Oligoadenylate synthetases (OASs) are interferon-inducible enzymes that participate in the first line of defense against a wide range of viral infection. Recent studies have determined that Oas1b, a member of the OAS gene family in the house mouse (Mus musculus), provides specific protection against flavivirus infection (e.g., West Nile virus, dengue fever virus, and yellow fever virus). We characterized the nucleotide sequence variation in coding and noncoding regions of the Oas1b gene for a large number of wild-derived strains of M. musculus and related species. Our sequence analyses determined that this gene is one of the most polymorphic genes ever described in any mammal. The level of variation in noncoding regions of Oas1b is an order of magnitude higher than the level reported for other regions of the mouse genome and is significantly different from the level of intraspecific variation expected under neutrality. Furthermore, a phylogenetic analysis of intronic sequences demonstrated that Oas1b alleles are ancient and that their divergence predates several speciation events, resulting in transspecific polymorphisms. The amino acid sequence of Oas1b is also extremely variable, with 1 out of 7 amino acid positions being polymorphic within M. musculus. Oas1b alleles are comparatively more divergent at synonymous positions than most autosomal genes and the ratio of nonsynonymous to synonymous substitution is remarkably high, suggesting that positive selection has been acting on Oas1b. The ancestry of Oas1b polymorphisms and the high level of amino acid polymorphisms strongly suggest that the allelic variation at Oas1b has been maintained in mouse populations by long-term balancing selection.

Introduction
Organisms and their pathogens are engaged in a coevolutionary process resulting from their conflicting interests: hosts must retain their ability to resist and eventually clear infection, and pathogens must retain their infectivity and ability to evade host defense. Host defense genes are in the forefront of this process and bear the signature of the interactions between the host and its pathogens. Many defense genes have been shown to evolve rapidly at the protein level because of the action of positive selection. This pattern of evolution is caused either by a succession of selective sweeps of new resistance alleles (the “arms race” model of evolution) or by the persistence of dynamic polymorphisms due to selective fluctuations in allelic frequency (the “Red Queen” model) (for a review, see Woolhouse et al. 2002). The selective mechanism responsible for the maintenance of polymorphism in populations is generally referred to as balancing selection and can be caused by heterozygote advantage, frequency-dependent selection (Kojima 1971), or selection that varies in time and space (Hedrick 1976). Although most balanced polymorphisms tend to be transient, in some very rare cases, balancing selection has been able to maintain alleles for very long periods of evolutionary time. At the major histocompatibility complex (MHC) in mammals (Hedrick and Thomson 1983; Figueroa et al. 1988) and at the plant resistance (R) genes (Bergelson et al. 2001), a remarkably large number of alleles have been maintained by balancing selection, and some of these polymorphisms are so ancient that they predate several speciation events, resulting in transspecific polymorphisms (Figueroa et al. 1988).

Key words: balancing selection, Mus musculus, Oas1b, West Nile virus, flavivirus.

E-mail: stephane.boissinot@qc.cuny.edu.

doi:10.1093/molbev/msn106
Advance Access publication May 5, 2008

© The Author 2008. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org
(e.g., Murray Valley encephalitis virus [Sangster et al. 1993]). It was found that these 2 resistance alleles differ at 14 amino acid positions (i.e., 4.2%) (Perelygin et al. 2002). This level of amino acid divergence is extremely high and is comparable with the divergence usually observed between distantly related species (e.g., mouse and rat proteins differ on average by 5% at the amino acid level; Gibbs et al. 2004). Although these data suggest that balancing selection could be acting on this gene, this pattern could also result from the complex evolutionary history of the house mouse, a species in which hybridization between subspecies (Ferris et al. 1983; Orth et al. 1998; Bonhomme et al. 2007) and with closely related species is known to occur (Greene-Till et al. 2000; Orth et al. 2002).

We assessed the variation of the Oas1b gene and we analyzed its molecular evolution in the house mouse and related species. Coding and noncoding regions of Oas1b were sequenced in a sample of wild-derived strains representative of the entire genetic diversity of this species. The Oas1b gene is extremely polymorphic within M. musculus and even within its subspecies, Mus musculus musculus and Mus musculus domesticus. Polymorphisms at Oas1b are very ancient and predate several speciation events. Our data demonstrate that the Oas1b gene has evolved under balancing selection for more than 2.8 Myr and constitute one of the very few cases of old transspecific polymorphisms maintained by long-term balancing selection in a mammal.

Materials and Methods

Sampling

The house mouse M. musculus is a complex species that originated in India less than 1 MYA (Boursot et al. 1996). From Northern India, this species gradually colonized the periphery of the Euro-Asiatic continent where 3 well-defined subspecies are found (M. m. musculus from Eastern Europe to Northern China; M. m. domesticus in Western Europe, North Africa, and the Middle East; and M. m. castaneus in South East Asia). These 3 subspecies still hybridize in natural conditions so that a clear-cut hybrid zone exists in Europe, whereas the subspecies seem to intermix to a greater extent in Asia. Mice from Japan are hybrids between M. m. musculus and M. m. castaneus and are sometimes referred to as “Mus musculus molossinus” (Yonekawa et al. 1986). We obtained DNA samples from 34 wild-derived strains of house mice, 23 from the “Conservatoire de la souris sauvage” (Université Montpellier II, Montpellier, France), and 11 from the Jackson laboratory (Bar Harbor, ME). These strains have been maintained in captivity for many generations so that they are homozygous at most loci, making the determination of haplotypes trivial. This sample includes strains from each of the 3 major subspecies, M. m. musculus (11 strains), M. m. domesticus (12 strains), and M. m. castaneus (4 strains); strains with undefined subspecific status from Iran, India, and Pakistan (4 strains); and hybrid strains from Japan (5 strains). In addition, we obtained DNA from other species within the genus Mus to determine the phylogenetic context of Oas1b evolution. These species include the 4 European species of mice, Mus spretus (3 strains), Mus mace-}

\textit{donicus} (3 strains), \textit{Mus spicilegus} (4 strains), and \textit{Mus cypricus} (1 strain), which collectively diverged from the house mouse ~1.4 MYA, the Indian species \textit{Mus famulus} (2.8 MYA), and the Southeast Asian species \textit{Mus caroli} (3.2 MYA). As an outgroup, we used the species \textit{Coelomys pahari} that diverged from the genus \textit{Mus} 6.5 MYA.

Molecular Analysis

Each of the 6 exons of Oas1b was amplified by polymerase chain reaction (PCR). We also amplified partial fragments from introns 2, 3, 4, and 5, from fragment containing the 3’ untranslated region (3’ UTR), and from a non-coding region located 10 kb downstream of the Oas1b gene. The products of amplification were purified and sent for sequencing to the company Macrogen (Seoul, Korea). To assess the variation of the Oas1b gene, we compared it with the level of variation in 2 neutral regions of the mouse genome. We selected 2 autosomal segments located at least 1 Mb from any known gene, 1 on chromosome 12 (from nucleotides 64,811,904 to 64,813,152 on the February 2006 assembly of the mouse genome at http://genome.ucsc.edu) and 1 on chromosome 19 (from nucleotides 19,726,706 to 19,727,769). These fragments were amplified by PCR in each wild-derived strain and sequenced. The list of primers used for amplification is available on request from the corresponding author. DNA sequences have been deposited at the EMBL database under GenBank accession numbers AM887890–AM887932.

Data Analysis

Sequences were aligned and manipulated using the BioEdit platform (Hall 1999). Distances, including synonymous (ds) and nonsynonymous (dn) distances, were calculated using the Nei and Gojobori (1986) method as implemented in the MEGA3 software package (Kumar et al. 2004). Phylogenetic analyses were performed using the Neighbor-Joining method (as implemented in MEGA3) and the maximum likelihood method (using PHYML [Guindon et al. 2005]). The nucleotide diversity was estimated using the parameter \(\pi\) and Watterson’s \(\theta\), and Tajima’s \(D\) (Tajima 1989) was calculated to assess the effect of selection on polymorphisms. We also tested for selection using the Hudson–Kreitman– Aguade test (known as the HKA test [Hudson et al. 1987]). The HKA test is based on the assumption that, in the absence of selection, silent site polymorphisms and divergence are expected to be the same across all loci. Basically, the test compares the ratio of polymorphism with divergence between a gene of interest (Oas1b in our case) and a neutral region of the genome (the neutral regions on chromosomes 12 and 19, see the Molecular Analysis). If the difference between the 2 ratios is statistically significant from the null hypothesis using a goodness of fit test, we can then reject the hypothesis of neutrality. We also used the McDonald and Kreitman (1991) test (known as the MK test): if variation at a locus is neutral, then the rate of substitution between species and the amount of variation within species are a function of the mutation rate. Therefore, under neutrality, the ratio of
nonsynonymous to synonymous fixed differences between species should be the same as the ratio of nonsynonymous to synonymous polymorphisms within species. If the 2 ratios differ significantly, we can reject the hypothesis of neutrality. The DnaSP program (Rozas et al. 2003) was used to perform the HKA and MK tests. We tested the possibility that gene conversion between alleles and/or between paralogous copies of \(Oas1b\) affected the sequence of \(Oas1b\) alleles using the method of Sawyer (1989) implemented in the GENECONV program.

Results

The coding sequence of \(Oas1b\) was obtained for 34 wild-derived strains of the house mouse (fig. 1). Fifty-five amino acid variants (at 51 positions) were detected within the species \(M.\) \textit{musculus}. This number is extremely large considering that this gene is only 377 amino acid long, that is, about 1 out of 7 amino acid positions is variable. As expected, we did not observe a single case of heterozygosity and the determination of linkage between polymorphisms was trivial. The lack of apparent heterozygosity strongly suggests that our PCR primers amplified only \(Oas1b\) and not other paralogous copies. The premature stop codon responsible for the susceptible phenotype of the laboratory strains was found in 3 wild-derived strains (SKIVE, SF, and DOT), and another stop codon, also in exon 4, was found in strain MBS. Single base pair deletions resulting in frameshifts were detected in exons 1 and 4 of strains CTP and 22MO, respectively, and an in-frame deletion of 39 nt resulting in a 13 amino acids deletion was found in exon 5 of the BZO strain. Out of the 55 variable amino acids, 24 polymorphisms are shared between at least 2 subspecies. These shared amino acid polymorphisms result from the same mutations at the DNA level indicating that
they are not the product of convergent evolution. In contrast, a single amino acid difference is fixed in one of the M. musculus subspecies (at amino acid position 45, in M. m. musculus, including M. m. molossinus). This pattern of variation extends to interspecific comparisons: there is not a single fixed amino acid that differentiates M. musculus from other species. In contrast, 5 polymorphisms are shared between M. musculus and either M. spicilegus or M. spreitus, and this number is presumed to be an underestimate due to the limited availability of strains of species other than M. musculus. The lack of fixed differences between taxa and the abundance of shared polymorphisms suggests that many polymorphisms are older than the origin of M. musculus and at least 5 of them are older than the split between M. musculus and other European species (M. spreitus, M. spicilegus, M. macedonicus, and M. cypriacus), which occurred ~1.4 MYA.

To confirm the ancestry of Oas1b alleles, the synonymous divergence (ds) between M. m. musculus and M. m. domesticus alleles was compared with the divergence of 183 autosomal genes for which a M. m. domesticus (from the C57BL/6J strain) and a M. m. musculus (from the CZECHI/EiJ strain) sequence are available in databases (fig 2A). On average, Oas1b alleles are much more divergent at synonymous sites than most autosomal genes and the ds between some Oas1b alleles can be as high as 4%. Only a handful of genes show such high ds, including several MHC genes. Figure 2A also shows that the distribution of intersubspecific divergences is almost indistinguishable from the distribution of interspecific divergences. On average, M. m. musculus and M. m. domesticus Oas1b alleles differ by 2.85% at synonymous sites, whereas they differ on average by 2.22% from European species. This clearly implies that allelic variation at the Oas1b gene is as old or older than the split between M. musculus and its European relatives.

The persistence of ancestral polymorphisms at Oas1b was further examined by performing a phylogenetic analysis on intronic sequences. Figure 3A shows a phylogenetic tree built using intron 5 sequences. This tree differs drastically from the species tree and shows strong evidence for transspecific polymorphisms. Two clades supported by high bootstrap values and by several indels are apparent on the tree. Each of these clades contains M. m. domesticus sequences and sequences from other species. M. m. domesticus sequences of clade 1 are more closely related to M. macedonicus, M. spicilegus, and M. spreitus than they are to M. m. domesticus sequences belonging to clade 2. Similarly, M. m. domesticus alleles in clade 2 are more closely related to M. spicilegus strain ZRU and to M. famulus than to clade 1. This clearly demonstrates that M. m. domesticus contains 2 deeply divergent allelic lineages at the Oas1b locus. These 2 lineages have separated before the split between M. musculus and M. famulus, which took place ~2.8 MYA. These lineages have also been maintained in M. spicilegus because sequences of this species are found in both clades.

Phylogenetic analyses were also performed on introns 2, 3, and 4 and on the 3' UTR (fig. 3B). The phylogenetic trees for each of these sections of Oas1b are drastically different. Trees based on intron 4 and on the 3' UTR are similar to the intron 5 tree and show that M. m. domesticus...
Long-Term Balancing Selection at *Oas* (*b*)

A. Intron 5

B. Intron 2

C. Intron 3

D. Intron 4

E. 3' UTR

[Diagram of genetic relationships and species distribution involving *M. domesticus* and *M. spicilegus*]
alleles, and to a lesser extent \( M. \) \( m. \) \( musculus \) alleles, fall into 2 distinct groups, supported by high bootstrap values. These groups are separated by long branches suggestive of their ancestry, although the phylogenetic analysis does not provide strong statistical support for transspecific polymorphism in intron 4 and in the 3' UTR. The 2 divergent \( M. \) \( m. \) \( domesticus \) clades differ by 2.35% (±0.52) and 2.70% (±0.61), for intron 4 and the 3' UTR, respectively, whereas they differ by 1.80% (±0.52) and 2.70% (±0.62) from \( M. \) \( f. \) \( monosulcus \), again suggesting that ancient alleles were maintained in \( M. \) \( musculus \) since before its split from \( M. \) \( f. \) \( monosulcus \). The composition of the \( domesticus \) groups is the same for intron 5 and the 3' UTR but is different from intron 4 tree. Strains DIK, PERC and PERA belong to the same clade as strains DGA and DJO for intron 5 and the 3' UTR, but these strains are closer to strains SF, SKIVE, and DOT on the intron 4 tree. This suggests that recombination is affecting the evolutionary history of \( Oas1b \) alleles by shuffling intronic sequences, although it appears the homogenizing effect of recombination has not been sufficient to mask the persistence of ancestral polymorphisms at \( Oas1b \). The trees built using introns 2 and 3 show a radically different pattern. The tree based on intron 3 sequences also shows some level of mixing between the subspecies of \( M. \) \( musculus \) and \( M. \) \( spicilegus \) and \( M. \) \( macedonicus \), but on this tree the \( M. \) \( f. \) \( monosulcus \) sequence falls clearly outside the diversity of \( M. \) \( musculus \) alleles. In addition, the divergences between \( M. \) \( musculus \) sequences are much smaller than the one observed for introns 4 and 5 and for the 3' UTR. The most divergent groups of \( M. \) \( m. \) \( domesticus \) sequences on the intron 3 tree differ by 0.68% (±0.19), which is almost 5 times smaller than the divergence between groups for intron 5 (3.29 ± 0.66%). On the trees based on intron 2, \( M. \) \( m. \) \( domesticus \) and \( M. \) \( musculus \) sequences form monophyletic groups of closely related sequences with no evidence of transspecific polymorphism, and the topology of this tree is very similar to the species tree.

For both \( M. \) \( m. \) \( domesticus \) and \( M. \) \( musculus \), the nucleotidic diversity in introns 4 and 5 and in the 3' UTR is 5–10 times higher than the nucleotidic diversity in introns 2 and 3 and in 2 putatively neutral regions of the mouse genome (table 1). This level of polymorphism is among the highest reported for any genomic region of the house mouse and is clearly outside the range reported by Zhang et al. (2005) for 44 genomic segments (~7 single nucleotide polymorphisms/kb within \( M. \) \( m. \) \( domesticus \)). The values of Tajima's \( D \) were negative for introns 2, 3, and 4 and for the 2 neutral regions. In introns 5 and in the 3' UTR, Tajima's \( D \) takes positive values, although we did not detect any significant deviation from neutrality.

### Table 1

<table>
<thead>
<tr>
<th>Regions</th>
<th>( L^* )</th>
<th>( N^b )</th>
<th>( S^c )</th>
<th>( \pi (%)^d )</th>
<th>( \theta (%)^e )</th>
<th>( D^f )</th>
<th>( \chi^2 )</th>
<th>( P ) Value</th>
<th>( \chi^2 )</th>
<th>( P ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intron 2</td>
<td>1525</td>
<td>13</td>
<td>23</td>
<td>0.309</td>
<td>0.486</td>
<td>-1.05</td>
<td>0.248</td>
<td>0.618</td>
<td>0.593</td>
<td>0.441</td>
</tr>
<tr>
<td>( M. ) ( m. ) ( musculus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M. ) ( m. ) ( domesticus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 3</td>
<td>725</td>
<td>11</td>
<td>12</td>
<td>0.144</td>
<td>0.238</td>
<td>-1.78</td>
<td>1.808</td>
<td>0.179</td>
<td>0.134</td>
<td>0.715</td>
</tr>
<tr>
<td>( M. ) ( m. ) ( musculus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M. ) ( m. ) ( domesticus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 4</td>
<td>710</td>
<td>13</td>
<td>13</td>
<td>0.152</td>
<td>0.285</td>
<td>-1.61</td>
<td>0.176</td>
<td>0.675</td>
<td>0.919</td>
<td>0.338</td>
</tr>
<tr>
<td>( M. ) ( m. ) ( musculus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M. ) ( m. ) ( domesticus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 5</td>
<td>690</td>
<td>13</td>
<td>13</td>
<td>1.356</td>
<td>1.649</td>
<td>-0.31</td>
<td>1.672</td>
<td>0.196</td>
<td>0.088</td>
<td>0.767</td>
</tr>
<tr>
<td>( M. ) ( m. ) ( musculus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M. ) ( m. ) ( domesticus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3' UTR</td>
<td>667</td>
<td>8</td>
<td>8</td>
<td>1.395</td>
<td>1.438</td>
<td>1.14</td>
<td>2.147</td>
<td>0.143</td>
<td>4.920</td>
<td>0.027</td>
</tr>
<tr>
<td>( M. ) ( m. ) ( musculus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M. ) ( m. ) ( domesticus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome 12</td>
<td>817</td>
<td>10</td>
<td>10</td>
<td>0.435</td>
<td>0.277</td>
<td>-0.99</td>
<td>1.121</td>
<td>0.289</td>
<td>0.109</td>
<td>0.742</td>
</tr>
<tr>
<td>( M. ) ( m. ) ( musculus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M. ) ( m. ) ( domesticus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome 19</td>
<td>1093</td>
<td>11</td>
<td>11</td>
<td>0.150</td>
<td>0.257</td>
<td>-1.46</td>
<td>0.109</td>
<td>0.742</td>
<td>2.833</td>
<td>0.092</td>
</tr>
<tr>
<td>( M. ) ( m. ) ( musculus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M. ) ( m. ) ( domesticus )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.**—\( P \) values <0.05 are indicated in bold.
  
1. \( L^* \) Length of the fragment in base pairs.
2. \( N^b \) Number of strain sequenced.
3. \( S^c \) Number of segregating sites.
4. \( \pi (%)^d \) Nucleotidic diversity.
5. \( \theta (%)^e \) Watterson's estimator of nucleotidic diversity.
6. \( D^f \) Tajima's \( D \).
7. \( \chi^2 \) Results of the HKA test using the region on chromosome 12 as neutral region of reference.
8. \( P \) Value Results of the HKA test using the region on chromosome 19 as neutral region of reference.
We tested if the excess of polymorphisms at the 3′ end of Oas1b was significantly different from the level of intraspecific variation expected under neutrality. To this end, an HKA test was performed using the 2 noncoding fragments on chromosomes 12 and 19 as neutral regions of reference (see Materials and Methods). We were able to reject the null hypothesis of neutrality for introns 4 and 5 and the 3′ UTR (bold in table 1), although not all comparisons yielded significant deviations from neutrality. In contrast, we cannot reject the neutrality hypothesis for introns 2 and 3 and for the region located 10 kb downstream of Oas1b.

We also examined the possibility that gene conversion (Pereluyt et al. 2006) could have affected the divergence of Oas1b alleles. Using GENECONV, we were unable to detect a single significant instance of gene conversion between Oas1b and any other member of the Oas1 gene family. Therefore, the large divergences between Oas1b intronic sequences result from the persistence of alleles over a long period of evolutionary time and not from the effect of gene conversion between paralogous copies. This analysis clearly indicates that Oas1b polymorphisms are ancient and strongly suggests that balancing selection has acted on Oas1b polymorphisms. It further suggests that the sites under balancing selection are located at the C-terminus of the Oas1b protein because the strongest evidence for transspecific polymorphism is found for introns 4 and 5 and for the 3′ UTR.

The high level of divergence of Oas1b alleles is also apparent at nonsynonymous sites (fig. 2B). The average divergence at nonsynonymous sites (dn) between M. m. musculus and M. m. domesticus is 2.00%. Very few genes show such a high dn and in our sample of 183 genes, only 5 genes (all involved in host–pathogens interactions) had a dn greater than 2.00%. This high level of dn is consistent with the exceptionally large number of amino acid polymorphisms at Oas1b. We then calculated the dn/ds ratio for all genes in our data set. A high value for the ratio dn/ds indicates that selection favors nonsynonymous substitutions, that is, amino acid changes, and is one of the predictions of balancing selection (Garrigan and Hedrick 2003). The dn/ds ratio is much higher for the Oas1b gene than it is for the vast majority of genes (fig. 2C). In general, the dn/ds ratio is very low (<0.1 and [Gibbs et al. 2004]) because most genes are evolving under strong purifying selection. In contrast, the dn/ds ratio is always higher than 0.4 for Oas1b and a number of pairwise comparisons are higher than 1. The Oas1b dn/ds ratio was also calculated between M. musculus and the other species in our sample (fig. 4), and values of dn/ds as high as 3.0 were obtained for some comparisons. Although it is plausible that an undetermined fraction of nonsynonymous changes resulted from a recent relaxation of selection on Oas1b (suggested by the presence of inactivating mutations in several strains) and caused an overestimation of dn, values of dn/ds higher than 1 are not consistent with neutral evolution. Instead, the high values of dn/ds suggest that selection in favor of nonsynonymous mutations is indeed acting on Oas1b and confirm the action of balancing selection on Oas1b polymorphisms.

The balancing selection hypothesis further predicts an excess of nonsynonymous polymorphisms within species. In M. m. domesticus and M. m. musculus, 65% and 82% of all polymorphisms are nonsynonymous, respectively (table 2). In contrast, the number of fixed nonsynonymous and synonymous differences between M. m. domesticus and either M. famulus or M. caroli are basically identical. For M. m. musculus, the difference is even more striking as the number of fixed synonymous differences is higher than the number of fixed nonsynonymous differences, for both comparisons. The MK test shows that the ratio of nonsynonymous to synonymous fixed difference differs significantly from the ratio of nonsynonymous to synonymous polymorphisms in M. m. musculus but not in M. m. domesticus, most likely because of the conservative nature of the

Table 2
MK Neutrality Tests

<table>
<thead>
<tr>
<th></th>
<th>Mus musculus</th>
<th>Mus caroli</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Synonymousa</td>
<td>Nonsynonymousb</td>
</tr>
<tr>
<td>Mus musculus musculus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polymorphic changes</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Fixed differences</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Mus musculus domesticus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polymorphic changes</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>Fixed differences</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

*Note.—Tests were performed with either M. famulus or M. caroli as outgroups.

a Synonymous (silent) mutations.
b Nonsynonymous (replacement) mutations.
Discussion

Our analysis revealed an extremely high level of variation at the 3' end of the Oas1b gene and provides strong evidence for transspecific polymorphism in the genus Mus. The nucleotide diversity in several introns of Oas1b is an order of magnitude higher than the diversity in putatively neutral regions of the mouse genome (the regions on chromosomes 12 and 19 and the 44 regions analyzed in Zhang et al. [2005]). The phylogenetic analysis of Oas1b introns further indicates that the allelic lineages found in M. musculus are older than the split between M. musculus and M. famulus, which occurred 2.8 MYA. Thus, this polymorphism is the second oldest reported in mice, after the genes of the MHC (Edwards et al. 1997). Although the presence of such divergent allelic lineages within a species could result from incomplete lineage sorting, this is very unlikely because the maintenance of alleles over 2.8 Myr would require an unrealistically large long-term effective population size. Assuming a conservative generation time of 2 generations/year, a long-term population size of more than one million individuals would be necessary to maintain the level of variation observed at the 3' end of Oas1b (using θ = 4Nₑμ, where θ is the nucleotide diversity, μ the mutation rate per site per generation, and Nₑ the effective population size). This value of Nₑ is 3–10 times higher than previous estimates (Eyre-Walker et al. 2002) and it is extremely unlikely that such a large effective population size was maintained for more than 2.8 Myr, considering that house mice went through several episodes of speciation that are typically associated with population bottlenecks. It is also unlikely that a recent hybridization event could account for the presence of divergent lineages in M. musculus. Although M. m. domesticus occasionally interbreed with M. spreitus under natural conditions (Greene-Till et al. 2000; Orth et al. 2002), the molecular divergence between these 2 species is much smaller than the divergence between Oas1b allelic lineages and experimental crosses between M. musculus and its Asian relatives fail to produce viable birth. Instead, our data strongly suggest that Oas1b polymorphisms have been maintained in populations for the last ~3 Myr by the action of balancing selection. We provide several lines of evidence that strongly support this hypothesis, including an extremely high nucleotide diversity in introns 4 and 5 and in the 3' UTR, a significant deviation from neutrality as revealed by the HKA test, a high dn/ds ratio and an excess of nonsynonymous polymorphisms. Balancing selection has been shown to act on a small number of loci in mammals (Hedrick and Thomson 1983; Figueroa et al. 1988; Shyue et al. 1995; Bamshad et al. 2002; Verrelli et al. 2002; Newman et al. 2006; Baysal et al. 2007) but rarely results in transspecific polymorphisms, probably because hosts eventually evolve resistance alleles that become fixed (Hedrick 2004). To our knowledge, there are only 2 well-documented cases of long-term (i.e., transspecific) balanced polymorphisms at host defense genes in mammal, the MHC (Figueroa et al. 1988; Edwards et al. 1997; Gutierrez-Espeleta et al. 2001) and the TRIM5α gene in Old World monkeys (Newman et al. 2006), making Oas1b only the third known case.

There are 3 main categories of mechanisms that can maintain allelic diversity for long periods of time: frequency-dependent selection, heterozygote advantage, and selection that varies in time and space. At this point, it is not possible to determine with certainty the mechanism responsible for the long-term maintenance of Oas1b alleles and more data about Oas1b's allelic diversity in natural populations are needed. However, our data point to some type of frequency-dependent mechanism. Under this model, the resistance allele that protects against the most common pathogenic genotypes will also be the most common one in the host population. In this situation, rare flaviviral genotypes that evade host defense will be advantaged and increase in frequency. Because hosts carrying rare alleles have in turn the highest fitness, the frequency of the rare allele will increase and the frequency of the common allele will decrease leading to a dynamic polymorphism. The dn/ds ratios were plotted against their ds values, which is a proxy of the divergence time between alleles (fig. 4). Less divergent (younger) alleles have higher dn/ds than older ones. This suggests that balancing selection has favored rare alleles over more common ones, assuming that young alleles are, in general, at lower frequency in populations than older alleles. A similar relationship is also expected if nonsynonymous sites become saturated faster than synonymous sites. However, this seems unlikely because most values of dn are lower than 3% and saturation is not expected within this range.

Whatever the mechanism involved in the maintenance of Oas1b alleles, a functional difference between the Oas1b alleles under balancing selection is implicit. Balancing selection also implies some type of interactions, direct or indirect between Oas1b and flaviviruses. The 2 alleles that have been functionally characterized (the major and minor resistance alleles) differ in the specificity of their response to flaviviruses, the minor resistance allele protecting against a smaller diversity of flaviviral genotypes than the major resistance allele. The amino acid polymorphisms responsible for these functional differences have yet to be mapped to specific regions of Oas1b as these 2 alleles differ at a number of sites spread across the entire gene. Although Oas1b retained a number of functional features found in other OAS, it has lost its synthetase activity and seems to protect against flaviviruses by a still unknown RNase L–independent pathway (Scherbik et al. 2006). As the double-strand RNA-binding motifs of Oas1b are located at the N-terminus of the protein, balancing selection is probably not caused by the interactions between flaviviral RNA and the Oas1b protein. The C-terminus of the Oas1b protein contains a domain with strong sequence similarity to an extracellular domain of interleukin receptors 3 and 5 (Ferguson W, Boissinot S, unpublished observation). This domain encompasses the CFK motif responsible for Oas1b tetramerization and plays an important role in protein–protein binding in the interleukin receptor family. This domain is in the immediate...
vicinity of a number of polymorphic sites including one of the polymorphisms shared between the different subspecies of *M. musculus* and between *M. musculus* and *M. spicilegus*. As this region has the potential to mediate protein–protein interactions, it is plausible that interactions between the *Oas1b* protein and a flaviviral protein, or a host protein which is itself under balancing selection, are responsible for the pattern of balancing selection at *Oas1b*, although this hypothesis needs to be tested experimentally.

**Acknowledgments**

The work was conducted in part with equipment from the Core Facilities for Imaging, Cellular and Molecular Biology at Queens College. We thank Else Fjerdingstad, George Jackman, and Peter Novick for their helpful comments on the manuscript. We thank Francois Bonhomme and the “Conservatoire de la souris sauvage” for providing DNA samples. This research was supported by National Institutes of Health grant AI058968-01A1 and PSC-CUNY grant 66253-0035 to S.B.

**Literature Cited**


Yoko Satta, Associate Editor

Accepted April 30, 2008