mashhad, Mashhad, Iran

*Email: Fm_yazdan@yahoo.com

Received: 12 January 2020 / Revised: 31 March 2020 / Accepted: 26 March 2020 / Published online: 20 April 2020. Ministry of Sciences, Research, and Technology, Arak University, Iran.

Abstract
Recent studies have shown that the house mouse (Mus musculus) has four subspecies in Iran. Although these four subspecies have been recognized, the house mouse of east Iran showed high heterozygosity in various markers like allozymes, nuclear, and mitochondrial gene sequences. Moreover, the taxonomy and diagnostic characters of mice populations in Iran and adjacent regions are poorly understood. To define evident characteristics for the subspecies described and identify the borders of Iranian subspecies, thirty-one populations were studied using three methods: chromosomal morphology, morphology, and molecular analysis of the mitochondrial Cytochrome b gene. Molecular analysis of the M. musculus samples revealed four clades: 1- clade M. m. isatissus (of Iran) and M. m. castaneus (of India), 2- clade M. m. bactrianus from eastern areas with high intrasubspecies genetic distance, 3- clade M. m. domesticus in the southern and western regions and 4- M. m. musculus in the northeastern region of Iran. Morphometric characters resulted in three groups that overlapped with each other. The morphological characters could not separate M. m. isatissus, and M. m. bactrianus, from each other. Analysis of cytogenetic variables showed four clear groups better than molecular clades. In these methods, the central and eastern clades are two distinct groups that are well supported by the difference in the size of centromeric heterochromatin and their patterns. These results showed that cytogenetic studies are useful and easy methods for identifying the diagnostic characters of Iranian subspecies.

Keywords: Banding, chromosomal variation, cytogenetic, mice

Introduction
So far, nine species have been recognized in the genus Mus. This taxon arose within the last 4 Myr (Bonhomme and Guénet 1996). Mus musculus was originally a Palearctic species, but it has now been spread throughout the world by humans and lives as a human commensal (Musser and Carleton 2005). Genetic studies have revealed three peripheral geographic populations of house mouse as Mus musculus musculus, M. m. domesticus and M. m. castaneus (Vanlerbergh & al. 1986, Orth et al. 1996, Darvish et al. 2006, Rajabi-Maham et al. 2012). M. m. isatissus was recognized in 2015 by Haddadian et al., in central areas of Iran. Numerous studies have indicated that subspecies of Mus musculus do not have complete reproductive isolation, and in regions of secondary contact, there is evidence of genetic exchanges from limited introgression (Guénet 2003). Preliminary molecular data have shown that the house mouse subspecies are highly divergent with geographic variability and the lack of unambiguous diagnostic characters (Phifer-Rixey et al., 2012). Chromosomal characteristics serve as useful
taxonomic markers in the genus *Mus* (Veyrunes et al. 2004). In recent years, cytogenetics has focused on chromosome architecture rather than the number, structure, and abnormalities in both plants and animals (Dhananjoy et al. 2014). The house mouse populations in the center of the original range (somewhere between the north of the Indian subcontinent and the adjacent regions of Iran and Afghanistan) (Din et al. 1996) have more diversity than other populations, and this diversity has remained mostly unknown. In the present study, we used the morphometric characters, chromosomal bands, and mtDNA variations of the subspecies of house mouse from Iran, aiming to appraise the patterns of phenotypic differentiation among the subspecies. Furthermore, we explored the patterns of phenotypic variations among different populations in *M. musculus* and evaluated centromeric band properties, molecular variations, and morphometric traits through the use of univariate and multivariate analyses in house mice.

**Material and methods**

**Sampling and DNA extraction**

A total of 167 mice belonging to the species *Mus musculus* were caught with Longworth live-traps in 31 localities from Iran (Fig. 1, Appendix Table 1). Samples were collected during the years 2012–2015, under the supervision of the Rodentology Research Group at Ferdowsi University of Mashhad. All 167 samples of mice were included in the morphometric analysis; 61 specimens were used in molecular analysis, and 133 specimens were used in the cytogenetic analysis.

**Molecular analyses**

**DNA amplification and sequencing**

DNA samples were extracted from fresh or ethanol-preserved tissues using the standard salt extraction method (Bruford et al. 1992). Complete Cytochrome b (1111 bp) sequences were amplified using modified universal primers; L7 and H6 for Cytb following Montgelard et al. (2002) protocol. PCR amplifications were performed according to Chevret et al., 2005. To explore the phylogenetic relationships through house mice representatives, *Mus spicilegus* was used as out-groups.

Figure 1. Collecting localities for *M. musculus* analyzed in this study (black boxes indicate the stations).

**Phylogenetic Analysis**

Phylogenetic analysis was performed using the Cytb sequences of the wild-caught specimens; the complete data set comprise 61 sequences. The alignment was performed with Clustal W (Thompson et al. 1997) algorithm using BioEdit 7.0.5 (Hall 1999). Haplotype (*h*) and nucleotide (*π*) diversity through populations were estimated using DnaSP version 5 (Librado et al. 2009). The MEGA 5.0 software package (Tamura et al. 2011) was used to calculate the basic diversity parameters among and within populations according to the maximum composite likelihood substitution model, with pairwise deletion and unequal rates among sites. The molecular phylogeny was reconstructed by Maximum Likelihood (ML) with RAxML v8.0 (Stamatakis 2006), and Bayesian analyses were performed using MrBayes v3.2.2 (Huelsenbeck and Ronquist 2001). The robustness of the nodes was estimated by bootstrap (BP) (1000
pseudoreplicates, starting tree: best tree obtained from 200 randomized MP starting tree) with RAxML and Posterior Probability (PP) using MrBayes - all parameters except topology are unlinked. Two simultaneous independent runs with 4 chains and 20 million generations each, were performed with one tree sampled every 500 generations and the burn-in of 5000 trees. Average genetic divergence between and within each clade was evaluated under the Kimura 2-parameter distance model of nucleotide substitution using MEGA 5.

**Chromosome analyses**

One hundred thirty-three specimens of *Mus musculus* were subjected to chromosome analyses. Chromosomes were obtained by the air-drying technique from bone marrow cells after yeast stimulation (Summer 1972). At least ten metaphases were analyzed for each population (including 2–8 individuals) using a 100x zoom digital CCD camera. The identification of chromosomal arms was performed by G-banding following the procedure of Seabright (1971) and according to the nomenclature of Cowell (1984). Means were compared by one-way ANOVA after Bartlett tested homogeneity. Tukey's test was carried out to measure differences between each pair of means. The C-banding method was performed according to the BSG (Barium/Saline/Giemsa) method of Summer (1972) with slight modifications. About 50 to 100 metaphase spreads from each specimen were examined, and at least 30 good chromosomal spreads were photographed using a 100x zoom digital CCD camera (Olympus, U-25ND25). The size of centromeric heterochromatin of all specimens was prepared by the Karyotype Analysis software, v: 1.2 (Yan Yu 2010).

**Morphometry**

We tested the classical biometric criteria for intraspecific distinction (Darvish 2008). Discriminative criteria can be identified using biometric studies on external and cranial characters of genetically typed specimens. Measurements of 167 adult specimens were used for biometric analyses. The biometric measurements classically employed to discriminate different Iranian mouse subspecies are:

1) The body length (BL)
2) The tail length (TL)
3) The foot length (FL)
4) The index of Tail length/ Body length (T/B).
5) The width of the molar process anterior part (A)
6) The width of the upper part of the zygomatic arc (B)
7) The zygomatic index on the skull (ZI: width of molar process anterior part/width of the upper part of the zygomatic arch).
8) The length of the lower tooth row (R.D.I)

These eight linear measurements were taken from 167 specimens (belonging to the four molecular clades; *M. m. bactrianus, M. m. musculus, M. m. domesticus* and *M. m. isatissus* from central Iran) using a digital caliper accurate to the nearest 0.05 mm. The data normality and homogeneity of variances (Levene's test) were checked. Then univariate and multivariate analyses such as ANOVA and Principle Component Analysis (PCA) were performed to extract significant segregation between the four clades, and sort out the studied clades based on morphometric differences. The significant level for all statistical analyses was set at $P < 0.05$. All statistical analyses were accomplished using SPSS ver. 16-2 (University of Bristol 2010), PAST ver. 2.08 (Hammer et al. 2011).

**Results**

**Molecular analysis**

In this study, 1111 nucleotides from the Cytochrome b (Cytb) gene belonging to 26 new specimens from this study, and 35 sequences retrieved from publications or GenBank were analyzed by molecular methods. The final aligned datasets were specified by 188 polymorphic sites defining 37 haplotypes, and 86 of these polymorphic sites were phylogenetically informative. The alignment
presented the following composition: A (31.6%), C (26.9%), T (29.1%), and G (13.6%), and the average of Transition/Transversion ratio was 4.9647. The lowest percentage of nucleotides, similar to previous works on mammals (Irwin et al. 1991), corresponds to guanines. Genetic distances based on p-distance among recognized subspecies (clade) varied between 1.92-3.32%. The smallest genetic distance was between M. m. bactrianus and M. m. isatisus (1.92%), and the highest genetic divergence between clades separated M. m. isatisus clade from M. m. domesticus (3.32%) (Table 1). The clade M. m. bactrianus had the highest diversity within subspecies (0.14%) and clade M. m. musculus showed the smallest intra genetic distance (0.04%) (Table 2).

**Phylogenetic analysis**

The maximum likelihood tree and Mrbey tree show typical result with two major clades (Southern-western and Eastern-central) and four main clades through the house mouse populations of the Iranian plateau; (1) M. m. musculus from the northeast of Iran, Uzbekistan and Kazakhstan (2) M. m. domesticus from the west and southwest of Iran (Ahvaz, Bandar-Abbas, Khalkhal, Uromia, Mammasani, Syria, and Turkey), (3) M. m. bactrianus from the east of Iran (Systan-va-Balouchestan, Neyshabur, Kerman, Tabas, Birjand, Sabzevar), and (4) M. m. isatissus from central Iran (Yazd, Shiraz, and Isfahan) (Fig. 2). Based on molecular trees, all clades emerged as monophyletic groups.

<table>
<thead>
<tr>
<th>Taxa</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Intraspecific</th>
<th>No. of Haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. M. m. isatisus</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>0.09 ± 0.01</td>
<td>7</td>
</tr>
<tr>
<td>2. M. m. bactrianus</td>
<td>1.92</td>
<td>± 0</td>
<td></td>
<td></td>
<td>0.14 ± 0.01</td>
<td>14</td>
</tr>
<tr>
<td>3. M. m. domesticus</td>
<td>3.32</td>
<td>± 3.10</td>
<td>± 0</td>
<td></td>
<td>0.07 ± 0.01</td>
<td>8</td>
</tr>
<tr>
<td>4. M. m. musculus</td>
<td>3.27</td>
<td>± 2.32</td>
<td>± 2.81</td>
<td>± 0</td>
<td>0.04 ± 0.01</td>
<td>9</td>
</tr>
</tbody>
</table>

**Table 1.** Genetic distances (p-distance ± SD) within and between Mus musculus subspecies based on Cytb sequences. Inter-intraspecific distances are shown, respectively.
showed the smallest size of CHR in all localities. Populations of *M. m. bactrianus* showed larger CHRs in their chromosomes as compared to other populations and populations of *M. m. musculus* showed smaller CHRs in their chromosomes compared to other populations. Comparative studies in CHRs populations showed four clear and sharp groups that can be easily recognized, (1) specimens of *M. m. domesticus* clade with small CHRs in almost all chromosomes except three chromosomes (4, 10 and 13), (2) specimens of *M. m. musculus* clade with the smallest CHRs in all chromosomes that have the same size in all chromosomes, (3) *M.

### Table 2. External (BL, TL, FL, T/B) and cranial (ZI, R.D.I, R.D.U) measurements of specimens belonging to the four sub-species. The values are displayed in millimeters. For all species, mean (M), range and sample size (n) is given.

|                | *M. m. isatissus* (n=14) | *M. m. bactrianus* (n=12) | *M. m. musculus* (n=16) | *M. m. domesticus* (n=4) | ANOVA P<  
|----------------|--------------------------|----------------------------|--------------------------|--------------------------|-----------
| **BL**        | 43 ± 76.16               | 56 ± 81.12                 | 58 ± 77.96               | 60 ± 70.91               | >0.05     |
| **TL**        | 48 ± 82.60               | 61 ± 83.06                 | 55 ± 70.91               | 58 ± 80.91               | >0.05     |
| **T/B**       | 0.82 ± 0.54              | 0.81 ± 0.09                | 0.76 ± 0.11              | 0.82 ± 0.58              | 0.05     |
| **FL**        | 22 ± 16.63               | 10 ± 15.83                 | 11 ± 14.59               | 14 ± 1.41               | >0.05     |
| **A**         | 0.29 ± 0.54              | 0.29 ± 0.13                | 0.45 ± 0.58              | 0.4 ± 0.26              | 0.05     |
| **B**         | 0.43 ± 0.92              | 0.44 ± 0.114               | 0.39 ± 0.211             | 0.39 ± 0.62             | 0.05     |
| **ZI**        | 0.39 ± 0.91              | 0.386 ± 0.241              | 0.59 ± 0.047             | 0.5 ± 0.26             | >0.05     |
| **R.D.I**     | 2.21 ± 3.05              | 3.8 ± 3.11                 | 3 ± 3.1                  | 3.36 ± 2.03             | 0.05     |

**Figure 2.** Phylogenetic consensus tree for 61 mtDNA (Cytb) sequences referred to *Mus musculus*, *Mus spicilegus* is used as out-groups. The color groups indicate Iranian specimens in this study.
m. isatissus clade with long and very heterogenous CHRs size in chromosomes and (4) specimens of M. m. bactrianus clade that showed uniform and long size in CHRs (Fig. 4). Groups 3 and 4 in CHRs size are different and distinguishable, before this, we did not have any clear characters for distinguishing these clades, but this study revealed that centromeric heterochromatin characters could be beneficial for showing borders in these subspecies.

**Morphometry**
Table 2 summarizes the range of morphometric variables used by Darvish (Darvish 2008, Darvish 1995). The mean ratio of body size/caudal length, Zygomatic index, lower row, of Mus m. domesticus (molecular clade) is 0.92, 0.59, 2.03, respectively. These parameters in M. m. musculus were 0.82, 0.67, and 3.05. In the subspecies M. m. isatissus and M. m. bactrianus were 1.19, 0.60, 3.05 and 1.02, 0.62, 3.1, respectively. These results showed that the importance of morphometric characters in house mice cannot show the difference between some subspecies, and in principal component analysis, almost all groups overlapped. The first two axes of the PCA explained 80% of the total variance. The characters T/B, ZI, and R.D.I, constituted most of the correlations with the PCA, and the characters BL and TL showed most correlation with the PCA2 (Fig. 5).

---

**Figure 3.** Haplotype network tree for 61 mtDNA (Cytb) sequences referred to *Mus musculus*
Figure 4. Phenotype and C-banded karyotypes of the house mouse of (A) Mashhad (*M. m. musculus*), (B) Zabol (*M. m. bactrianus*), (C) Shiraz (*M. m. isatissus*) and (D) Uromia (*M. m. domesticus*). Scale bar = 40 μm
Discussion

*M. musculus* is not the best-known mammal species that have at least three main subspecies: *M. m. domesticus; M. m. musculus;* and *M. m. castaneus* (Bonhomme et al., 2012). Ancestors of the house mouse originated in 0.5 Mya and diversified approximately 250,000 years ago into three main lineages (Cucchi et al. 2012). There are many grey zones in *M. musculus*; these populations remain relatively poorly described (Bonhomme et al. 2012). The house mouse is located somewhere between the north of the Indian subcontinent and the adjacent regions of Iran and Afghanistan (Boursot et al. 1996). Populations in these areas are considered as more polymorphic (Phifer-Rixey et al. 2012, Rajabi-Maham et al., 2012). Some studies showed that this region had been the cause of speciation for rodents' genus like *Allactaga, Jaculus, Meriones, Gerbillus* and *Calomyscus* (Dianat et al. 2013, Shahabi et al. 2013). Moreover, there are many cases of house mouse populations in this region that cannot be described in the three major subspecies, which were designated as 'oriental lineages' by Boursot et al. (1996). In Iran *M. m. castaneus* has been described in vast regions covering the border of Afghanistan to the north, south, and west of Iran (Guenet 2003, Darvish 2008, Rajabi-Mahan et al., 2012). In recent years, researchers have shown new subspecies in the central regions of Iran (Hardouin 2015, Darvish 2014). Our analysis of mitochondrial sequences and chromosomal investigation from house mice identified two different clades in samples named previously as *M. castaneus*. The phylogenetic analysis of the *Iranian sample* sequences clearly shows that all but two of the sequences cluster together within a single major clade with a very low divergence level. The low level of divergence within the major clade may be indicative of a recent common origin for the insertion event. There are no morphologically diagnostic differences between *M. m. bactrianus* and *M. m. isatisissus*. Our results show that the eastern/central clade consists of two distinct groups. Some morphometric characters used before for recognition of house mouse subspecies in this among some of them were landar in different analysis, for example, ZI values, T/B and R.I.D (Auffray et al. 1990), their results indicated that these characters show different morphologic groups but these groups overlap with each other. Moreover, morphometric characters could not identify two

Figure 5. Principal component analysis (PCA) based on First (1) and Second (2)
central-eastern subspecies (M. m. isatiss and M. m. bactrianus).

**Conclusion**

All subspecies are strongly supported in centromeric heterochromatin size. The nature and extent of chromosome variability within this subclade should be assessed by standard cytogenetic methods. These comparative studies showed that we can use chromosomal characters, especially centromeric heterochromatin variations for differentiation between house mouse subspecies. This character is clear and easy, and in the Iranian house mouse, these patterns were enough for the identification of subspecies.

**Acknowledgments**

The authors express their gratitude to Dr. Malekzadeh for his generous help and M. Mahmoodi, M. Mohammadyari, M. Habbibi, A. Khajea and A. Khosravi for their great efforts in collecting the specimens. This study was supported by a grant (no. 27137) from Ferdowsi University of Mashhad.

**References**


