Analysis of four polymorphic minisatellites suggests frequent genetic exchanges among House Mouse subspecies

Eric Rivals, François Bonhomme, Annie Orth, Gemma R Grant, Alec J Jeffreys, Philippe Rj Bois

Analysis of four polymorphic minisatellites suggests frequent genetic exchanges among House Mouse subspecies

Eric Rivals[1], François Bonhomme[2], Annie Orth[2], Gemma Grant[3], Alec Jeffreys[3], Philippe Bois[4]

[1] LIRMM, CNRS Université de Montpellier 2, UMR 5506, 161 rue Ada, Montpellier 34392 Cedex 5, France
rivals@lirmm.fr

[2] Biologie Intégrative, ISEM CNRS Université de Montpellier 2, UMR 5554, Montpellier 34095, France
{bonhomme,a-orth}@univ-montp2.fr

[3] Department of Genetics, University of Leicester, Leicester LE1 7RH, UK, ajj@leicester.ac.uk

[4] The Scripps Research Institute, Department of Cancer Biology, Genome Plasticity Laboratory, Parkside Drive, Jupiter, Florida 33458, USA, pbois@scripps.edu

Polymorphic genetic features, such as variable number of tandem repeats (VNTRs), are regarded as the most informative markers due to their intrinsic high variability. Minisatellites are particularly informative, as shown by their early use in forensics and paternity testing in humans [7]. DNA typing of minisatellites provides a powerful and highly discriminating tool. Unlike microsatellites that are comprised of short repeats (1 to 6 bp), minisatellites are intermingled arrays of usually GC-rich variant repeats ranging from $[10, 100]$ bp and with array lengths in $[0.1, 20]$ kilobases. Intermingled patterns of variant repeats along the array can be charted by Minisatellite Variant Repeat mapping by PCR (MVR-PCR) to provide detailed information on internal allele structure [1]; see examples on Figure 1. Compared to human, the situation is more favourable in mouse for pedigree and genealogy analysis. Systematic isolation has identified human-like minisatellite loci (i.e., GC-rich, highly polymorphic; [2]). However, none were found to be hypermutable. Moreover, mouse sperm mutants arise by simple intra-allelic duplication and deletion [3]. This combination of high polymorphism, lower mutation rate, and relatively simple intra-allelic turnover mechanisms make mouse minisatellites potentially highly informative for species-wide population studies. Nevertheless, reconstructing the genealogy of alleles is hampered by the fact that aligning their sequences is difficult. Recently however, development of new algorithms specifically designed to treat tandem repeat data enable the analysis of large MVR datasets (MS_ALIGN [5]). This allows quantification of molecular divergence between alleles and renders these information-rich loci amenable to phylogenetic analysis.

We therefore used MVR-PCR together with the MS_ALIGN algorithm to study for the first time the distribution of allelic variants at four different minisatellite loci in the House Mouse (*Mus musculus*). This species has radiated outside its cradle within the last 0.5 MA, leaving at its periphery three well recognised subspecies with recent ancestry (*M. m. domesticus*, *M. m. musculus* , and *M. m. castaneus*, abreviated by DOM, MUS, and CAS resp.) and populations of a more ancient descent at its centre [6]. It has more recently expanded outside Eurasia because of commensalism with man. The question of allele circulation throughout the species range is crucial for understanding the impact of selective forces that shape complex eukaryotic genomes. However, for a standard nuclear DNA sequence the intra-specific nucleotide divergence is generally small, resulting in very short and poorly informative coalescent branches within subspecies. To characterise allele circulation among House Mouse subspecies, we report intra-specific coalescence analysis at four minisatellite loci: MMS 24, 26, 80, and 30 [2] located resp. on chromosomes 7 (22cM), 9 (68cM and 79 cM), and X (43 cM) on a panel of 116 mice. We analysed at four loci 116 individuals originating from 40 geographical locations and representing the diversity present in House Mouse subspecies. Internal structures of alleles were determined by MVR-PCR to produce one map per allele. We estimated molecular divergences between pair of alleles with the alignment program MS_ALIGN [5]. From the resulting distance matrices, we inferred alleles genealogies (*i.e.*, trees) with a Neighbour-Joining method.

Figures 6 – 9 of [4] show the coalescence trees observed at all loci. One striking feature is the variable degree of subspecific coalescence observed, which goes from almost complete resolution of the *domesticus*, *musculus*, and *castaneus* clades for the X chromosome locus MMS30 to a more interspersed situation for MMS24. Nevertheless in all four trees, small clades of almost pure subspecific composition could be identified. We found many examples where identical or nearly identical haplotypes are shared among geographically distant subspecies. These alleles represent "intruders" in the abovementioned subspecific clades. The most noticeable case is the one of five CAS/central haplotypes placed in the DOM subtree at locus MMS30. When looking at the mutiple alignment in Figure 4 of [4] (or Figure 1 below), it is striking that these intruder haplotypes differ considerably from the typical CAS/central MVR codes, and resemble much more the DOM or MUS haplotypes. This is confirmed by conserved sequence motifs and average distances to the DOM vs CAS sets of alleles. Given the alleles similarity, the absence of long branches for these intruders, and the mutation mode of mouse minisatellites, homoplasy cannot explain such observations. The level of subspecific resolution in the MMS30 tree suggests that such haplotypes are introgressed alleles due to recent genetic exchanges between subspecies. Such evidence also underline the ability of MS_ALIGN to correctly handle complex cases. It exploits the properties of rapid simple mutation and complex internal structure at minisatellites to provide far more informative systems compared to classical markers.

Our findings suggest that, at least for the chromosomal regions under scrutiny, wild House Mice subspecies constitute a set of interrelated gene pools still connected through long range gene flow or genetic exchanges occurring in the various contact zones existing nowadays or having existed in the past. Identifying genomic regions that do not follow this pattern will be a challenging task for pinpointing genes important for speciation. This work is published in [4] and evaluated [8].

```
DOM  Geor  GYKKKWGKLKY oWKGWGK o GGYWYKKoK ˜ ˜ ˜YYY˜KG
DOM  Espa  GYKKKWGKLKW oWKGWGK o GGYWYKKoKKK ˜YYY˜KG
DOM  Fran  GYKKKWGKLKY LWKGWGKWGGYWYKKoKKK ˜YY ˜ ˜ ˜ G
DOM  Turq  GYKKKWGKLKY oWKGWGLWGGYWYKKoKKK ˜YYY˜KG
CEN  Paki  GYKKKWGK o KY KWKGWGK o GGYWYKKoKKK ˜YYY˜KG
CEN  Inde  GYKKKWGLWKY o o Ko WGLWGGYWYKKoKKKKYY ˜ ˜KG
CAS  Mada  GYKKKWGL LKY L o KGWGL o GGKWKKKoKKKKYY ˜ ˜KG
           * * * * * * *    * .    * . * *    * * . * . * * * * . .   * *    . *
```

Figure 1: Multiple alignment of DOM haplotypes and CAS/central intruders from MMS30.

Acknowledgements: E.R. is supported by is supported by the ACI IMPBIO REPEVOL.

References

[1] P. Bois and A. J. Jeffreys. Minisatellite instability and germline mutation. *Cellular and Molecular Life Sciences*, 55(12):1636–1648, 1999.

[2] P. Bois, J. D. H. Stead, S. Bakshi, J. Williamson, R. Neumann, B. Moghadaszadeh, and A. J. Jeffreys. Isolation and Characterization of Mouse Minisatellites. *Genomics*, 50(3):317–330, June 1998.

[3] P.R.J. Bois, G.R. Grant, and A.J. Jeffreys. Minisatellites show rare and simple intra-allelic instability in the mouse germ line. *Genomics*, 80(1):2–4, July 2002.

[4] F. Bonhomme, E. Rivals, A. Orth, G. R. Grant, A.J.J. Jeffreys, and P. Bois. Species-wide distribution of highly polymorphic minisatellite markers suggests past and present genetic exchanges among House Mouse subspecies. *Genome Biology*, 8:R80, 2007.

[5] S. Bérard and E. Rivals. Comparison of minisatellites. *J. Comp Biol*, 10(3-4):357–372, 2003.

[6] J-L. Guénet and F. Bonhomme. Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet*, 19(1):19–31, 2003.

[7] A. J. Jeffreys, V. Wilson, and S. L. Thein. Hypervariable 'minisatellite' regions in human DNA. *Nature*, 314(6006):67–73, Mar 1985.

[8] Faculty of 1000 Biology. `http://www.f1000biology.com/article/id/1084935/evaluation`.