The Microevolution of Mouse Salivary Androgen-Binding Protein (ABP) Paralleled Subspeciation of *Mus musculus*

J. M. Hwang, J. R. Hofstetter, F. Bonhomme, and R. C. Karn

Mouse salivary androgen-binding protein (ABP; Dlouhy and Karn 1983) is widely distributed among rodent species although its precise function remains to be determined (Karn 1991; Karn and Dlouhy 1991). The Alpha subunit, which is common to all forms of ABP observed so far, is encoded by *Abpa* located on chromosome 7 (Dlouhy et al., 1987). The Alpha subunit shares significant identity with chain 1 of the cat allergen Fel d1, rabbit uteroglobin, and human lung Clara 10 (Karn 1994). Karn and Russell (1993) reported limited identity between a second subunit, Beta, and helospectin, a gila monster venom protein.

Microevolution of the ABP Alpha subunit gene (*Abpa*) appears to have occurred in the course of the mouse subspeciation that gave rise to the *Mus musculus* complex [Karn and Dlouhy 1991; see Bonhomme (1986a,b) and Boursot et al. (1993) for a description of the complex]. Three electrophoretic variants [electromorphs a (*Alpha* or A*), b (A') and c (A'')] of the Alpha subunit of ABP have been found in wild mouse populations, with allelic distributions corresponding to the following *Mus musculus* subspecies: *domesticus* (*Abpa*), *musculus* (*Abpoc*), and *castaneus* (*Abpa*) (Karn and Dlouhy 1991). The allelic distribution amounts to distinct intersubspecies variation, each subspecies essentially having its own variant, but little intrasubspecies variation (Karn and Dlouhy 1991). Based on electromorphic data, it was unclear whether this circumstance is the result of selection or genetic drift. In an attempt to distinguish between these two possibilities, and to draw generalizations about the structure and function of ABP, we examined patterns of nucleotide substitution between polymorphic Alpha subunit alleles. Most coding sequences evolve under a high purifying selection pressure so that most missense mutations are readily eliminated. This results in a rate of synonymous substitution that is on the average fivefold higher, with a large variance between genes (Nei 1986). If, on the other hand, amino acid substitutions in a particular protein were all neutral, the genetic code predicts approximately a threefold excess of nonsynonymous substitutions for a protein with average amino acid composition since most substitutions on the first and second base pairs are nonsynonymous and about two-thirds of the mutations on the third base pair are synonymous. Therefore the synonymous/nonsynonymous substitutions ratio reflects the intensity of the constraints exerted on a protein's sequence to keep it functional. Thus, if ABP is subject to the constraints of an average protein, we expect mostly synonymous substitutions in the coding sequences of Alpha subunit variants.

We have examined the three *Abpa* alleles found in *M. musculus* subspecies at the level of the cDNA sequence that codes for the Alpha subunit secreted protein and the putative protein sequence that it determines. cDNA sequences were also produced for three other full species of *Mus*. The divergence of these sequences involved numerous nucleotide substitutions...
and amino acid replacements. We speculate both on what these findings may mean for ABP evolution, in conjunction with *M. musculus* subspeciation, and on the potential functional significance of the changes.

**Materials and Methods**

Male C3H/St (source of the Abpa* allele), DBA/2J (source of the Abpa allele; see Karn and Dlouhy 1991) for an explanation of how this allele was fixed in a few inbred strains), and CAST/Ei (source of the Abpa allele) mice were obtained as described previously (Karn and Dlouhy 1991). *Mus caroli*, *M. spreptus*, and *M. spicilegus* were obtained from Health Research Inc., Rockwell Park Division, Buffalo, New York (note that *M. spicilegus* was sold under the earlier species name *M. hortulanus*), where they are maintained as randomly bred wild strains.

An mRNA isolation kit (Invitrogen Corp., San Diego, California) was used to recover poly(A+) submaxillary gland RNA. cDNA was synthesized using 2–10 µg poly(A+) RNA (Bethesda Research Laboratories, Gaithersburg, Maryland, cDNA Synthesis System). Primers were derived from the sequence of an ABP Alpha subunit cDNA (Karn and Russell 1993) and purchased from Indiana University. They were

**Primer Code Sequence (5’ to 3’)**

S1083 AGATTTGTCGCTTTTGCCAGCTTT
S1078 CTGTTGTCATCTATGCTGTTGAGGA

Symmetric polymerase chain reactions (SPCR) were performed by a modification of Saiki et al. (1988). Unpurified SPCR product was the template in an asymmetric reaction (APCR) with only one amplification primer. After 50 cycles, the APCR product was the counterpart to the one used in APCR. An alternative sequencing strategy used the New England Biolabs (Beverly, Massachusetts) CircumVent kit with end-labeled primers.

PCR was run more than once on each cDNA to detect and avoid errors that could have been mistaken for real substitutions, but none were encountered. We also determined the sequences of seven *M. m. domesticus* and seven *M. m. musculus* *Abpa* coding regions for evaluation of intraspecies variation (see Results section).

Base substitutions were analyzed with DNA parsimony algorithms in the PHYLIP program to produce an unrooted dendrogram.

**Results**

Figure 1 shows the cDNA sequences, and the amino acid sequences they specify, of the three *Abpa* alleles found in the three subspecies of *M. musculus* (*Abpa* in *domesticus*, *Abpa* in *musculus*, and *Abpa* in *castaneus*) and those representing *M. caroli*, *M. spreptus*, and *M. spicilegus*. The nucleotide and amino acid replacements are summarized in Table 1. In the 210 bp region that codes for the secreted protein sequence, there are 16 positions at which base substitutions result in amino acid variability. In each case, the protein’s net charge is consistent with its electrophoretic mobility as reported by Karn and Dlouhy (1991). Our data suggest that the mobilities of ABP a and ABP b are due to amino acid replacements that do not result in a full charge change between the two Alpha subunits, however the difference of nearly 2 pK units between Lys and Arg is sufficient to account for the difference in electrophoretic mobility. A similar difference is seen between normal hemoglobin and the phoeretic mobility. A similar difference is seen between normal hemoglobin and the phoeretic mobility.

In order to examine within-subspecies variation and compare it to between-subspecies variation (McDonald and Kretman 1991), we amplified the Alpha subunit coding region from genomic DNAs from mice trapped in widely separated areas on

### Table 1

<table>
<thead>
<tr>
<th>Position</th>
<th>Subspecies</th>
<th>Nucleotide</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td><em>M. m. domesticus</em></td>
<td>GAG</td>
<td>Asp</td>
</tr>
<tr>
<td>20</td>
<td><em>M. m. musculus</em></td>
<td>GAG</td>
<td>Asp</td>
</tr>
<tr>
<td>30</td>
<td><em>M. m. castaneus</em></td>
<td>GAG</td>
<td>Asp</td>
</tr>
<tr>
<td>40</td>
<td><em>M. caroli</em></td>
<td>GAG</td>
<td>Asp</td>
</tr>
<tr>
<td>50</td>
<td><em>M. spreptus</em></td>
<td>GAG</td>
<td>Asp</td>
</tr>
<tr>
<td>60</td>
<td><em>M. spicilegus</em></td>
<td>GAG</td>
<td>Asp</td>
</tr>
</tbody>
</table>

**Figure 1.** Sequences of the Alpha subunits for *Mus musculus domesticus* (a), *musculus* (b) and *castaneus* (c); *M. caroli* (car); *M. spreptus* (spr); and *M. spicilegus* (spi). Base substitutions and corresponding amino acid replacements for nonsynonymous substitutions are shown at the positions of variability.
either side of the European hybrid zone and maintained as moderately inbred strains by Bonhomme (Karn and Dlouhy 1991). The sequences of seven *M. m. domesticus* and seven *M. m. musculus* *Abpa* coding regions (not shown) were the same as those for the a and b alleles, respectively (Figure 1).

Figure 2 shows an unrooted tree representing the maximum likelihood estimate of the phylogenetic relationships between *Abpa* genes. Substitutions are plotted on the tree proportionally to deduced branch lengths. Accordingly, 24 of them are necessary to explain the character states at the 20 variable positions, implying four homoplastic events on sites 12, 14 and 15 on the reduced data set. Taking into account these double events, this allows us to estimate the exact count of each type of substitution along the branches of the tree. We find 20 nonsynonymous substitutions for only 4 synonymous substitutions on the whole tree. Within the European species, we find only 1 nonsynonymous substitution for 11 nonsynonymous substitutions. At this point, it is clear that *Abpa* is very far from an average gene where nonsynonymous substitutions are rare. Here these amino acid substitutions appear to be the rule, how abnormally frequent are they? If we consider that one would have a 3:1 ratio under no selective constraint at all (for the 70 residues in the Alpha sequence, the ratio is actually 3.3:1), an exact binomial test yields a probability of 214 of getting such an excess of substitutions.

Another rather intriguing outcome of this analysis is the topology of the tree itself: it is highly incongruent with the now well-established consensus phylogeny of the genus [see Boursot et al. (1993) for a review] which, in all instances with all molecular markers used thus far, indicates that *domesticus*, *musculus*, and *castaneus* form a monophyletic group. On this basis, we believe that our results show that *Abpa* is a gene evolving very far from a typical canonical coding sequence and we strongly question the neutrality of the predominant amino acid substitutions that we see.

**Discussion**

Karn and Dlouhy (1991) noted that *Abpa* alleles have apparently been fixed, whether by drift or natural selection, each in a different mouse subspecies (*Abpa* for *domesticus*, *Abpa* for *musculus*, and *Abpa* for *castaneus*) of the *M. musculus* complex. We are interested in determining how and why this microevolution occurred. Here we report the base differences for the *Abpa*, *Abpb*, and *Abpc* alleles of *M. musculus*, as well as those representing three other full species: *M. caroli*, *M. spretus*, and *M. spicilegus*.

The nonsynonymous/synonymous substitutions ratio we observe—5:1—is slightly higher than what it would be—3:1—in the case of pure neutral evolution with no constraint at all. If purifying selection was acting, as it does in many proteins, the ratio would be much smaller (1:5 in an average protein, in other words 25-fold less). Moreover, we think that substitution mutations could have been favored during the course of recent evolution, for several indirect reasons.

The first argument pertains to rates of evolution. We find an average of 17 sub-
stitutions over 210 sites between the European taxa and *M. caroli* whose time of divergence is about 2.5 million years [see Boursot et al. (1993) for a review]. This allows us to estimate a minimum substitution rate of 1.6%/Myr in each lineage. This is higher than the values obtained for various mammalian pseudogenes in the same set of species with the same phylogenetic calibration, for which we have 1.14%, 1.31%, and 1.15%/Myr/lineage for the L1, alpha-globin, and LDH pseudogenes, respectively (Garcia-Meunier et al. 1993). It seems that there is even an excess of substitutions as compared to freely diverging sequences that can arise only by selective advantage of the mutants.

The second is that ABP has a somewhat unusual distribution of its polymorphisms across the *M. musculus* subspecies. Each of the three main subspecies has fixed its own allele, with very little variation in each group. Seven individual genes have been sequenced for each of *AbpA* and *Abpa*, always with the exact same sequence for each allele. In order for this to have occurred by drift alone, one must posit a strong bottleneck at the time of subspecies divergence. Furthermore, the bottleneck had to be followed by either too little time for the restoration of neutral variation or by the population remaining quite small so that the intraspecies diversity subsequently remained very limited. This scenario is both unlikely [the divergence of the three subspecies has been estimated through mtDNA analysis by Boursot et al. (1996) to be on the order of 0.5 million years] and incongruent with what actually has been observed in *M. musculus*, where no other genetic system in *M. musculus* shows such a clear-cut subspecies-specific distribution (Bonhomme 1986a; Moriwaki et al. 1986) and where abundant sequence polymorphism has been found intraspecies each time it has been studied, a good example of this being provided by mtDNA (see for instance Boursot et al. 1996).

The majority of fixed mutations in structural genes appear to be selectively neutral (Kimura 1983), however, notable examples of positive Darwinian selection are known, especially at functionally important sites (Fitch et al. 1991; Hughes and Nel 1988, 1989; Laskowski et al. 1987). Balancing selection, specifically overdominance (heterozygote superiority in fitness), has been suggested to account for a rate of nonsynonymous substitution in excess of synonymous substitution (Hughes and Nel 1988, 1989; Takahata et al. 1992), especially in the major histocompatibility complex (Mhc). Maruyama and Nel (1981) suggest that overdominance is probably not operating to maintain most polymorphisms except in isolated instances, such as the loci for histocompatibility and immunoglobulins which are usually polymorphic. They suggest that, in most species, the observed average heterozygosity is lower than the neutral expectation. They cite bottleneck effect and random fluctuations of selection intensity, but not overdominance, as possible factors. Our position, as already described, is that a bottleneck was unlikely and, in contrast to the two systems they describe, *Abpa* heterozygosity is abnormally low in wild populations of house mice, having been encountered in only a few near the European hybrid zone (Karn and Dlouhy 1991). Thus overdominance seems an unlikely explanation for the ratio of nonsynonymous to synonymous base substitutions observed in *Abpa* alleles. While we cannot rule out balancing selection, it seems to us that directional selection is a sufficient explanation, given our present state of knowledge, both for the unusual population distribution of *Abpa* alleles and for the large ratio of non-synonymous to synonymous substitutions we have observed in their coding sequences. It appears that a good deal of homoplasy occurred in the phylogeny of the *Abpa* haplotypes, as shown by the incongruence we found with the canonical phylogeny of the genus (Figure 2). We therefore envision the possibility of cyclical selection of certain amino acid variants that became advantageous at some stage. Once one was fixed, there was an opportunity for the reappearance of an earlier one at another site. This appears to concern a limited number of sites, especially amino acids 32, 33, 36, and 39 (Figure 1). Bonhomme and his colleagues have recently proposed that *M. musculus* is an incipient ring species with a double *Rassenkreis* (Boursot et al. 1996; Din et al. 1996). Three chains of radiation of the species from its origin in the north of the Indian subcontinent are envisioned, one branch spreading westward which eventually became the *domesticus* subspecies, one flowing north which eventually became the *musculus* subspecies, and one migrating southeastward which eventually became the *castaneus* subspecies. The *domesticus* and *musculus* subspecies made secondary contact in Europe where they established a hybrid zone, and the *musculus* and *castaneus* subspecies made a broad secondary contact in China. Apparently *Abpa* underwent microevolution in the course of radiation of these three branches, resulting in the distinct subspecies variation noted by Karn and Dlouhy (1991). Karn and Dlouhy (1991) and Karn and Russell (1993) have suggested that ABP may function in assortive mate selection, possibly resulting in prezygotic isolation where radiating subspecies made secondary contact. Dod et al. (1993) recognized the possibility that prezygotic isolation was contributing to the stability of the current hybrid zone between *M. m. domesticus* and *M. m. musculus*, but pointed out that no data bearing on that possibility existed.

Species recognition genes have been identified in sea urchins (Metz and Palumbi 1996; Palumbi and Metz 1991) and abalones (Swanson and Vaccieri 1995). The gene products mediate sperm attachment to eggs of the same species and inhibit heterospecific fertilization. Those studies demonstrate that polymorphism in mate recognition loci resulting in evolution of sexual isolation can arise within natural populations (Palumbi and Metz 1991) even among species that differ very little in their gene pools (Metz and Palumbi 1996). An intriguing functional parallel between these invertebrate species recognition systems and immune recognition has been drawn by Palumbi and Metz (1991). This is an exciting suggestion in light of the recent observation that the Alpha subunit of mouse ABP shares significant homology with chain 1 of the major cat allergen, Fel d 1 (Karn 1994). Since others have observed that cats contaminate their environments heavily with Fel d 1 (Morgenstern et al. 1991), perhaps mice also deposit ABP in their environment. We speculate that the human allergic reaction to such proteins in the dander of cats and mice is an abnormal version of a normal recognition response those animals make when they encounter such markings in their environment.

Could functional interactions between *Abpa* alleles contribute to subspecies recognition and therefore stability of hybrid zones between subspecies? The observation that *Abpa* is more different from *Abpa* and *Abpa* than those two alleles are from each other is consistent with the more clearly defined hybrid zone between the *domesticus* and *musculus* subspecies than that between *musculus* and *castaneus* (Din et al., in press). Future studies of *Abpa* microevolution will include a detailed analysis of the transition of alleles
across the *domesticus/musculus* hybrid zone as a natural test of compatibility of the a and b alleles in interspecific crosses.

References


Received December 20, 1995
Accepted July 29, 1996

Corresponding Editor: Christine Kozák

Hwang et al. Microevolution of Mouse Salivary ABP 97