High Mutation Rate of a Long Microsatellite Allele in *Drosophila melanogaster* Provides Evidence for Allele-Specific Mutation Rates

C. Schlötterer, R. Ritter, B. Harr, and G. Brem
Institut für Tierzucht und Genetik, Vienna, Austria

Within recent years, microsatellite have become one of the most powerful genetic markers in biology. For several mammalian species, microsatellite mutation rates have been estimated on the order of $10^{-3}$–$10^{-5}$. A recent study, however, demonstrated mutation rates in *Drosophila melanogaster* of at least one order of magnitude lower than those in mammals. To further test this result, we examined mutation rates of different microsatellite loci using a larger sample size. We screened 24 microsatellite loci in 119 *D. melanogaster* lines maintained for approximately 250 generations and detected 9 microsatellite mutations. The average mutation rate of $6.3 \times 10^{-6}$ is identical to the mutation rate from a previous study. Most interestingly, all nine mutations occurred at the same allele of one locus (*DROYANETS*). This hypermutable allele has 28 dinucleotide repeats and is among the longest microsatellite reported in *D. melanogaster*. The allele-specific mutation rate of $3.0 \times 10^{-4}$ per generation is within the range of mammalian mutation rates. Future microsatellite analyses will have to account for the dramatic differences in allele-specific mutation rates.

**Introduction**

Microsatellites are highly polymorphic DNA regions that have become the marker of choice in several fields of biology, such as behavioral ecology (Schlötterer and Pemberton 1998), linkage analysis (Dib et al. 1996; Dietrich et al. 1996), and phylogeny reconstruction (Bowcock et al. 1994). Despite their widespread use, very little is known about the evolution of microsatellites. Although it is generally accepted that DNA slippage is the predominant mutation mechanism creating new microsatellite alleles (Tautz and Schlötterer 1994), the exact mutation patterns generated through this mechanism are not yet known. The simplest model assumes that the size of a microsatellite changes by only a single repeat unit. Some work on *Escherichia coli* (Levinson and Gutman 1987a) and yeast (Henderson and Petes 1992), as well as the analysis of natural variation in human populations (Di Rienzo et al. 1994), suggests that mutations of more than a single repeat unit also occur. Among other problems, this uncertainty about the mutation patterns makes it difficult to obtain good mutation rate estimates from population data. Thus, direct observations of mutations are preferable. The scoring of microsatellites for human pedigrees has identified some microsatellite mutations. However, given the substantial typing effort needed, this approach cannot be transferred to many other organisms. An alternative strategy would be to use inbred lines which have been propagated for several generations. This would reduce the typing effort and should still result in a reliable mutation rate estimate.

Recently, a study using inbred *Drosophila melanogaster* lines reported microsatellite mutation rates of $6.3 \times 10^{-6}$, which is at least one order of magnitude lower than rates for mammals (Schug, Mackay, and Aquadro 1997). At present, it is not clear whether the reported low mutation rates in *D. melanogaster* are specific to the lines used by Schug, Mackay, and Aquadro (1997). To test this, we used different lines and a set of 24 microsatellites which share only one locus with the microsatellites used by Schug, Mackay, and Aquadro (1997).

**Materials and Methods**

**Fly Stocks**

The lines used in this study were originally generated by mutagenesis of a second chromosome carrying the phenotypic markers *cn*, *bw*, and *sp*. Before mutagenesis, the chromosome had been isogenized; thus, all mutants are derived from the same chromosome (Nüsslein-Volhard, Wieschaus, and Kluding 1984). The chromosomes carrying a recessive embryonic lethal mutation were maintained over a *CyO* balancer chromosome. This balancer carries multiple inversions to prevent recombination. As individuals homozygous for one of the chromosomes are not viable, the flies were maintained in a heterozygous state. The 123 independent lines with zygotic mutations on the second chromosome that were scored for microsatellite mutations were obtained from the UMEA stock center, the Nüsslein-Volhard laboratory, and the Bloomington stock center. Wild-type flies were from Kenya (25 chromosomes), Hungary (60 chromosomes), and the Netherlands (40 chromosomes). The African flies were obtained from the Species Stock Center, where they had been maintained as isofemale lines. We therefore accounted for the effects of inbreeding by randomly selecting one chromosome from heterozygous individuals. Both chromosomes were analyzed in the populations from Hungary and the Netherlands; we screened the F1 generation from freshly collected flies.

**Microsatellite Analysis**

DNA was extracted from single individuals by a high-salt procedure (Miller, Dykes, and Polesky 1988). Twenty-four microsatellites were amplified by PCR from 50 ng genomic DNA in 10-μl volumes using γ-$^{32}$P end-labeled primers, 1.5 mM MgCl₂, 200 μM dNTPs, 1 μM of each primer, and 0.5 U *Taq* polymerase.
(Schlötterer 1998b). Cycling conditions were 30 cycles of 1 min at 94°C, 1 min at 49–57°C (depending on the primer combination; see table 1), and 1 min at 72°C. PCR products were separated on a 7% denaturing polyacrylamide gel (32% formamide, 5.6 M urea) at 60–80 W. Mutations were easily identified by comparison with the other lines which have the identical genotype, because they were derived from the same inbred line (Nüsslein-Volhard, Wieschaus, and Kluding 1984). Sequences of DROYANETS alleles were determined by cloning and subsequent sequencing of PCR products.

Mutation Rate Calculations

Average mutation rates were calculated as \( m/(2GLK) \), where \( n \) = number of mutations, \( G = \) number of generations, \( L = \) number of lines, and \( K = \) number of loci. The factor two was used to account for the fact that two independent second chromosomes were surveyed in each line. To test whether the observed distribution of mutations is compatible with equal mutation rates for all loci, we used the binomial distribution and calculated the probability of only one locus having mutated given the average mutation rate per line \( (m/L) \). The probability of all nine mutations occurring on the same allele was also calculated using the binomial distribution. Upper confidence limits for the mutation rates of the microsatellite loci with no mutations were calculated from a Poisson distribution, with the probability of observing exactly zero mutations equal to \( e^{-m} \). Solving for \( m \) such that the probability of zero mutations is 0.05 gives \( m = 2.996 \). The upper confidence limit was obtained by \( m/(2GLK) \). The 95% confidence intervals for the mutation rates averaged over all loci were also calculated by assuming a Poisson distribution. The values for \( m \) corresponding to nine mutations were obtained from Rohlf and Sokal (1995) and divided by 2GLK.

Exclusion of Contaminants

In total, four lines which had 3–10 mutations of a microsatellite allele were observed. Because of the low mutation rate of Drosophila microsatellites, they were treated as suspected contaminants and excluded from further analysis. For three lines, we had additional evidence for contamination. Phenotypic inspection verified two lines as contaminants (S. Luschnik, personal communication). For the third line, we compared the genotypes in an independent replicate and could not confirm the four mutations observed. As the two replicate lines were maintained independently for only 1 year, this result corroborates the suspected contamination.

Results

We screened 24 microsatellites for mutations in 119 lines which were generated in September 1979 (C. Nüsslein-Volhard, personal communication) during a large mutagenesis screen for embryonic lethal mutations (Nüsslein-Volhard, Wieschaus, and Kluding 1984). Assuming a generation time of 25 days, the lines were roughly 250 generations old. All microsatellites were located on the second chromosome, which was maintained over a CyO balancer chromosome. Of the 24 microsatellite loci (22 dinucleotide and 2 trinucleotide repeats) screened, 13 were homozygous and 11 were heterozygous (table 1). Two types of mutations were observed: loss of one allele and mutation to a novel allele. The loss of one allele was detected in one line at the loci G411 and DS08011 and in another line at locus DS08011 only. Both loci are located distal to the CyO break at 58A and are free to recombine. Thus, the loss of one allele is most likely caused by recombination between the balancer chromosome and the other second chromosome rather than by a mutation in the microsatellite. All other mutations were detected at locus DROYANETS. Of the two ancestral alleles with 84 and 120 bp, only the longer allele was altered; this allele, which is located on the balancer chromosome, has 28 CA repeats. With a reported average microsatellite length of 10 repeats (Schug et al. 1998b), DROYANETS represents one of the longest microsatellites in D. melanogaster. Four of the nine observed mutations changed the repeat number by one. The remaining mutations altered allele size by at least three repeat units (fig. 1), indicating that the majority of the mutations do not follow a strict stepwise mutation process. Sequence analysis of two mutations demonstrated that allele size variation is caused by changes in the microsatellite stretch only. The flanking region was identical for the alleles on the balancer chromosome and on the mutagenized chromosome. No base substitution was detected in the microsatellite stretch of the 94-bp allele, ruling out the possibility that the mutation rate of the shorter allele was influenced by imperfections in the microsatellite.

Assuming that microsatellite variation is neutral and that each line is in mutation-drift equilibrium for both chromosomes, the mutation rate of the 28-repeat allele can be calculated as \( 9/(250 \times 119) = 3.0 \times 10^{-4} \). This allele-specific mutation rate is comparable with the mutation rate estimates for mammals, which are on the order of \( 10^{-3} \)–\( 10^{-5} \) (Dallas 1992; Weber and Wong 1993; Ellegren 1995). The mutation rate averaged across all loci is \( 6.3 \times 10^{-6} \) (\( = 9/(2 \times 250 \times 119 \times 24) \); 95% confidence interval \( = [1.2 \times 10^{-5}, 3.1 \times 10^{-6}] \)), which is similar to that reported by Schug, Mackay, and Aquadro (1997).

Exchange of Balancer Chromosomes

It should be noted that during the propagation of the lines, balancer chromosomes may have been exchanged. An exchange of balancer chromosomes in some of the lines used in our study would affect all microsatellite alleles on them. As we did not detect any evidence of an unusual pattern apart from the four contaminant lines, we think that balancer exchange either did not occur or did not affect the outcome of our typings.

Discussion

Mutation Rate Estimates

Assuming that microsatellite variation is neutral and each line is in mutation-drift equilibrium for both
### Table 1
Microsatellite Loci Analyzed for Mutations

<table>
<thead>
<tr>
<th>Locus</th>
<th>Repeat Unit Length</th>
<th>Chromosomal Location</th>
<th>Annealing Temperature (°C)</th>
<th>Primer Sequences (5’–3’)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z50409</td>
<td>12</td>
<td>Allele 1</td>
<td>22B1–22B9</td>
<td>GCAAACGGAAAGCAGAAGA</td>
<td>Harr et al. (unpublished)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Allele 2 (balancer)</td>
<td></td>
<td>GTGTTGTCACCTTCCTCTCAC</td>
<td></td>
</tr>
<tr>
<td>DS01340</td>
<td>14</td>
<td>24A1–24A2</td>
<td>55</td>
<td>GAGTAATGTCCTGCTCAGGAC</td>
<td>Harr et al. (unpublished)</td>
</tr>
<tr>
<td>DROGPDH</td>
<td>7</td>
<td>25F5–26A</td>
<td>53</td>
<td>CTTGCATATCCTCTACAGGAGG</td>
<td>Schug et al. (1998)</td>
</tr>
<tr>
<td>G410</td>
<td>6</td>
<td>33E9–33E10</td>
<td>53</td>
<td>AACATATACGGCAAAAGACC</td>
<td>Harr et al. (1998)</td>
</tr>
<tr>
<td>674A</td>
<td>8</td>
<td>35E6–35F5</td>
<td>54</td>
<td>TGTGACGTCAGTCAAAAGC</td>
<td>Harr et al. (1998)</td>
</tr>
<tr>
<td>DS00762</td>
<td>12</td>
<td>37B1–37B2</td>
<td>52</td>
<td>GTGACGTGTCACATTCAGGCG</td>
<td>Harr et al. (1998)</td>
</tr>
<tr>
<td>DS08513</td>
<td>0</td>
<td>37B8–37C2</td>
<td>54</td>
<td>CATCACATGCTCTGAAGGGAAG</td>
<td>Schlotzerer, Vogl, and Tautz (1997)</td>
</tr>
<tr>
<td>DS00665/1</td>
<td>16</td>
<td>38A1–38A2</td>
<td>55</td>
<td>CCTTCATGTGCTAGGCTCAGCA</td>
<td>Harr et al. (unpublished)</td>
</tr>
<tr>
<td>5915</td>
<td>10</td>
<td>44D5–44E2</td>
<td>54</td>
<td>GAGGCGTGTGACATGAAAGC</td>
<td>Harr et al. (1998)</td>
</tr>
<tr>
<td>DROGPAD</td>
<td>18</td>
<td>47A</td>
<td>56</td>
<td>AAATAGAAGCTTCTTTGTGAATGCC</td>
<td>Schug, Mackay, and Aquadro (1997)</td>
</tr>
<tr>
<td>DMMASTER</td>
<td>8</td>
<td>50C</td>
<td>55</td>
<td>TATTTTCACATTTCCAATCTCG</td>
<td>Goldstein and Clark (1995)</td>
</tr>
<tr>
<td>DMELF</td>
<td>8</td>
<td>54F</td>
<td>55</td>
<td>TCGGACACCCTGGAGGCTAG</td>
<td>Goldstein and Clark (1995)</td>
</tr>
<tr>
<td>DS00144</td>
<td>5</td>
<td>55A2–55B1</td>
<td>52</td>
<td>AAATCAAAACTCTCAAGGCG</td>
<td>Harr et al. (unpublished)</td>
</tr>
<tr>
<td>DS08687a</td>
<td>4</td>
<td>57C5–57D1</td>
<td>54</td>
<td>GAGGAGAGGCGATCTTCCAGC</td>
<td>Schlotzerer, Vogl, and Tautz (1997)</td>
</tr>
<tr>
<td>DS08687b</td>
<td>12</td>
<td>57C5–57D1</td>
<td>52</td>
<td>ATGTCCTTCCCTTTTACAGCA</td>
<td>Schlotzerer, Vogl, and Tautz (1997)</td>
</tr>
<tr>
<td>Z32225</td>
<td>6</td>
<td>58C1–58C7</td>
<td>52</td>
<td>ATTGTCGTCGTCACATCGCTCG</td>
<td>Harr et al. (unpublished)</td>
</tr>
<tr>
<td>G411</td>
<td>Not sized</td>
<td>+2</td>
<td>58C2–58C5</td>
<td>GGAAGGGGAGACAAATCACAGCA</td>
<td>Harr et al. (1998)</td>
</tr>
<tr>
<td>DROMHIC</td>
<td>10</td>
<td>60E9</td>
<td>57</td>
<td>GACATTTACGATATGAGGCTCA</td>
<td>Schug et al. (1998)</td>
</tr>
</tbody>
</table>

*Note:* The size differences are given for the ancestral alleles on both chromosomes. All repeat numbers, except for DROGPDH, were inferred from allele sizes given in the original publication. The repeat number for DROGPDH was corrected according to sequencing of Drosophila melanogaster alleles (B. Harr, personal communication).
chromosomes, the average mutation rate across all loci is $6.3 \times 10^{-6}$. The good agreement of our mutation rate estimate with that of Schug, Mackay, and Aquadro (1997) is interesting, as we used an almost completely different set of microsatellite loci as well as different *D. melanogaster* lines. Both studies screened the same number of loci and observed mutations at a single locus. While Schug, Mackay, and Aquadro (1997) observed only one mutation, we detected nine at one locus. This distribution is extremely unlikely to occur if all loci have the same mutation rate ($P < 0.0002$, binomial). Hence, one highly variable locus significantly affected the average mutation rate of *D. melanogaster* microsatellites, while a large fraction of microsatellite loci remained invariant. Our calculated upper 95% confidence limit for the mutation rate per locus per generation is $2.1 \times 10^{-6}$ for the 23 loci at which no mutation was detected. This confidence limit is one order of magnitude lower than the estimate obtained by Schug, Mackay, and Aquadro (1997). This discrepancy results from the larger number of meiotic divisions analyzed in our study. The fact that an increase in meiotic divisions did not result in mutations at more loci indicates that most loci have very low mutation rates.

If a high mutation rate is characteristic of the *DROYANETSB* locus, then both alleles have the same probability of mutation. We observe, however, that all mutations occurred at a single allele only, rejecting the simple model of a highly variable locus ($P = 0.017$, binomial). All mutations occurred at the *DROYANETSB* allele with 28 repeats, which is the longest allele in our study. The second longest microsatellite allele (*DROGANETSB*) is 10 repeats shorter. In their survey of *D. melanogaster* microsatellites, Schug et al. (1998) detected only two microsatellite loci with more than 21 repeats, indicating that the *DROYANETSB* allele with 28 repeats is among the longest reported for *D. melanogaster*.

Mutation Rate Estimates from Natural Populations

To compare our results with the variability of *DROYANETSB* in natural populations, we screened microsatellite variability in three *D. melanogaster* populations. Variances in repeat number ranged from 19.1 to 41.0, with an average of 29.9. Previously, variances in repeat number were found to vary between loci. The reported variances range from 0.0 to 43.0, with an average of 7.6 (Goldstein and Clark 1995; Schlötterer, Vogl, and Tautz 1997; Harr et al. 1998). Thus, the variances observed for *DROYANETSB* are higher than those for most other loci in *D. melanogaster*. By assuming an $N_e$ of $10^5$, and the stepwise mutation model, the average microsatellite mutation rates can be estimated as $\mu = \mu/4N_e$ (Moran 1975). For the observed variances of *DROYANETSB*, we obtain a mutation rate of $7.5 \times 10^{-6}$ ($=29.9/[4 \times 10^5]$). Interestingly, this value is lower than our estimates for the mutation rate of the long *DROYANETSB* allele but is close to the mutation rate averaged across 24 *Drosophila* microsatellites.

Parameters Influencing Microsatellite Mutation Rate

Recently, it has been proposed that large size differences between microsatellite alleles result in an increased mutation rate. Amos et al. (1996) tested for the influence of size differences on the mutation rate by asking if more mutations occurred in offspring of parents with a larger size difference than in offspring of parents with a smaller size difference. A significant deviation from the 1:1 ratio was interpreted as support for an increasing mutation rate with increasing size difference between alleles. In agreement with this model, the highly variable *DROYANETSB* locus has the largest size difference between alleles. While this model of heterozygote instability should predict that both alleles mutate at the same rate, our distribution of mutations (9:0) does not support this ($P = 0.017$, binomial), suggesting causes other than size difference between alleles. Additional evidence against influence of heterozygosity on microsatellite mutation rates is provided by the lines used in our study. They carried two different second chromosomes which have been maintained in permanent heterozygous state, as both chromosomes are lethal in the homozygous state. If only heterozygous loci are considered, the distribution of the observed mutations is not compatible with equal mutation rates across loci ($P < 0.001$, binomial). Excluding the microsatellite locus *DROYANETSB*, the upper 95% confidence limit for the mutation rate is $5 \times 10^{-6}$, which is still lower than that obtained in a study using homozygous lines (Schug, Mackay, and Aquadro 1997). Nevertheless, it should be considered that the lines used to measure mutation rates in *D. melanogaster* carried a balancer chromosome, which suppresses recombination. If the heterozygosity effect suggested by Amos et al. (1996) depends on recombination, then our experimental design is not well suited to test this hypothesis. On the other hand, there is no experimental support that recombination or other interchromosomal exchange mechanisms are involved in microsatellite mutations. In two organisms, yeast and *E. coli*, it has been shown that deficiencies in the recombination pathway do not result in an increased microsatellite mutation rate (Levinson and Gutman 1987a; Henderson and Petes 1992). Furthermore, studies of humans on polymorphisms flanking a microsatellite in the Cystic Fibrosis Transmembrane Conductance Regulator
(CFTR) gene demonstrated a microsatellite mutation without exchanges in the flanking regions (Morral et al. 1991).

An alternative explanation for our data is that mutation rate is dependent on the repeat number at a given allele. This interpretation is consistent with recent studies of yeast (Wierdl, Dominska, and Petes 1997). Using plasmids carrying microsatellites of different lengths, Wierdl, Dominska, and Petes (1997) showed that microsatellite mutation rate increases more than linearly with repeat number. Further support for a higher mutation rate for longer microsatellites can be found in a recent study of barn swallows (Primmer et al. 1996). Similarly, a dinucleotide microsatellite locus tightly linked to HLA-DQB1 in humans provides support for a length-dependent mutation rate of microsatellites (Jin et al. 1991). Given the wealth of support for the length dependence of microsatellite mutation rates, we conclude that our data are also best explained by a higher mutation rate of the longer allele.

Mutational Bias

Two recent studies have described a mutational bias for microsatellite mutations (Rubinsztein et al. 1995; Primmer et al. 1996). Both studies demonstrated a bias toward longer alleles. If this bias is maintained during microsatellite evolution, an infinite growth of the microsatellite is predicted. Because no microsatellite has been observed that is close to the predicted infinite number of repeats, several hypotheses of long-term microsatellite evolution have been discussed. The spectrum ranges from a neutral explanation that assumes that microsatellite evolution have been discussed. The spectrum ranges from a neutral explanation that assumes that microsatellite beyond a certain length collapse (Levinson and Gutman 1987b), to selection against long microsatellite alleles (Garza, Slatkin, and Freimer 1995). Recently, experimental data from yeast demonstrated that microsatellite longer than 25 repeats have a different mutation spectrum than do shorter ones. Long microsatellites in yeast lose on average about 4 bp per mutation (Wierdl, Dominska, and Petes 1997; Schlötterer 1998a). This suggests that a reversal in the mutational bias of long alleles could be responsible for the maintenance of a certain allele size range for microsatellites. Interestingly, our observed mutations in D. melanogaster show a similar trend. The average loss of 3.3 bp per mutation is very close to the yeast results. We were interested in studying the allele distribution of DROYANETS B to see if long alleles are observed in natural populations. Figure 2 shows that in all three populations, long microsatellite alleles are present at a low frequency. Interestingly, the longest allele is only a few repeats longer than the 120-bp allele in our study. It is almost impossible to prove from population data that this distribution results from a high mutation rate of long alleles combined with a mutational bias toward shorter alleles. A similar distribution may be expected if the long allele is of recent origin. Nevertheless, a mutational bias of long alleles toward shorter alleles provides an interesting model of long-term microsatellite evolution, as it may explain why several studies failed to find a correlation between microsatellite variability and average microsatellite length (Michalakis and Veuille 1996; Harr et al. 1998; Schug et al. 1998a).

Acknowledgments

We are grateful to the laboratory of C. Nüsslein-Volhard, the UMEA Stock Center, the Bloomington Stock Center, R. C. Woodruff, and D. Slezak for providing flies. S. Luschnik performed the phenotypic identification of the contaminated lines. Many thanks to D. Bachtrog for her help in sequencing the microsatellite alleles. M. Schug and S. Weiss provided many helpful comments on the manuscript. This work is supported by grants of the Fonds zur Förderung der Wissenschaften (FWF) and the European Community to C.S.

LITERATURE CITED


CHARLES F. AQUADRO, reviewing editor

Accepted June 26, 1998