An Analysis of Signatures of Selective Sweeps in Natural Populations of the House Mouse

Sonja Ihle, Iary Ravaoarimanana,† Meike Thomas, and Diethard Tautz
Institut für Genetik der Universität zu Köln, Köln, Germany

Population and locus-specific reduction of variability of polymorphic loci could be an indication of positive selection at a linked site (selective sweep) and therefore point toward genes that have been involved in recent adaptations. Analysis of microsatellite variability offers a way to identify such regions and to ask whether they occur more often than expected by chance. We studied four populations of the house mouse (Mus musculus) to assess the frequency of such signatures of selective sweeps under natural conditions. Three samples represent the subspecies Mus mus domesticus and came from Germany, France, and Cameroon. One sample came from Kazakhstan and constitutes a population of the subspecies Mus musculus. Mitochondrial D-loop sequences from all animals confirm their respective assignments. Approximately 200 microsatellite loci were typed for up to 60 unrelated individuals from each population and evaluated for signs of selective sweeps on the basis of Schlötterer’s In RV and In RH statistics. Our data suggest that there are slightly more signs of selective sweeps than would have been expected by chance alone in each of the populations and also highlights some of the statistical challenges faced in genome scans for detecting selection. Single-nucleotide polymorphism typing of one sweep signature in the M. m. domesticus populations around the β-defensin 6 locus confirms a lowered nucleotide diversity in this region and limits the potential sweep region to about 20 kb. However, no amino acid exchange has occurred in the coding region when compared to M. m. musculus. If this sweep signature is due to a recent adaptation, it is expected that a regulatory change would have caused it. Our data provide a framework for conducting a systematic whole genome scan for signatures of selective sweeps in the mouse genome.

Introduction

Understanding the genetic basis of adaptations remains one of the largest challenges in current biological research. Many aspects of the phenotype can be the target of selection, including morphology, physiology, immune response, or traits relevant for reproduction. Mapping of quantitative trait loci is one way to approach this problem but requires an a priori knowledge about the phenotypic characteristics of the trait to be studied. Genome scans for detecting signatures of selective sweeps in natural populations have been proposed as a phenotype independent approach to identify adaptive trait loci (reviewed in Schlötterer 2003). The basic assumption is that a locus that has been under positive selection will lead to a loss of neutral polymorphism in the closely linked region of the genome (Maynard Smith and Haigh 1974). Such reduced polymorphism can be detected by typing neutral markers in different populations. This approach has been applied to populations from Drosophila (Schlötterer, Vogl, and Tautz 1997; Harr, Kaiser, and Schlötterer 2002; Kaiser, Dieringer, and Schlötterer 2003; Schöff and Schlötterer 2004), humans (Diller, Gilbert, and Kocher 2002; Paysier, Cutter, and Nachman 2002; Bamshad and Wooding 2003; Kayser, Brauer, and Stoneking 2003; Storz, Paysier, and Nachman 2004), maize (Vigouroux et al. 2002), and Plasmodium (Wootton et al. 2002).

We have chosen here the house mouse (Mus musculus) as a system for studying this question because its evolutionary history is well known (Boursot et al. 1993; Guenet and Bonhomme 2003) and because the excellent genetic and genomic resources for the mouse will make it possible to study candidate loci further.

Mus musculus originated on the Indian subcontinent and the subspecies Mus m. musculus and Mus m. domesticus split less than 1 MYA. Mus m. musculus has colonized much of Asia and Eastern Europe, while M. m. domesticus has invaded the Near East and from there Western Europe less than 4,000 years ago (Cucchi, Vigne, and Auffrey 2005). A hybrid zone between the subspecies exists roughly from Denmark along the eastern part of Germany into the Balkan. For our study, we have chosen one population from Western Germany and one from Southern France, which represent pure M. m. domesticus populations. In the past few hundred years, M. m. domesticus has spread further across the rest of the world. We have sampled a population from Africa (Cameroon) as a representative of such a newly established lineage. As an outgroup, we have taken samples from a population in Kazakhstan, which represents M. m. musculus close to its area of origin.

Our results suggest that genomic regions of reduced variability occur more often than expected by sampling error, implying that these are true signatures of selective sweeps. However, it is also clear that larger numbers of loci will need to be screened before this can be statistically confirmed.

Methods

Samples

Mouse populations were sampled in houses, barns, or stables, and care was taken that only single animals were used from any given sampling site. The different sampling sites were at least 300 m apart from each other but often many kilometers (see Supplementary Table 1a, Supplementary Material online). Additional animals for outgroup comparisons were provided by Francois Bonhomme and Pierre Boursot from the wild mice breeding center “Conservatoire
genetique de souris sauvage’ in Montpellier and by Pavel Munclinger (Czech Republic).

Isolation of DNA and Preparation for Genotyping

For DNA isolation mice were dissected in the field. The spleen was directly dissolved in 5 ml of sodium dodecyl sulfate (SDS) buffer (80 mM ethylendiaminetetraacetic acid (EDTA) 0.5% SDS, 100 mM Tris pH 8). Liver, kidneys, heart, lung, and testis were stored in 100% ethanol as tissue backup. DNA was isolated by standard salt extraction procedures or by using the extraction kit NucleoSpin Blood XL (Macherey–Nagel, Düren, Germany). Dried DNA pellets were resolved in TE (10 mM Tris pH 8, 0.1 mM EDTA). The DNA was stored at −20°C. For genotyping, the DNA was diluted to 5 ng/µl and stored in 96 well-plates for further processing.

D-Loop Sequencing

D-loop sequences were obtained for all samples used for genotyping, employing the following primers (Prager et al. 1993): forward: 5’-CAT TAC TCT GGT CTT GTA AAC C; reverse: 5’-GCC AGG ACC AAA CCT TTG TGT. Fifty µl polymerase chain reactions (PCR) were performed by the following standard PCR conditions: final concentration of 0.5 ng/µl DNA template, 0.06 µmol/µl deoxynucleoside triphosphate, 0.045 µmol/µl MgCl₂, 0.1 µl/µl of 10× PCR buffer, 0.4 pmol/µl of forward and reverse primers, and 0.04 U/µl of Taq polymerase. For amplification, the DNA template was denatured for 5 min at 94°C followed by 35 cycles of 45 s at 94°C denaturation, 45 s at 59°C annealing, 1 min and 30 s at 72°C extension, and a final extension step of 5 min at 72°C. PCR products were purified using ultra-free filters from Millipore according to the manufacturers protocol. Products were rediluted in 10 mM Tris (pH 8), and 100 ng PCR product was directly added to a 10-µl cycle sequencing reaction using 3 µl of ET-Terminator Ready-Reaction Mix (Amersham Biosciences, Buckinghamshire, UK) and 5 pmol of either the reverse or the forward primer. Cycle sequencing was performed with 35 cycles with 20 s at 95°C, 15 s at 58°C, and 1 min at 60°C. Sephadex G-50 columns were used for cleanup of the reaction. Sequencing reactions were run on a MegaBace 1000 capillary sequencer (Amersham–Molecular Dynamics, Buckinghamshire, UK), and the data were transferred into an Excel spreadsheet for further analysis with Arlequin (Schneider, Roessli, and Excoffier 2000). General gene diversity estimates were calculated by using the microsatellite toolkit (Park 2001) and the program MS analyser (Dieringer and Schlo¨tterer 2003). In RV (Schlo¨tterer 2002) and ln RH (Kauer, Dieringer, and Schlo¨tterer 2003) statistics were used to identify loci with significant signatures of selective sweeps.

A subset of loci was chosen from the vicinity of genes which could be considered as candidate loci for positive selection because of their assumed involvement in innate immunity, resistance, or hybrid sterility. The genes were chosen from the databases Pubmed, Locuslink, and Nucleotide database (http://www.ncbi.nlm.nih.gov/) by initial screening with the following keywords: “interferon,” “resistance,” “speciation,” “behaviour,” “imprinting,” “olfaction,” and “saliva.” Genes with a suitable microsatellite in their vicinity or the introvn were then further used (see Supplementary Table 1b and c [Supplementary Material online] for a list of all loci studied).

Analysis of Nucleotide Diversity

The genomic sequence of the region representing the β-defensin 6 from chromosome 8 locus was obtained from the ENSEMBL database. Sequencing primers from nonrepetitive regions were designed to amplify stretches of 500–1,000 bp from the region. About 4,000 bp of the genomic region around the two β-defensin 6 exons was contiguously sequenced. The fragments were amplified and directly sequenced from 28 animals: five each from Kazakhstan, Czech Republic, Cameroon, France, and Germany and three animals from the United States. For outgroup comparisons, four animals from M. spretus were sequenced for each fragment. Heterozygous positions were identified through manual inspection of all sequence positions for which the automatic nucleotide calls were ambiguous. The sequences were aligned with the program Seqman of the Lasergene expert sequence analysis software (DNASTAR, Inc., Madison, Wisc.) and imported into the program DNASP 3.51.
(J. Rozas and R. Rozas 1999). To deal with heterozygous positions, all sequences were duplicated, and each nucleotide state was randomly assigned to one of the duplicated sequences; that is, no haplotype information was recovered. Gene diversity and the number of polymorphic sites were calculated separately for the two lineages *M. musculus* and *M. domesticus* with the program DNASP 3.51. By comparing all polymorphic sites against the *M. spretus* sequences, the ancestral or derived state for each polymorphism was inferred.

**Results**

Our sampling regime was designed to obtain a representative sample of each of the local populations in the study. Because mice are known to live in extended family groups (Berry and Bronson 1992), we have used only single individuals from each sampling site and took care that sampling sites were at least 300 m apart of each other. All sampling sites were within a radius of at most 50 km at the respective locality. Thus, we can assume that all individuals of this study are unrelated to each other and represent local populations. A total of 204 microsatellite loci were typed for 20–60 individuals from each of these populations. The loci were chosen in two different ways. One set of loci ($N = 76$) was chosen from the vicinity of genes which could be considered as possible candidates for frequent adaptations, such as genes involved in innate immunity (i.e., inducible by interferon) or hybrid sterility (see Methods for a full set of criteria). The second set of loci ($N = 128$) was taken from the mouse mapping panel, whereby we focused on chromosomes 1, 9, and 15. These loci can be considered to represent a random sample with respect to their location in the vicinity to genes. Of all loci typed, 18 had to be discarded because they were either too difficult to type or because they had too few ($<4$) alleles.

**Population Structure**

Because the verification of the assumptions of population structure is important for the assessment of the results from the polymorphism analysis, we have obtained mitochondrial D-loop sequences for all individuals used in the study. Figure 1a shows the respective tree, including several outgroups. This confirms the clear separation of *M. m. domesticus* and *M. m. musculus* populations and shows that the D-loop lineages found in Cameroon are more closely related to lineages found in the German than in the French population. However, in Germany, we also find a second group of haplotypes that is basal to the French haplotypes. This could be due to either a loss of haplotypes in the French population or an admixture or subdivision in the German population. Figure 1b shows the allele sharing tree that is based on all microsatellite genotypes. This tree is expected to better reflect the current patterns of gene flow. It shows a clear separation of the populations with no overlap. There is also no indication of substructure in the German population. We assume, therefore, that the two types of mitochondrial haplotypes that we find in the German sample are remnants of the colonization process and that the population is currently homogeneous and interbreeding.

Table 1 lists some basic population genetic parameters. Expected heterozygosities are lowest for the Cameroon population, consistent with the history of a recent colonization. The number of alleles is highest in the Kazakhstan population, which is in line with the assumption of a more ancient population. Intriguingly, the observed heterozygosities

![Fig. 1.](https://academic.oup.com/mbe/article-abstract/23/4/790/1008112)
are always lower than the expected, that is, there is an excess of homozygous individuals. We interpret this as a consequence of the social system of mice where inbreeding within nests is not rare (Berry and Bronson 1992). However, because we use only one mouse per nest, we can still assume that our population sample is not biased by this effect.

Detection of Sweep Signatures

To assess the number of signatures of selective sweeps in each of the populations, we employed Schlötterer’s \( \ln \text{RV} \) and \( \ln \text{RH} \) statistics, which have been shown to provide a reliable estimator that is largely independent of population size differences, bottlenecks, and mutation rates (Schlötterer 2002; Kauer, Dieringer, and Schlötterer 2003; Schöfl and Schlötterer 2004). \( \ln \text{RV} \) measures \( \log \) of the ratios of the variances of repeat numbers for each locus between two populations, and \( \ln \text{RH} \) measures the \( \log \) of the ratios of heterozygosities (Kauer, Dieringer, and Schlötterer 2003). This results in an approximately normal distribution of values for which the mean and the variance are calculated. Loci outside of the chosen cutoff value (here 5% and 1%) are candidates for selective sweeps.

We computed six pairwise comparisons of the \( \ln \text{RV} \) and \( \ln \text{RH} \) statistics for the four populations studied. In addition, we compared whether we find different results when the set of random loci was treated separately from the candidate loci or when all loci were analyzed jointly. For each pairwise comparison, we tested whether the distribution of \( \ln \text{RV} \) or \( \ln \text{RH} \) values was significantly different from the expected normal distribution. This was not the case in any of the tests (Kolmogorov-Smirnov test—data not shown). This implies that there is not a large number of values that fall outside the distribution, that is, we can employ the respective distributions to apply significance tests.

Supplementary Table 2 lists all loci identified by the different test criteria. Table 2 provides the summary of the number of loci found. Two general conclusions can be drawn from this table. First, there is no obvious difference in the numbers found for the candidate loci versus the random loci. Second, the total number of loci found is slightly, but not significantly, above the number that would be expected. This is particularly clear when one looks at the whole sample of all loci. For each statistic, we should have expected to find 9.3 or 1.9 loci at the 5% and 1% cutoff level, respectively. Instead, we find on average (over both statistics) 10.8 and 3.8, respectively. Although this difference is not statistically significant, it is nonetheless a consistent pattern in all six pairwise comparisons.

The different statistics and the different pairwise comparisons identify only partially overlapping sets of loci (see Supplementary Table 2, Supplementary Material online). This is not unexpected, given the fact that most

<table>
<thead>
<tr>
<th>Population</th>
<th>Animals Analysed</th>
<th>Observed Average Heterozygosity</th>
<th>Expected Average Heterozygosity</th>
<th>Average Number of Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cameroon</td>
<td>68</td>
<td>0.45</td>
<td>0.60</td>
<td>6.9</td>
</tr>
<tr>
<td>France</td>
<td>66</td>
<td>0.53</td>
<td>0.72</td>
<td>10.4</td>
</tr>
<tr>
<td>Germany</td>
<td>57</td>
<td>0.49</td>
<td>0.73</td>
<td>10.3</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>60</td>
<td>0.60</td>
<td>0.75</td>
<td>12.3</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Kaz vs. Fra</th>
<th>Kaz vs. Ger</th>
<th>Kaz vs. Cam</th>
<th>Fra vs. Ger</th>
<th>Fra vs. Cam</th>
<th>Ger vs. Cam</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P &lt; 0.05 )</td>
<td>( \ln \text{RV} )</td>
<td>( \ln \text{RH} )</td>
<td>( \ln \text{RV} )</td>
<td>( \ln \text{RH} )</td>
<td>( \ln \text{RV} )</td>
<td>( \ln \text{RH} )</td>
<td>( \ln \text{RV} )</td>
</tr>
<tr>
<td>Candidate</td>
<td>67</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Random</td>
<td>119</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>All</td>
<td>186</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>13</td>
<td>13</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( P &lt; 0.01 )</th>
<th>( \ln \text{RV} )</th>
<th>( \ln \text{RH} )</th>
<th>( \ln \text{RV} )</th>
<th>( \ln \text{RH} )</th>
<th>( \ln \text{RV} )</th>
<th>( \ln \text{RH} )</th>
<th>( \ln \text{RV} )</th>
<th>( \ln \text{RH} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate</td>
<td>67</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Random</td>
<td>119</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>All</td>
<td>186</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

NOTE.—Kaz, Kazakhstan; Fra, France; Ger, Germany; Cam, Cameroon.

* “Candidate” and “random” refers to the separate analysis of the loci in the respective classes (see Supplementary Table 2a, Supplementary Material online).

* “All” refers to the calculation where all loci were analyzed jointly (see Supplementary Table 2b, Supplementary Material online)—the calculation for “all” is not simply a sum of candidate and random because the variance estimates for each calculation are derived from the respective pools of the loci in each class, which can lead to shifts of significance score for some loci.
of the regions of reduced variability are attributable to sampling effects and because the different statistics measure somewhat different parameters. Accordingly, it is not directly possible to identify good candidate loci for true selective sweeps. However, one can focus on loci that show a particularly strong and consistent signal. We chose one such locus for further analysis.

**Analysis of a Candidate Locus**

Locus P104 lies close to the β-defensin 6 gene, which has an important role in microbial defence in the mouse. The locus shows a strong sweep signature for the *M. m. domesticus* populations from France and Germany in all respective pairwise comparisons and for both statistics. The region upstream and downstream of this locus was scanned for singe-nucleotide polymorphisms (SNPs), using animals from each of the surveyed populations, plus an additional *M. m. musculus* population from the Czech Republic and *M. spretus* samples as outgroup. Comparisons focussed on *M. m. domesticus* versus *M. m. musculus* because the pattern suggested that this sweep would have occurred after the separation of these subspecies.

To obtain the SNP data, a total of 20 fragments scattered along a region of 1.5 Mb were sequenced in each of about 56 chromosomes from the *M. m. musculus* and *M. m. domesticus* animals, as well as from eight chromosomes from *M. spretus* for outgroup comparisons. The distribution along the chromosome, the length of the fragments, and the diversity estimates per fragment and per lineage are listed in Supplementary Table 3 (Supplementary Material online) and graphically displayed in figure 2.

The *M. m. domesticus* samples show a somewhat lower nucleotide diversity in the region from 300 kb upstream to 45 kb downstream of the β-defensin 6 gene. Outside of this region, as well as at single fragments within the region, a higher nucleotide diversity is found in *M. m. domesticus*, indicating that the lower nucleotide diversity is not a general feature of the *M. m. domesticus* samples. The largest difference in nucleotide diversity between the subspecies is found from 10 kb upstream to 10 kb downstream of the β-defensin 6 gene. The fragments approximately 100 kb downstream show an unusually high nucleotide diversity and a very high number of heterozygous individuals. It seems possible that there is a hidden duplication in this area, which causes this effect, but we have not further explored this and do not discuss this further.

Applying Tajima’s D (Tajima 1989) and the H-test of Fay and Wu (Fay and Wu 2000) to these data does not yield significant results for most of the fragments analyzed (Supplementary Table 4, Supplementary Material online). We have also used the multilocus HKA-test (Hudson, Kreitman, and Aguadé 1987) developed by

---

**Fig. 2.—**Nucleotide diversity calculated per fragment. *(a)* *Mus m. musculus*, *(b)* *Mus m. domesticus*, and *(c)* the difference in nucleotide diversity for each locus. The location of β-defensin exons and other annotated potential genes in this chromosomal region are depicted at the bottom.
Hey (http://lifesci.rutgers.edu/~heylab), which is expected to deal much better with such situations. The overall results show a significant deviation from the expectation of neutrality in the comparison of *M. spretus* to *M. m. domesticus* ($\chi^2 = 71.7$; df = 19; $P < 0.001$) but not in the comparison between *M. m. musculus* and *M. spretus* ($\chi^2 = 21.1$; df = 19; $P = 0.33$). The results for the individual fragments are listed in Supplementary Table 5 (Supplementary Material online). For *M. m. domesticus*, the largest deviations are found for β-defensin 6 exon 1 fragment and the fragments 100 kb downstream. However, the latter exhibit an excess of polymorphism, rather than a loss, caused by high numbers of nucleotide exchanges at low frequency. Because the frequencies themselves are not considered in the HKA-test, the exclusion of these fragments from the analysis does not change the overall test statistic (data not shown).

β-Defensin 6 codes only for a short peptide of 63 amino acids in length. No polymorphisms or exchanges were found in this coding region between *M. m. domesticus* and *M. m. musculus*, but several replacements and polymorphisms were found in the *M. spretus* samples (fig. 3). The Ka/Ks ratios of the *M. m. domesticus* lineage against the three *M. spretus* sequences range from 0.7 to 1.8. This is in line with the notion that positive selection acts frequently on this gene (Maxwell, Morrison, and Dorin 2003). It is therefore surprising that we find no coding substitutions in the *M. musculus* subspecies. However, there are several fixed differences in the upstream region in *M. m. domesticus*. Preliminary experiments suggest indeed that there are differences in expression levels between the subspecies (Ihle 2004). Thus, if positive selection has acted on this locus, it would have been mediated via a regulatory change and not via a protein sequence change. Still, it should be emphasized that we have currently no independent criteria that positive selection has indeed acted at this locus.

Fig. 3.—Alignments of the coding region of the β-defensin 6 gene from the *Mus m. domesticus* and *Mus m. musculus* samples, as well as from three alleles from *Mus spretus*. Note that all alleles of the former were identical. Upper part shows the nucleotide alignment, the lower part the aminoacid alignment. Identical positions are indicated by dashes. Note that the third allele of *M. spretus* has a mutation in the ATG start codon, that is, it is probably nonfunctional.
signatures under the ln RV or ln RH statistics, we would expect to see fewer signatures in Kazakhstan and more in Cameroon. This is not the case (table 2). There are no obvious differences with respect to the number of sweep signatures detected for the four populations. This finding is reassuring because it confirms the results from the simulation runs which suggested that the ln RV statistic is relatively refractory to population size differences (Schlötterer 2002).

Another important question with this approach concerns the appropriate null distributions of ln RH and ln RV values. Ideally, one should use a set of loci that can be considered to evolve under neutral conditions. However, there are no a priori criteria for choosing such loci with confidence. We have therefore taken the opposite approach, namely, to choose loci that are candidates for positive selection and compared these to randomly chosen loci. We find that the probability distributions for both sets of loci are not significantly different from a normal distribution. Accordingly, there are only minor differences with respect to the numbers of outliers detected when one uses the distributions separately or jointly (see table 2). We conclude from this that most of the patterns detected in this study are likely to be due to random fluctuations. An excess of about 1% of the loci tested in pairwise comparisons appear to fall in the tails of the distributions.

Given the small number of possible true selective sweeps among the much larger number of false positives, it is important to find additional criteria of how the true cases can be identified. Schlötterer (2002) has suggested that signatures that are found in at least two population comparisons and/or with both statistics might be considered to be more reliable. The β-defensin 6 locus fulfills these criteria. It shows highly significant ($P < 0.001$) ln RV and ln RH values in pairwise comparisons of the French and German populations with respect to the Kazakhstan population. The analysis of nucleotide diversity around this locus shows a significantly lowered diversity for the M. m. domesticus populations compared to M. m. musculus populations. Hence, the SNP data confirm the sweep signature and limit it to a window of about 20 kb. However, we currently have only limited reference data that would allow us to assess whether this loss of nucleotide diversity is within the sampling error or not. Within the 1.5 Mb window that we screened, we did not find another region with similarly large differences in nucleotide diversity (ignoring the unusual pattern of the fragments at the right end of the window), although fluctuations are evident. Still, more reference data will be required before the statistical significance of the finding at the β-defensin 6 locus can be properly assessed.

In a previous study of assessing nucleotide diversity in natural population samples from M. m. domesticus, Nachman (1997) found an average $\pi = 0.00078$, which is about three times lower than our average of $\pi = 0.0025$ for the M. m. domesticus populations (see Supplementary Table 3, Supplementary Material online). However, the studies differ in several respects. Nachman (1997) focussed on short introns from genes on the X chromosome, while we studied mainly random fragments in nongenic regions on an autosome. Furthermore, our samples came from three separate regions (Germany, France and Cameroon), while Nachman sampled only in Italy. Using the same population sample that we have studied here, Tina Harr and John Baines have surveyed several random fragments from various chromosomes for nucleotide diversity and found a similar average $\pi$ for M. m. domesticus populations to that which we obtained (personal communication).

There is another interesting observation in our data set which is of relevance for whole genome screens. Typing of additional microsatellite loci around the β-defensin 6 locus did not reveal any that showed a sweep signature (data not shown). This includes one locus only 4 kb upstream, that is, in the region with reduced nucleotide diversity. However, it turns out that this locus has a greater number of repeats and a larger number of alleles than the P104 locus that shows the sweep signature. It is known that short microsatellite alleles have lower mutation rates (Schlötterer et al. 1998) and would therefore retain the signature of a sweep for a longer period of time. Loci with longer alleles, on the other hand, would quickly recover their variability. This observation suggests that sweep signatures that are detected with loci with greater numbers of repeats would reflect relatively recent events, while older events could only be traced with loci of shorter length (i.e., reduced mutation rates). This is in line with the theoretical expectations proposed by Wiehe (1998).

### Supplementary Material

Supplementary Tables 1–5 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

### Acknowledgments

We are particularly grateful to Francois Bonhomme and Pierre Boursot for their support in initiating this work, as well as Barbara Dod, Louis Rouquet, Paul Pfänder, Jaroslav Pialek, Pavel Munclinger, Barbora Bimova, Eva Bozikova, Solo Robespierre, and Uli Schliewen for their help with obtaining the wild-type samples. We thank Christian Schlötterer, Bettina Harr, Thomas Wiehe, John Baines, and Wolfgang Stephan for discussions and suggestions and Trevor Price for suggestions on the manuscript. This study was supported by a grant of the Volkswagen Stiftung and through contributions from the Verband der Chemischen Industrie and the International Graduate School in Genetics and Functional Genomics in Cologne.

### Literature Cited


