Rapid Divergence of Microsatellite Abundance Among Species of Drosophila

Charles L. Ross, Kelly A. Dyer,*[1] Tamar Erez,* Susan J. Miller,† John Jaenike,*[1] and Therese A. Markow**

*Department of Ecology and Evolutionary Biology, University of Arizona; and †BioComputing Facility, Arizona Research Labs, University of Arizona

Among major taxonomic groups, microsatellites exhibit considerable variation in composition and allele length, but they also show considerable conservation within many major groups. This variation may be explained by slow microsatellite evolution so that all species within a group have similar patterns of variation, or by taxon-specific mutational or selective constraints. Unfortunately, comparing microsatellites across species and studies can be problematic because of biases that may exist among different isolation and analysis protocols. We present microsatellite data from five Drosophila species in the Drosophila subgenus: *D. arizonae*, *D. mojavensis*, and *D. pachea* (three cactophilic species), and *D. neotestacea* and *D. recens* (two mycophagous species), all isolated at the same time using identical protocols. For each species, we compared the relative abundance of motifs, the distribution of repeat size, and the average number of repeats. Dimers were the most abundant microsatellites for each species. However, we found considerable variation in the relative abundance of motif size classes among species, even between sister taxa. Frequency differences among motifs within size classes for the three cactophilic species, but not the two mycophagous species, are consistent with other studied Drosophila. Frequency distributions of repeat number, as well as mean size, show significant differences among motif size classes but not across species. Sizes of microsatellites in these five species are consistent with *D. virilis*, another species in the subgenus Drosophila, but they have consistently higher means than in *D. melanogaster*, in the subgenus *Sophophora*. These results confirm that many aspects of microsatellite variation evolve quickly but also are subject to taxon-specific constraints. In addition, the nature of microsatellite evolution is dependent on temporal and taxonomic scales, and some variation is conserved across broad taxonomic levels despite relatively high rates of mutation for these loci.

Introduction

Microsatellites are regions of the genome comprising a variable number of repeats of simple base pair sequences. Allelic variants at microsatellite loci differ primarily in the number of these repeats, with such variation arising as a result of slippage during DNA replication (Sia, Jinks-Robertson, and Petes 1997). Such changes at microsatellite loci occur at rates orders of magnitude greater than base substitution mutations in nonrepetitive DNA. The high rate of mutation and presumably low selection coefficients associated with variant alleles results in high levels of heterozygosity and allelic diversity at microsatellite loci. Thus, the combined presence of thousands of microsatellite loci throughout eukaryotic genomes and the ease with which microsatellite variation can be scored have made these loci increasingly important in studies of the genetic structure of populations, parentage analysis, and genetic mapping.

Beside their importance in genetic and evolutionary studies, factors governing microsatellite variation are also of direct biological importance. For instance, elevated rates of mutation at microsatellite loci are associated with certain hereditary diseases (Gryfe et al. 1997). More important, mutations that affect microsatellite stability, such as those involved in DNA mismatch repair, can also influence overall levels of genomic stability (Degtyareva et al. 2002). Consequently, microsatellite stability may be correlated with overall levels of genomic stability.

Recent hypotheses regarding variation at microsatellite loci (Kruglyak et al. 1998; Kruglyak et al. 2000; Calabrese, Durrett, and Aquadro 2001) have shown that the distribution of allele sizes at a locus may be explained by a balance between slippage rates, which tend to increase mean and variance of allele size, and point mutations within loci that break up continuous segments of repeats and thus decrease allele size. If this is so, then differences between species in patterns of microsatellite variation are likely to result from differences in the rates and nature of mutation associated with microsatellite variation. Given the potential biological importance of these loci, selection may also affect overall allele size. Because these processes are likely to be of considerable importance to the organism, one may expect that closely related species would exhibit similar patterns of microsatellite variation.

Previous surveys of microsatellite variation have focused largely on variation at two levels—within species (both allelic variation and among loci) and among higher taxa. The high levels of allelic variation within species have sparked the widespread use of these markers in recent evolutionary studies. At the other end of the hierarchical spectrum, comparative analyses of microsatellites have uncovered substantial differences among higher taxonomic groups (Goldstein and Schlötterer 1999; Katti, Ranjekar, and Gupta 2001; Neff and Gross 2001). Two observations are particularly noteworthy. First, different microsatellite motifs (the repeated unit) are not equally represented in species belonging to different groups (Katti, Ranjekar, and Gupta 2001). For instance, the most abundant motif in many insects is AC, but in honeybees and wasps AG is the most abundant (Estoup et al. 1993*b*; Thoren, Paxton, and Estoup 1995). Similarly, the relative numbers of monomers, dimers, and trimers differ among vertebrate classes,

[1] Present address: Department of Biology, University of Rochester.

Key words: microsatellite evolution, Drosophila, *D. mojavensis*, *D. arizonae*, *D. pachea*, *D. neotestacea*, *D. recens*.

E-mail: clross@email.arizona.edu.

Mol. Biol. Evol. 20(7):1143–1157. 2003
DOI: 10.1093/molbev/msg137
Molecular Biology and Evolution, Vol. 20, No. 7

whereas vertebrates as a whole have a greater abundance of tetramers than many other groups of organisms (Toth, Gáspári, and Jurka 2000; Neff and Gross 2001). These disparities in microsatellite representation among higher taxa cannot be explained solely by differential base composition of genomes (Field and Wills 1998; Bachtrog et al. 2000; Kruglyak et al. 2000; Katti, Ranjekar, and Gupta 2001).

A second observation is that taxa vary greatly with respect to average microsatellite length, the number of repeated units. Drosophila, for instance, have shorter microsatellite alleles than many other organisms, such as mammals, fish, and some other insects (Schug et al. 1998*b*; Toth, Gáspári, and Jurka 2000; Neff and Gross 2001). Given the assumption that patterns of microsatellite variation reflect a balance between mutational processes within a species, it is not surprising that these processes differ among species that have been separated for hundreds of millions of years.

Although microsatellite variation has been well characterized within species and among higher taxa, there has been very little comparative study of genome-wide patterns of microsatellite variation among closely related species. Such a comparative approach would provide insight into the stability of the mutational (and perhaps selective) processes acting on genomes over relatively short macroevolutionary time periods. Specifically, if closely related species experienced identical mutational and selective processes, then we would predict that they would exhibit similar patterns of microsatellite variation. If these processes varied among species, then we would expect this to be manifested in differences among species in their patterns of microsatellite variation. Given the possible connection between microsatellite stability and overall genomic stability, such interspecific comparisons might also shed light on the general malleability of the genomes in these lineages.

One potential problem in making such comparisons among species is the biases inherent in various methods of microsatellite characterization. That is, if different investigators used different methods to identify and characterize microsatellite loci in their particular study species, then some of the variation among species could be due to methodological biases rather than true genomic differences among the species. For example, studies of microsatellite variation in several Drosophila species support the possibility that microsatellites evolve quickly but are subject to taxon-specific constraints (Hutter, Schug, and Aquadro 1998; Schug et al. 1998*a*; Pascual, Schug, and Aquadro 2000; Schlötterer and Harr 2000; Noor, Kliman, and Machado 2001). Nonetheless, many such comparisons among species suffer from a lack of consistency in the methods of collection and interpretation of microsatellite loci data sets. Two common methods to isolate and characterize microsatellites, searching DNA databases (e.g., GenBank) and constructing genomic DNA libraries, each have their own sets of biases (Schug et al. 1998*b*; Schlötterer and Harr 2000). Furthermore, microsatellite variation found by these two methods can differ from variation found in natural populations. Consequently, if different methods are used in different studies, the inherent biases may contribute to the apparent variation among related species in microsatellite patterns. Hence, one way to get better estimates of the evolutionary stability of microsatellite patterns is to use similar or identical methods on different species.

The present study was undertaken to ask whether patterns of microsatellite variation are stable among congeneric species of Drosophila. The species we examined diverged from each other between 1 and ~56 MYA, i.e., over relatively short macroevolutionary time spans. The particular species we consider include *D. arizonae*, *D. mojavensis*, *D. pachea*, *D. neotestacea*, and *D. recens*. These five species belong to three major radiations within the subgenus Drosophila (Powell 1997). *Drosophila arizonae* and *D. mojavensis* are sister taxa within the *repleta* group and are capable of hybridization. *Drosophila pachea* belongs to the *nannoptera* species group, and *D. recens* and *D. neotestacea* belong to the closely related *quinaria* and *testacea* groups, respectively. These species are ecologically diverse, including both mushroom-feeding (*D. recens* and *D. neotestacea*) and cactophilic species (*D. arizonae*, *D. mojavensis*, and *D. pachea*).

To minimize the effects of methodological bias in the identification and characterization of microsatellite variation, we used identical laboratory and analytical methods for all five species. Thus, any variation we see among species reflects genome level differences in microsatellite variation, rather than a methodological artifact. Our study did not examine allelic variation within microsatellite loci; rather, by sampling across loci, our intention was to obtain a random sample across the genome of motif representation and the frequency distributions of repeat lengths within motif classes.

Our results reveal that the abundance of various microsatellite motifs varies considerably among the five Drosophila species, but the length-frequency distributions are relatively constant across all species. Thus, genome-wide patterns of microsatellite variation among these species, and presumably the mutational and selective processes underlying those patterns, are relatively stable when considering microsatellite length, but these patterns show considerable evolutionary lability in the ''birth'' and ''death'' of microsatellite loci.

Materials and Methods

The Species

Drosophila arizonae is a generalist cactophilic fly found in Arizona south to Guatemala (Fellows and Heed 1972; Ruiz, Heed, and Wasserman 1990; Markow and Hocutt 1998). *Drosophila mojavensis* and *D. pachea* are endemic to the Sonoran desert (Fellows and Heed 1972; Heed and Mangan 1986; Markow and Hocutt 1998); these species both specialize on necrotic tissue of specific cacti (Fellows and Heed 1972; Ruiz and Heed 1978; Heed and Mangan 1986; Pitnick 1993). Both *D. neotestacea* and *D. recens* are generalist mycophagous, or mushroom-feeding, flies. *Drosophila neotestacea* is common in temperate and boreal forests across northern North America (Grimaldi, James, and Jaenike 1992), and *D. recens* is restricted to the cooler forests of central and eastern North

America (Jaenike 1978; Shoemaker and Jaenike 1997). Phylogenetically, these species represent both closely and distantly related species within the subgenus Drosophila (fig. 1).

Isolation and Characterization of Microsatellites

For each species, we extracted genomic DNA from about 150 individuals (males and females) using Phenol/Chloroform (Sambrook, Fritsch, and Maniatis 1989). All fly stocks originated from multiple inseminated females and were maintained as multi-female stocks. Specific stocks and collection dates are as follows: *D. arizonae* (PERA497) were collected in 1997 from the Superstition Mountains in Arizona; *D. mojavensis* (GU500) and *D. pachea* (GU500) were both collected in 2000 from Guaymas, Sonora; *D. neotestacea* were collected in 1990 from Rochester, N.Y., and *D. recens* were collected in 1996 from Big Moose, N.Y.

Microsatellite sequences were isolated from an enriched genomic library using the protocol of Hamilton et al. (1999). Briefly, genomic DNA for each species was completely digested with *Hae*III, *Nhe*I, and *Sau*3a (ratio of enzymes: 5 Units: 2.5U: 0.5U) to give DNA fragments between 200–1,000 bp long. These fragments were ligated to linker DNA and then amplified using polymerase chain reaction (PCR) to give numerous copies of each fragment. The fragments were enriched for microsatellite sequences by hybridizing the DNA pool in separate reactions to biotinylated $(AC)_{15}$, $(CG)_{15}$, $(AGC)_{10}$, and $(ATC)_{10}$ oligonucleotides. The hybridized sequences enriched for microsatellites were recovered using steptavidin-coated iron beads. DNA enriched for microsatellites was PCR amplified, cloned into a pBluescript II sk+ vector (Stratagene), and transformed into *E. coli* (XL2 MRF Ultracompetent cells, Stratagene).

Plasmid inserts were amplified using colony PCR and then sequenced on an ABI 3700 capillary automated sequencer at the GATC core facility at the University of Arizona. The Hamilton enrichment protocol results in a high percentage of positive colonies in the primary screen so that no secondary screen is normally necessary. In our screens of colonies, approximately 70%–80% contained sequences with microsatellites. It is also important to note that this protocol results in potential opportunities to duplicate alleles at loci through a number of PCR steps, transformation and plating *E. coli*, and using hundreds of flies as source DNA.

Processing Sequences and Identification of Microsatellites

Sequences were processed using Faktory (Miller and Myers 1999, downloads available at http://bcf.arl.arizona.edu/faktory/) to remove poor quality sequence as called by Phred (Ewing et al. 1998) and to identify and remove vector sequence. Next, UniqFlank, a program in Perl developed by S.J.M., created a Blast database of input sequences and then found duplicates using the entire sequence, including any microsatellites that may be contained within the sequences. We used the following Blast parameters: gap formation penalty (G) = –22, gap extension penalty (E) = –2, expect value cutoff (e) = 10^{-10}, low complexity filter off. For those sequences that match, UniqFlank randomly chose one to keep and removed the others. This step eliminated any clonal or PCR duplicate sequences in the data set while allowing for different alleles at the same locus. In UniqFlank, two matching criteria are used: the minimum percentage match between two sequences and the minimum percentage length between subject and query sequences. At this step, we used the following matching criteria: minimum percentage match (min%) = 98%, minimum length match (minlen%) = 95%.

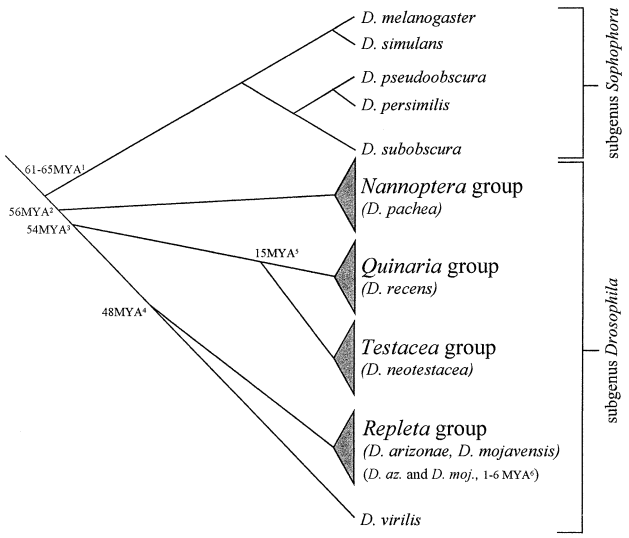

FIG. 1.—Phylogenetic relationship among study species as well as representatives in *Sophophora* subgenus. Note that phylogeny is not to scale and divergence times are approximate. Divergence estimates: (1) Beverley and Wilson (1984); (2) Pitnick, Spicer, and Markow (1997); (3) Pitnick, Spicer, and Markow (1997); (4) Pitnick, Spicer, and Markow (1997); (5) Spicer (in preparation); (6) undetermined, see Ruiz, Heed, and Wasserman (1990).

Next, UniqFlank used the pattern-finding program, Tandem Repeats Finder (TRF, Benson 1999), to find all repeated patterns in the sequences, including microsatellites. TRF has major advantages over some pattern-finding programs in that it does not require a pattern library to recognize patterns. A pattern library can potentially limit recognition of some patterns. Additionally, TRF recognizes both simple and complex patterns, which allows it to readily identify any size repeat motifs, and the statistically based recognition criteria of TRF enable it to identify ''imperfect'' microsatellite runs without the limitation of an arbitrary and restrictive rule for imperfections in a patterned run of DNA. We used the following TRF parameters: match reward = +2, mismatch penalty = –8, indel penalty = –8, minimum alignment score = 20, maximum pattern = 6.

After identifying microsatellites, UniqFlank created a Blast database of the flanking region around each microsatellite, and then found matches using only the flanking sequence. In this step, UniqFlank can identify different alleles of the same locus, as well as loci shared among species. Here, we used the following parameters: Blast – G = 5, E = 2, e = 10^{-10}; UniqFlank – min% =

95%, minlen% = 95%. Lastly, UniqFlank scored microsatellites for motif and allele length—both total, or ''imperfect'', as well as ''perfect'' runs within the total length. Further information about this program is available from the authors.

Data Analysis

All motif size classes from one to six base pairs in length (i.e., microsatellites) were scored from the sequences isolated from the five species. We adopt the convention of reporting motifs that represent the same microsatellite by the motif sequence that gives the first alphabetical motif. For example, the motif AC = AC, CA, GT, TG. Likewise, AGC = GCA, CAG, GCT, CTG, TGC.

We scored the ''total'' length of a microsatellite allele, which allows for ''imperfections'' such as point mutations and indels. The inclusion of imperfections in the microsatellite length depended on the microsatellite's overall alignment score with a consensus pattern determined from the sequence itself. TRF identifies a potential repeat pattern by first analyzing the sequence of interest. Once a consensus pattern is determined, the actual sequence is aligned to that pattern and the microsatellite length is scored as the longest amount of sequence (or number of repeat units in the microsatellite) that gives a minimally acceptable alignment score based on Smith-Waterman style local alignment. In this way, alleles may contain numerous imperfections if their alignment scores are still high. For our analysis, match, mismatch, and indel weights were +2, –8, and –8, respectively, which produce relatively restrictive conditions for allowing imperfections. Microsatellites must have had an alignment score of at least 20 to be accepted. In addition, UniqFlank also records ''perfect'' runs of alleles, which allows for comparison between perfect and imperfect lengths. For the analysis, we only used microsatellite alleles of at least five repeat units with the exception of monomers, which must have been 10 bp long before they were included.

We performed three general analyses to determine the consistency with which microsatellites are represented among the five species. First, we compared the relative abundance of motifs and motif classes within and among the five species using Fisher exact tests. Second, we compared the overall shape of the size frequency distributions for motifs and motif classes within and among species using Kolmogorov-Smirnov tests. Third, we compared the length of alleles for motifs and motif classes within and among species using Tukey-Kramer HSD tests. All statistics were calculated using the software package JMP version 3.1.6 (SAS Institute) or DataDesk version 6.1 (Data Description), except Kolmogorov-Smirnov tests, which were calculated using StatView version 4.1 (Abacus Concepts).

Results

Abundance of Motifs Within and Among Species

We recovered a large assortment of microsatellite motifs from each species (table 1; Total microsatellites recovered/sequences screened: *D. arizonae* – 245/366, *D. mojavensis* – 350/368, *D. pachea* – 275/558, *D. neotestacea* – 169/186, *D. recens* – 90/128). Overall, the abundance of motif classes differs within and among species (Pearson $\chi^2_{12} = 76.75$, $P \leq .0001$; table 2). If monomers are excluded, microsatellite abundance is inversely related to the size of the motif class (fig. 2). Relative to the other species, *D. arizonae* has a high abundance of trimers (46.1%) and a low abundance of dimers (46.5%) (tables 1 and 2, fig. 2). However, *D. mojavensis*, the sister taxa to *D. arizonae*, has a high abundance of dimers (73.7%) and a low abundance of trimers (16.7%) relative to the other species. *Drosophila recens* has a relatively high abundance of microsatellites with long (4–6) motifs (12.2%), and *D. neotestacea* has a marginally lower abundance of trimers (20.7%). Thus, our species vary in the abundance of microsatellite loci as a function of motif size. It should be noted that our data set might overrepresent dimers and trimers, because these size classes were the targets of our oligonucleotide probes.

Within the dimer class, we compared the relative abundance of all dimer motifs across species (fig. 2). *Drosophila mojavensis* has a relatively low abundance of AT compared to other species (0.39% of all dimers, Fisher exact test, $P = .0305$; fig. 2), and tends toward a relatively high frequency of AC (90.31%, $P = .0625$; fig. 2). *Drosophila recens* has a relatively high frequency of AT (9.80%, Fisher exact test, $P = .0013$; fig. 2), and tends toward a relatively low frequency of AC (78.43%, $P = .0792$; fig. 2). Within trimers, only the motifs, AGC, ATC, and AAC were sufficiently abundant to compare species (fig. 2). *Drosophila pachea* has a relatively high proportion of the motif AGC and a relatively low proportion of AAC (fig. 2). In both *D. neotestacea* and *D. recens*, AGC is relatively rare compared to AGC in the other species, and ATC and AAC are both relatively abundant (fig. 2). Additionally, the motif ATC is marginally low in *D. mojavensis*. Consequently, as with motif size classes, motifs differ in their abundance among species.

Overall, the four motifs that were presumably enriched because of our protocol varied in frequency. AC was highly abundant (87.1% of all dimers, $\chi^2_1 = 6.818$, $P = .009$) for all species and was also the most commonly found motif from all four of our oligonucleotide hybridization reactions for each species. AGC was highly abundant in *D. arizonae*, *D. mojavensis*, and *D. pachea*, but much less abundant in *D. neotestacea* and *D. recens* (76.9% for the former three species, 25.4% for the latter two species). CG and ATC, however, were relatively rare (0% of all dimers and 8.1% of all trimers, respectively).

Frequency Distributions of Microsatellite Length

For all species combined, the difference in the size frequency distributions between dimers and trimers is highly significant ($\chi^2_2 = 170.23$, $P = <.0001$; table 3). This difference between these motif classes overall also is apparent when dimers and trimers are compared within each species (table 3). In general, dimer frequency distributions tend to have a higher occurrence of rare very long alleles, resulting in longer tails in the dimer distributions than in the trimer distributions (fig. 3). Furthermore, the ''peak,''

or most common allele, for dimer distributions is shifted farther from the minimum microsatellite size than for trimers. This trend—as motif size increases, distributions become more truncated and have shorter alleles—is also consistent with the distributions of longer motif classes (fig. 3). It is important to note that these frequency distributions represent one randomly chosen allele from each of many microsatellite loci within a species. Consequently, comparisons across species do not represent assessments of orthologous alleles or loci.

For every pairwise comparison, dimer frequency distributions and trimer frequency distributions differed significantly among species only when comparisons involved either *D. pachea* or *D. neotestacea* (table 3, fig. 3). Indeed, the contrast with the greatest difference is that between these two species (table 3). By inspection, the distributions of these two species appear to represent the most deviant distribution shapes for our dimer data set: *D. pachea* has a peak shifted farthest from the minimum size, and *D. neotestacea* has a peak closest to the minimum size of five repeats (fig. 3). No pairwise species comparisons for trimers were significant. These results show that most variation in the frequency distributions of microsatellite sizes occurs among motif size classes. In contrast, among species, differences in size frequency distributions are generally low.

Additionally, we compared the frequency distributions within and among species for the two most abundant dimer and trimer motifs, AC and AGC. As in the general case with all dimers and all trimers, the frequency distributions for the motifs, AC and AGC, are highly significantly different ($\chi^2_2 = 135.78$, $P = <.0001$) when all species are combined. This difference is also highly significant within each species comparison, except in *D. neotestacea* (data not shown). As with dimers and trimers in general, AC distributions show longer rare alleles (long tails) and modes shifted farther away from the minimum allele size than AGC distributions. As indicated, *D. neotestacea* appears to be the exception to this trend, as this species has fewer long alleles.

Pairwise species comparisons for the AC motif show similar significance patterns as when all dimers are combined (data not shown). Again, all significant comparisons involve either *D. pachea* or *D. neotestacea*, and the distributions most deviant are from these two species, but in opposite ways. No pairwise species comparisons for AGC were significant. These AC versus AGC comparisons suggest that the differences observed between dimers versus trimers as a whole extend to specific motifs within each size class.

Mean Allele Lengths of Microsatellites

Because many of the frequency distributions are different in overall shape, an investigation is warranted into one specific aspect of shape, namely, the ''location'' of the distributions. Mean allele size can be used to summarize the location of a frequency distribution, though this certainly does not capture the entire spectrum of variation in these non-normal distributions. Nonetheless, this value is of interest because it is useful not only for interpreting the evolutionary and mutational properties of microsatellite loci (Kruglyak et al. 1998; Harr and Schlötterer 2000; Kruglyak et al. 2000) but also for describing variation in natural populations of organisms (Schug et al. 1998*b*; Noor, Schug, and Aquadro 1999; Pascual, Schug, and Aquadro 2000).

Over all species, motif classes differ in mean allele lengths (Kruskal-Wallis rank sum test: $\chi^2_5 = 252.75$, $P = <.0001$; table 4). Pairwise motif class comparisons of mean length indicate the following relationship: [monomers = dimers] > [trimers = tetramers = pentamers = hexamers] (table 5, Tukey-Kramer HSD). Long tails (maximums), as well as distributions shifted to higher allele lengths (median scores), account for the difference of monomers/dimers compared to other motif classes. By inspection, however, there is a tendency for smaller allele lengths as motif size increases (fig. 3).

When variation due to motif class and the interaction between motif class and species are partitioned, species do not differ significantly in mean allele length (Two-way analysis of variance [ANOVA] with interaction: species F-ratio = .02034, n.s. Species effect was dynamically transformed to normal approximation). This result is apparent despite marked differences among dimers.

Discussion

Microsatellite loci have been identified in the genome of every organism that has been searched (Goldstein and Schlötterer 1999; Katti, Ranjekar, and Gupta 2001; Neff and Gross 2001). Based on nucleotide composition in genomes, microsatellite sequences are greatly overrepresented. These loci have been found to vary considerably within species and among large taxonomic groups (Field and Wills 1998; Schug et al. 1998*b*; Goldstein and Schlötterer 1999; Bachtrog et al. 2000; Kruglyak et al. 2000; Toth, Gáspári, and Jurka 2000; Neff and Gross 2001). In this study, we isolated and characterized microsatellites from five species in the subgenus Drosophila that are related on a rarely studied taxonomic scale—i.e., among relatively closely related species. From our sequences, we compared the abundance and size variation of motifs and motif classes among the five species in order to document differences within Drosophila and to explore the macroevolutionary tempo and mode of microsatellite evolution. In our analysis, we found extensive differences among species, but we also found similarities both across our species and across broader taxonomic groups.

In this study our approach to understanding variation at microsatellite loci involved isolating microsatellite sequences from a cloned library. Our isolation protocol included a number of DNA duplication steps (PCR steps and a cloning step into a bacterial vector) and an enrichment technique using four oligonucleotide repeat motifs (30-mers) to hybridize to potential microsatellite sequences. For each species of interest, we performed four separate hybridization reactions, two dimers and two trimers, one traditionally abundant and one rare motif from each class. Unlike most previous studies where microsatellites have been isolated, we isolated sequences for multiple species at the same time, allowing us to compare

Table 1
Microsatellite Motifs Recovered

Species	Motif	*N*	% of Total	Mean[a]	(Perfect, *N*)[b]	SD	SE	Median	Min	Max	Range
D. arizonae	A	6	2.45	13.67	(12.5, 4)	3.50	1.43	13.5	10.0	18.0	8.0
	AAC	20	8.16	7.73	(7.17, 14)	1.88	0.42	7.5	5.0	11.7	6.7
	AACAGC	4	1.63	6.55		1.37	0.68	6.2	5.3	8.5	3.2
	AAG	1	0.41	10.30		—	—	10.3	10.3	10.3	0.0
	AC	99	40.41	12.96	(9.12, 110)	6.39	0.64	12.0	5.0	49.0	44.0
	ACACTC	1	0.41	10.50		—	—	10.5	10.5	10.5	0.0
	ACAG	1	0.41	8.20		—	—	8.2	8.2	8.2	0.0
	ACATAT	1	0.41	5.20		—	—	5.2	5.2	5.2	0.0
	ACGC	1	0.41	6.50		—	—	6.5	6.5	6.5	0.0
	AG	14	5.71	12.21	(8.80, 15)	4.31	1.15	11.5	5.5	22.0	16.5
	AGC	83	33.88	8.64	(6.89, 61)	2.68	0.29	8.0	5.0	18.3	13.3
	AGCTC	1	0.41	6.20		—	—	6.2	6.2	6.2	0.0
	AGGC	1	0.41	10.00	(6.00, 1)	—	—	10.0	10.0	10.0	0.0
	AT	1	0.41	5.00	(5.00, 1)	—	—	5.0	5.0	5.0	0.0
	ATC	8	3.27	9.80	(6.80, 8)	2.63	0.93	9.2	6.7	13.3	6.6
	C	2	0.82	12.00	(12.0, 2)	2.83	2.00	12.0	10.0	14.0	4.0
	CCG	1	0.41	6.70	(6.70, 1)	—	—	6.7	6.7	6.7	0.0
D. mojavensis	A	12	3.43	15.75	(12.6, 8)	9.33	2.69	13.0	10.0	43.0	33.0
	AAATGC	1	0.29	5.50	(5.50, 1)	—	—	5.5	5.5	5.5	0.0
	AAC	12	3.43	7.43	(6.85, 8)	1.80	0.52	6.7	5.0	11.0	6.0
	AACAC	1	0.29	5.20		—	—	5.2	5.2	5.2	0.0
	AACAGC	2	0.57	8.00		3.82	2.70	8.0	5.3	10.7	5.4
	AAGTCG	1	0.29	5.70		—	—	5.7	5.7	5.7	0.0
	AATGGG	1	0.29	11.20		—	—	11.2	11.2	11.2	0.0
	AC	233	66.57	13.82	(8.92, 276)	5.78	0.38	13.0	5.0	46.0	41.0
	ACACGC	5	1.43	6.06		0.56	0.25	6.0	5.5	7.0	1.5
	ACACTC	4	1.14	9.23		0.38	0.19	9.4	8.7	9.5	0.8
	ACAG	3	0.86	7.80		0.69	0.40	8.2	7.0	8.2	1.2
	ACC	1	0.29	5.00	(5.00, 1)	—	—	5.0	5.0	5.0	0.0
	ACTCG	1	0.29	5.80		—	—	5.8	5.8	5.8	0.0
	ACTGCG	1	0.29	9.50		—	—	9.5	9.5	9.5	0.0
	AG	24	6.86	9.06	(8.31, 24)	3.05	0.62	8.8	5.0	17.0	12.0
	AGC	45	12.86	9.16	(6.60, 34)	3.12	0.47	8.3	5.0	18.7	13.7
	AGCAGG	1	0.29	6.20		—	—	6.2	6.2	6.2	0.0
	AT	1	0.29	5.00	(5.00, 1)	—	—	5.0	5.0	5.0	0.0
	ATC	1	0.29	7.00	(7.00, 1)	—	—	7.0	7.0	7.0	0.0
D. pachea	A	8	2.91	20.75	(11.50, 2)	7.94	2.81	22.0	10.0	28.0	18.0
	AAAAG	1	0.36	6.40		—	—	6.4	6.4	6.4	0.0
	AAAT	1	0.36	5.00		—	—	5.0	5.0	5.0	0.0
	AAC	6	2.18	9.83	(5.43, 4)	4.71	1.92	9.0	5.0	17.0	12.0
	AAG	3	1.09	7.03	(5.70, 1)	1.15	0.67	7.7	5.7	7.7	2.0
	AATAGC	1	0.36	5.80		—	—	5.8	5.8	5.8	0.0
	AATTC	1	0.36	5.20	(5.20, 1)	—	—	5.2	5.2	5.2	0.0
	AC	150	54.55	15.21	(10.1, 172)	6.03	0.49	14.5	5.0	41.0	36.0
	ACACGC	4	1.45	7.00		0.73	0.36	6.9	6.3	8.0	1.7
	ACAG	1	0.36	5.80		—	—	5.8	5.8	5.8	0.0
	ACAGAT	1	0.36	6.20		—	—	6.2	6.2	6.2	0.0
	AG	21	7.64	12.24	(8.33, 21)	5.39	1.18	10.0	5.0	22.0	17.0
	AGAGCC	1	0.36	5.80	(5.80, 1)	—	—	5.8	5.8	5.8	0.0
	AGC	65	23.64	9.66	(7.96, 41)	3.55	0.44	8.7	5.0	18.7	13.7
	AGG	2	0.73	6.15		0.21	0.15	6.2	6.0	6.3	0.3
	AGGCGG	1	0.36	5.30	(5.30, 1)	—	—	5.3	5.3	5.3	0.0
	AT	3	1.09	7.50	(5.17, 3)	2.18	1.26	8.5	5.0	9.0	4.0
	ATC	3	1.09	7.33	(8.15, 2)	2.33	1.34	6.3	5.7	10.0	4.3
	C	2	0.73	13.50	(11, 1)	3.54	2.50	13.5	11.0	16.0	5.0
D. neotestacea	A	9	5.33	12.89	(12.6, 8)	2.85	0.95	12.0	10.0	19.0	9.0
	AAAC	1	0.59	7.20		—	—	7.2	7.2	7.2	0.0
	AAAT	1	0.59	5.00	(5.00, 1)	—	—	5.0	5.0	5.0	0.0
	AAC	17	10.06	9.31	(5.00, 1)	3.61	0.88	9.0	5.0	18.3	13.3
	AACAGC	1	0.59	8.30		—	—	8.3	8.3	8.3	0.0
	AAG	1	0.59	5.70		—	—	5.7	5.7	5.7	0.0
	AAT	1	0.59	5.70		—	—	5.7	5.7	5.7	0.0
	AC	97	57.40	11.69	(8.31, 97)	5.46	0.55	11.0	5.0	28.0	23.0
	ACAG	3	1.78	5.17		0.29	0.17	5.0	5.0	5.5	0.5
	ACAGCT	1	0.59	5.50		—	—	5.5	5.5	5.5	0.0
	ACAT	3	1.78	7.33		4.04	2.33	5.0	5.0	12.0	7.0
	AG	14	8.28	10.64	(7.50, 16)	5.30	1.42	10.5	5.0	23.5	18.5
	AGC	8	4.73	7.75	(6.00, 5)	2.82	1.00	6.7	5.3	13.7	8.4
	AGGG	1	0.59	5.20		—	—	5.2	5.2	5.2	0.0

Table 1
Continued

Species	Motif	*N*	% of Total	Mean[a]	(Perfect, *N*)[b]	SD	SE	Median	Min	Max	Range
	AT	3	1.78	5.50	(5.50, 3)	0.50	0.29	5.5	5.0	6.0	1.0
	ATC	8	4.73	7.54	(6.05, 6)	1.31	0.46	7.5	5.7	9.3	3.6
D. recens	A	4	4.44	15.75	(14.5, 2)	3.30	1.65	17.0	11.0	18.0	7.0
	AAAGG	1	1.11	5.20	(5.2, 1)	—	—	5.2	5.2	5.2	0.0
	AAC	10	11.11	8.84	(5.75, 1)	4.21	1.33	7.0	5.0	17.7	12.7
	AACAAG	1	1.11	5.70		—	—	5.7	5.7	5.7	0.0
	AACTC	1	1.11	5.80		—	—	5.8	5.8	5.8	0.0
	AATCAG	1	1.11	7.30		—	—	7.3	7.3	7.3	0.0
	AC	40	44.44	15.40	(7.75, 57)	7.62	1.20	15.0	5.0	41.5	36.5
	ACAG	3	3.33	7.60	(6.5, 1)	1.65	0.95	6.8	6.5	9.5	3.0
	ACAT	1	1.11	6.00	(6.00, 1)	—	—	6.0	6.0	6.0	0.0
	ACG	1	1.11	7.70		—	—	7.7	7.7	7.7	0.0
	ACTCTG	1	1.11	6.80		—	—	6.8	6.8	6.8	0.0
	ACTG	1	1.11	7.80	(7.70, 1)	—	—	7.8	7.8	7.8	0.0
	AG	6	6.67	9.50	(7.83, 3)	2.41	0.98	9.0	6.5	12.5	6.0
	AGATGC	1	1.11	7.50		—	—	7.5	7.5	7.5	0.0
	AGC	7	7.78	8.19	(6.05, 6)	3.31	1.25	6.7	5.3	14.7	9.4
	AGG	1	1.11	9.70		—	—	9.7	9.7	9.7	0.0
	AT	5	5.56	5.10	(5.10, 5)	0.22	0.10	5.0	5.0	5.5	0.5
	ATC	5	5.56	8.86	(7.30, 1)	2.28	1.02	8.3	6.3	11.7	5.4

NOTE.—Data only represent microsatellites ≥5 repeats long.

[a] Non-integer numbers represent partial repeats of a motif at the end of the allele (e.g., AGCAGCAGCAG = 3.67 repeats for AGC).

[b] Perfect alleles include only those microsatellites (or perfect runs within imperfect alleles) with no imperfections.

these species with each other directly while avoiding many potential biases associated with ''meta-analyses.'' As a result, we argue that comparisons across our five species are consistent in a relative manner.

Our isolation strategy and general comparative approach to understanding variation among species at microsatellites have a number of advantages over comparing data sets from numerous different studies. As mentioned,

Table 2
Microsatellite Motif Classes Recovered

Species	Motif Class	*N*	% of Total	Mean[a]	(Perfect, *N*)[b]	SD	SE	Median	Min	Max	Range
D. arizonae	Monomer	8	3.27	13.25	(12.3, 6)	3.24	1.15	13.0	10.0	18.0	8.0
	Dimer	114	46.53	12.80	(9.05, 126)	6.18	0.58	12.0	5.0	49.0	44.0
	Trimer	113	46.12	8.56	(6.93, 84)	2.57	0.24	7.7	5.0	18.3	13.3
	Tetramer	3	1.22	8.23	(6.00, 1)	1.75	1.01	8.2	6.5	10.0	3.5
	Pentamer	1	0.41	6.20		—	—	6.2	6.2	6.2	0.0
	Hexamer	6	2.45	6.98		2.09	0.85	6.2	5.2	10.5	5.3
D. mojavensis	Monomer	12	3.43	15.75	(12.6, 8)	9.33	2.69	13.0	10.0	43.0	33.0
	Dimer	258	73.71	13.34	(8.86, 301)	5.76	0.36	12.5	5.0	46.0	41.0
	Trimer	59	16.86	8.70	(6.62, 44)	2.96	0.39	8.0	5.0	18.7	13.7
	Tetramer	3	0.86	7.80		0.69	0.40	8.2	7.0	8.2	1.2
	Pentamer	2	0.57	5.50		0.42	0.30	5.5	5.2	5.8	0.6
	Hexamer	16	4.57	7.58	(5.50, 1)	2.09	0.52	6.6	5.3	11.2	5.9
D. pachea	Monomer	10	3.64	19.30	(11.3, 3)	7.73	2.45	16.0	10.0	28.0	18.0
	Dimer	174	63.27	14.72	(9.80, 196)	6.05	0.46	14.0	5.0	41.0	36.0
	Trimer	79	28.73	9.39	(7.71, 48)	3.56	0.40	8.7	5.0	18.7	13.7
	Tetramer	2	0.73	5.40		0.57	0.40	5.4	5.0	5.8	0.8
	Pentamer	2	0.73	5.80	(5.20, 1)	0.85	0.60	5.8	5.2	6.4	1.2
	Hexamer	8	2.91	6.39	(5.55, 2)	0.84	0.30	6.3	5.3	8.0	2.7
D. neotestacea	Monomer	9	5.33	12.89	(12.6, 8)	2.85	0.95	12.0	10.0	19.0	9.0
	Dimer	114	67.46	11.40	(8.13, 116)	5.44	0.51	10.8	5.0	28.0	23.0
	Trimer	35	20.71	8.34	(5.94, 12)	3.04	0.51	7.7	5.0	18.3	13.3
	Tetramer	9	5.33	6.10	(5.00, 1)	2.32	0.77	5.0	5.0	12.0	7.0
	Hexamer	2	1.18	6.90		1.98	1.40	6.9	5.5	8.3	2.8
D. recens	Monomer	4	4.44	15.75	(14.5, 2)	3.30	1.65	17.0	11.0	18.0	7.0
	Dimer	51	56.67	13.70	(7.55, 65)	7.60	1.06	13.0	5.0	41.5	36.5
	Trimer	24	26.67	8.64	(6.05, 11)	3.30	0.67	7.7	5.0	17.7	12.7
	Tetramer	5	5.56	7.32	(6.73, 3)	1.38	0.62	6.8	6.0	9.5	3.5
	Pentamer	2	2.22	5.50	(5.20, 1)	0.42	0.30	5.5	5.2	5.8	0.6
	Hexamer	4	4.44	6.83		0.81	0.40	7.1	5.7	7.5	1.8

NOTE.—Data only represent microsatellites ≥5 repeats long.

[a] Non-integer numbers represent partial repeats of a motif at the end of the allele (e.g., AGCAGCAGCAG = 3.67 repeats for AGC).

[b] Perfect alleles include only those microsatellites (or perfect runs within imperfect alleles) with no imperfections.

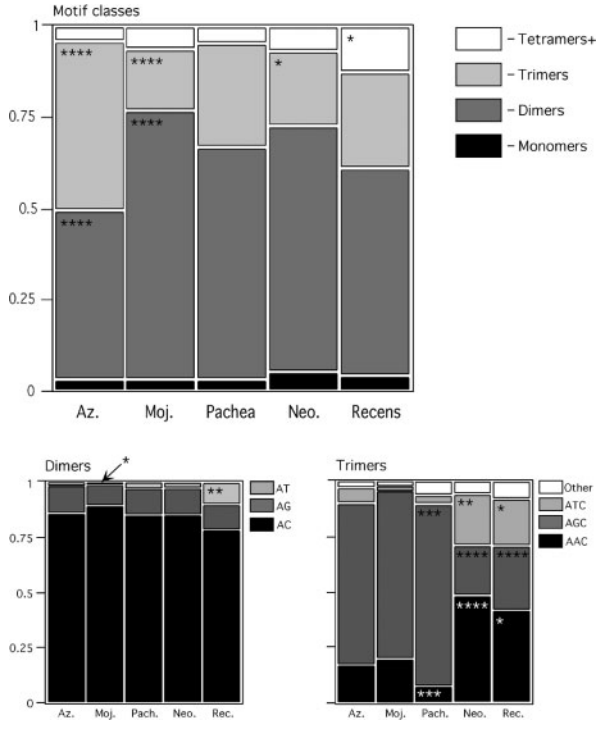

FIG. 2.—Relative abundance of motif classes, and motifs within dimers and trimers, for each species. Relative frequency for each level is measured within each species. Because relatively few microsatellites of higher motif classes were recovered, we grouped tetramers, pentamers, and hexamers into one class for analysis, ''tetramers+.'' Overall significance of motif class by species was tested using Pearson Chi square test: $\chi^2_{12} = 76.75$, $P < .0001$. Differences in specific motif class and motif abundance were tested using Fisher 2×2 exact tests. The Fisher test construction compares specific species/motif class categories (e.g., *D. arizonae*-trimers) with the combined values of all other species and the combined values of all other motif classes. Only those tests were performed where chi square values had high standardized residuals. Asterisks represent level of significance: $*P < .05$, $**P < .01$, $***P < .001$, $****P < .0001$.

by screening all five species at once, we have been able to standardize and control the molecular isolation procedure and provide an internal consistency to our data set. Also, the searching and scoring method of the TRF program employs a standardized criterion for identifying and scoring microsatellites by employing a minimum cutoff score based on local alignment of a pattern identified de novo. This approach is not based on an arbitrary criterion of how to accept imperfections in the microsatellite sequence. Consequently, repeated patterns are scored through a more realistic array of imperfections.

Furthermore, we used four separate hybridization reactions per species. These reactions served in part as a control for bias introduced by any one reaction. The four motifs that the reactions targeted (AC, CG, AGC, ATC) were most successfully retrieved from each motif's respective reaction (except CG loci, which were absent in every reaction), but the relative abundance of motifs and size classes were consistent in all reactions (data not shown). For example, AC was most successfully pulled out of the fly genomes with the AC hybridization reaction, but it was also the most successful dimer retrieved from all four reactions for each species, indicating that AC is highly abundant. Similarly, compared to other species, high numbers of trimers were retrieved from all *D. arizonae* hybridization reactions for all trimer motifs recovered and not just AGC and ATC motifs. Hence, our choice of hybridization reactions at worst may distort the frequencies of motifs, but it will not change our basic conclusions. Therefore, though we cannot completely rule out biases from undetected problems, it is unlikely that differences within and among our species are due to protocol biases.

Are Microsatellites Equally Represented in Species?

The abundance of motifs and motif classes is not equally represented across our five species or within the genus Drosophila. For example, *D. mojavensis* has over four times as many dimers (73.7%) as trimers (16.9%), yet *D. arizonae*, sister taxon of *D. mojavensis*, has almost identical equal proportions of the two—46.5% dimers and 46.1% trimers. Additionally, *D. recens* differs from other Drosophila in that it has a relatively high abundance of microsatellites with long motifs (tetramers-hexamers, 12.2%). The differences we find are pronounced, even between sister taxa, and they represent a genome-wide sampling of these species. Across species, these biases in motifs are consistent for each motif within a size class (table 1). For example, in *D. mojavensis* we recovered more loci for each dimer motif detected (AC, AG, AT) than in *D. arizonae*. The biases we find are present despite a presumed consistent base pair composition across Drosophila (Powell and Moriyama 1997). Consequently, the mechanisms behind these motif size class differences among species, and thus microsatellite evolution in general, must operate independently of the motif composition itself. Furthermore, these are not the same mechanisms that control genome-wide base pair composition, such as codon usage bias (Powell and Moriyama 1997).

Assuming our sampling is representative, these patterns may represent biases in the ''birth'' and ''death'' of microsatellite loci as a function of motif size class over relatively short periods in evolutionary history. If the evolution of length variation at microsatellite loci is controlled primarily by a balance between slippage rates at these loci and point mutations (Kruglyak et al. 1998), then the break up of repeated patterns caused by point mutations alone can be used to explain the ''death'' of the locus. However, it is difficult to understand how slippage rates may cause the ''birth'' of loci under a certain allele size threshold, because mutation rates due to slippage rates at these sites would be nearly equivalent to other, nonrepetitive sites when the number of repeats is low (Kruglyak et al. 2000; Calabrese, Durrett, and Aquadro 2001). Other factors, such as selection from upper size constraints (Garza, Slatkin, and Freimer 1995; Feldman et al. 1997) may be important as well. Regardless of the mechanism, the bias in motif size classes would require species-specific differences in mutation rates as a function of motif class and/or selection in these species.

Patterns of motif class abundance within the entire genus Drosophila reveal that dimers are the predominant motif size, and there is a general decrease in abundance as

Table 3
Kolmogorov-Smirnov Tests for Differences in Distribution Shape

Comparison	df	Count A	Count B	Max diff.	χ^2	*P* Value
Dimers vs. Dimers (pairwise across species)						
arizonae vs. mojavensis	2	114	258	0.106	3.531	0.3421
arizonae vs. pachea	2	114	174	0.178	8.774	0.0249
arizonae vs. neotestacea	2	114	114	0.158	5.684	0.1166
arizonae vs. recens	2	114	51	0.150	3.156	0.4128
mojavensis vs. pachea	2	258	174	0.140	8.155	0.0339
mojavensis vs. neotestacea	2	258	114	0.205	13.267	0.0026
mojavensis vs. recens	2	258	51	0.162	4.463	0.2147
pachea vs. neotestacea	2	114	174	0.257	18.169	0.0002
pachea vs. recens	2	174	51	0.217	7.431	0.0487
neotestacea vs. recens	2	114	51	0.197	5.447	0.1313
Trimers vs. Trimers (pairwise across species)						
arizonae vs. mojavensis	2	113	59	0.081	1.107	>.9999
arizonae vs. pachea	2	113	79	0.160	4.752	0.1858
arizonae vs. neotestacea	2	113	35	0.173	3.188	0.4063
arizonae vs. recens	2	113	24	0.233	4.314	0.2314
mojavensis vs. pachea	2	59	79	0.121	1.978	0.7438
mojavensis vs. neotestacea	2	59	35	0.128	1.436	0.9753
mojavensis vs. recens	2	59	24	0.189	2.426	0.5945
pachea vs. neotestacea	2	35	79	0.208	4.210	0.2437
pachea vs. recens	2	79	24	0.173	2.204	0.6645
neotestacea vs. recens	2	35	24	0.133	1.102	>.9999
Dimers vs. Trimers (pairwise within species)						
arizonae	2	114	113	0.446	45.163	<.0001
mojavensis	2	258	59	0.482	44.563	<.0001
pachea	2	174	79	0.456	45.202	<.0001
neotestacea	2	114	35	0.364	14.166	0.0017
recens	2	51	24	0.483	15.219	0.0010
All species	2	711	310	0.444	170.225	<.0001

NOTE.—Kolmogorov-Smirnov tests the maximum unsigned difference between two cumulative frequency distributions. It is sensitive to differences between two distributions in location, dispersion, skewness, and other shape parameters.

motif size increases, although hexamers tend to be more frequent than either tetramers or pentamers (Schug et al. 1998*b*; Toth, Gáspári, and Jurka 2000; Katti, Ranjekar, and Gupta 2001). This inverse relationship between motif size and abundance generally may be explained by higher slippage rates in dimers compared to other size classes (Kruglyak et al. 1998, 2000). In *Drosophila melanogaster*, for instance, dimers, trimers, and tetramers represent 66%, 30%, and 4% of microsatellites among these size classes, respectively (Schug et al. 1998*b*). Our five species are consistent with this trend in general, but we find considerable variation in motif class abundance for each species in particular (fig. 2). For instance, while *D. mojavensis*, *D. pachea*, *D. neotestacea*, and *D. recens* are similar to *D. melanogaster* in relative motif class abundance when only dimers, trimers and tetramers are considered, *D. arizonae* is not. Also, no other study of which we are aware has found the high abundance of loci with long motifs as *D. recens* in this study (fig. 2). This may be a function of the different approaches of different studies; whereas we had an aggressive protocol for identifying long microsatellite motifs in DNA sequences, most other studies do not look for motifs larger than trimers or tetramers. Thus, species within the genus Drosophila show a general trend in motif class abundance but significant variability among even closely related species.

Patterns of motif class abundance across Drosophila reflect those of other insects and many organisms in general in that motif classes are not equally represented across different taxa or even within the genome of any one species (Schug et al. 1998*b*; Bachtrog et al. 2000; Kruglyak et al. 2000). Birds and bats, for example, have fewer microsatellites than other animals, perhaps because of constraints of flight on genome size (Van Den Bussche, Longmire, and Baker 1995; Primmer et al. 1997). Primates in general show evidence of a large number of monomer loci (Jurka and Pethiyagoda 1995; Toth, Gáspári, and Jurka 2000); the human genome has over twice as many monomer microsatellites as dimers, but trimers are relatively rare (Katti, Ranjekar, and Gupta 2001). Rodents appear to have a low incidence of trimers compared to other species (Beckmann and Weber 1992; Stallings et al. 1994), but relatively high levels of dimers (Toth, Gáspári, and Jurka 2000). Vertebrates in general contain a large proportion of tetramers relative to other motif classes (Toth, Gáspári, and Jurka 2000). Overall, many species show a high incidence of dimers compared to other motif classes, but yeast and fungi seem to be notable exceptions to this general rule (Toth, Gáspári, and Jurka 2000). Hence, as with Drosophila, there appears to be considerable variation among organisms in motif class abundance.

As with motif classes, motifs within each class vary considerably in their abundance among organisms. To the extent that comparisons among studies using different procedures are valid, we can compare our results to those of Schug et al. (1998*b*), who examined microsatellite patterns in *D. melanogaster*. In *D. melanogaster*, AC is the

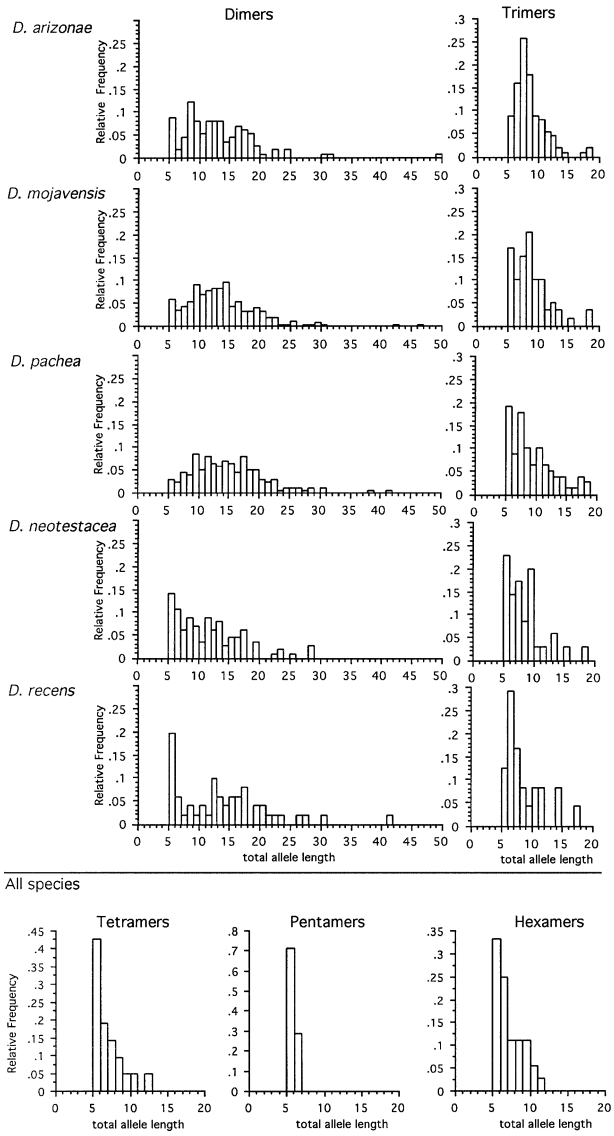

FIG. 3.—Frequency distribution of allele length for motif classes. Each graph shows the relative frequency of clones as a function of the number of repeats (total allele length) for each species. Tetramers, pentamers, and hexamers were grouped among all species. Total allele length includes allowed ''imperfections'' in the allele based on minimum alignment scores with a consensus pattern (see text for details).

most abundant dimer (49%), followed by AT (33%), AG (16%), and CG (1%), respectively (Schug et al. 1998*b*). The five species we studied also show an abundance of AC repeats (mean for five species = 85.4%) and a paucity of CG repeats, but the remaining dimers do not follow the same pattern as *D. melanogaster*, as AG is more common than AT (11.5% to 3.1%, respectively). As with *D. melanogaster*, AGC is the most frequent trimer in three of the five fly species considered here (46% for *D. melanogaster*, 77.3% = mean of *D. arizonae*, *D. mojavensis*, and *D. pachea*), but relatively low in the two mycophagous species (mean = 26% for *D. neotestacea* and *D. recens*). Likewise, AAC and ATC exhibit similar taxonomic patterns of abundance for *D. melanogaster* and the three cactophilic species; AAC occurs at moderate frequencies and ATC is rare. The two mycophagous species, however, deviate from these patterns. Finally, the tetramer AGAT is the most abundant motif found in *D. melanogaster* (Schug et al. 1998*b*), but ACAG is the most abundant for our five species. The comparison among the five species in this study and *D. melanogaster* demonstrates both similarities and differences among species within the Drosophila genus.

The generally high abundance of AC and AGC, and the rarity of CG, in Drosophila are consistent with a broader taxonomic pattern across many animals. However, the two mycophagous species of Drosophila are notable exceptions to this generality. Likewise, the frequencies of AGC and other trimers are not consistent across our study species or among Drosophila more generally. Other organisms also vary in motif abundance. In plants, the motifs AT and AG occur with relatively high frequency compared to AC, but the trimer AGC is less frequent (Lagercrantz, Ellegren, and Andersson 1993). In yeast (*Saccharomyces cerevisiae*), the motif AT is found at high frequency (Field and Wills 1998; Kruglyak et al. 2000). Numerous studies have shown that AC is the most abundant motif in many vertebrates and insects (Lagercrantz, Ellegren, and Andersson 1993; Toth, Gáspári, and Jurka 2000; Katti, Ranjekar, and Gupta 2001). In honeybees (*Apis mellifera*) and yellow jacket wasps (*Vespula rufa*), however, AG is the most abundant dimer (Estoup et al. 1993*b*; Thoren, Paxton, and Estoup 1995). These observations indicate that there is extensive variation in motif abundance across taxa despite some broad similarities.

It should be noted that the large frequency of AC and AGC in the five species in this study might be inflated because these motifs were targets for two of our enrichment reactions. The rarity of CG, which also was an enrichment target, as well as the lower abundance of AGC in *D. neotestacea* and *D. recens*, cannot be explained by this bias. Though the rarity of CG might be explained by the tendency of CG oligonucleotide probes to form hairpin structures during hybridizations, thus leading to the recovery of few microsatellites, we believe there is a real paucity of CG microsatellites in Drosophila for three reasons. First, we recovered many microsatellite loci with our CG hybridizations (data not shown), indicating that the CG oligonucleotide is capable of operating as a probe for microsatellites. Second, whereas we recovered many AC loci using the CG probe (and the other probes as well), which suggests that AC is an abundant microsatellite in Drosophila genomes, we recovered no CG loci with any of the probes. Finally, when large amounts (100–1,000 kb) of the *D. melanogaster* genome from the Drosophila Genome Project (http://www.fruitfly.org/) are assayed for microsatellites with our identification protocol, AC is abundant and CG is rare (data not shown). These large sequenced sections of the Drosophila genome should be unaffected by an enrichment bias with respect to microsatellite motif. Thus, AC appears to be very common in the entire genus Drosophila and CG is rare. It is clear that species exhibit differences in motif and motif class abundance, but it is as yet not understood why this abundance varies across species. Though a number of hypotheses may explain

Table 4
Kruskal-Wallis Test for Differences in Allele Length

Level	Count	Score Sum	Score Mean	(Mean-Mean0)/Std0	
Monomer	43	33819	786.488	4.543	
Dimer	711	470761.5	662.112	13.056	
Trimer	310	119082	384.135	−11.47	
Tetramer	22	4710	214.091	−5.099	
Pentamer	7	750	107.143	−3.727	
Hexamer	36	8762.5	243.403	−6.016	
			1-way Test, Chi-Square Approximation		
			Chi square	df	Prob > Chi square
			252.7515	5	<.0001

NOTE.—Kruskal-Wallis test for overall differences in allele length among motif classes.

some of this variation (base composition, for example), none is sufficient to explain many of the differences among species.

Microsatellite motif classes are not equally represented across chromosomes and are highly overrepresented in noncoding regions (Hancock 1995; Bachtrog et al. 2000; Katti, Ranjekar, and Gupta 2001). Trimers, and to a lesser extent hexamers, are the exception to this observation, presumably because trimers can integrate into reading frames in exons. For trimers, then, codon bias may be important in determining the abundance of motifs within this class. In Drosophila, which exhibits a general bias toward G and C at synonymous sites (Powell and Moriyama 1997), this could favor the abundance of certain trimers, such as AGC, over others, such as ATC, especially if the trimer motifs are out of phase with gene codons such that G is in the more commonly synonymous third position. We have not localized our microsatellite loci within the genomes of our five species, but the high frequency of trimers in *D. arizonae* is curious. Locating these loci, and loci in the sister taxon, *D. mojavensis*, may prove useful for understanding the differences in microsatellite representation within the genome of species, especially because these species are closely related (divergence estimated at ~1 MYA) yet they show markedly different frequencies of dimers and trimers.

Do Microsatellites Vary in Length Among Species?

Allele length, or the number of repeats, comprises the operational unit of variation at microsatellite loci (though see Ortí, Pearse, and Avise 1997). Unfortunately, length variation can be difficult to compare among taxa and across studies because of inherent biases in the collection and interpretation of these data. Nevertheless, certain trends are apparent. For example, fish appear to have long microsatellites (Estoup et al. 1993*a*; Brooker et al. 1994; Neff and Gross 2001), ''simple'' organisms contain a relatively large number of long trimers (Field and Wills 1996), and mammals and insects appear to have a wide range of average lengths (Beckmann and Weber 1992; Stallings et al. 1994; Schug et al. 1998*b*).

In Drosophila, a number of studies have shown that microsatellites are shorter than in either mammals or fish (Hutter, Schug, and Aquadro 1998; Kruglyak et al. 1998; Schug et al. 1998*b*; Bachtrog et al. 2000). Microsatellites in *D. melanogaster* appear to be especially short in length, and this effect is consistent across all motif classes. Drosophila *subobscura*, a distantly related species to *D. melanogaster* within the subgenus *Sophophora*, contains longer microsatellites (Pascual, Schug, and Aquadro 2000), although other species in the obscura group (*D. pseudoobscura* and *D. persimilis*) do not (Noor, Schug, and Aquadro 1999). In the subgenus Drosophila, microsatellites in *D. virilis* are reported to be long on average (Schlötterer and Harr 2000). Compared to variation in other Drosophilids, considerable variation in size lengths exists among our study species, even though only *D. pachea* and *D. neotestacea* differ from each other significantly ($P = .0003$, Scheffe post-hoc contrast). *Drosophila pachea's* average dimer length, 14.7 repeats, is similar to that of other Drosophila species with long microsatellites. *Drosophila neotestacea* has relatively short dimers (mean = 11.4) and more closely approximates lengths in *D. melanogaster*. Thus, Drosophila harbors marked length variation both within and between its two subgenera, and many Drosophila species, but not all, do indeed have shorter microsatellites than other groups, such as mammals and fish.

Table 5
Pairwise Comparisons of Mean Allele Length

Abs(Dif)-LSD	Monomer	Dimer	Trimer	Tetramer	Pentamer	Hexamer
Monomer		2.20572	6.73066	8.67526	9.82591	8.41718
Dimer	−0.15597		4.52494	6.46954	7.62019	6.21146
Trimer	4.28341	3.5014		1.9446	3.09525	1.68652
Tetramer	4.73325	3.21408	−1.37345		1.15065	0.25808
Pentamer	3.69665	1.90823	−2.65261	−5.37533		1.40873
Hexamer	5.01988	3.64236	−0.96144	−3.81158	−4.8034	

NOTE.—Pairwise comparisons among motif size classes. Above diagonal are differences in means, below diagonal are Tukey-Kramer HSD. Positive values for Tukey-Kramer HSD show pairs of means that are significantly different.

Variation in allele length within and among species may have important implications. Microsatellite length may be influenced by size-dependent slippage rates (Kruglyak et al. 2000), effective population size and population structure (Slatkin 1995), genome structure and size (Hancock 1996), and selection on length itself (Garza, Slatkin, and Freimer 1995). Across species with different average microsatellite lengths, these hypotheses must propose that any or all of these factors must differ across species as well. Unfortunately, a circularity develops because these hypotheses estimate values, such as slippage rates, as a result of observed patterns of microsatellite variation in organisms, so any differences among species must imply differences in, for example, size-dependent slippage rates or selection on length. Independent assessments of these mechanisms and characters, as by experimentally testing slippage rates of polymerases in different species, would prove beneficial. Though the Drosophila in this study likely have greater population structure and smaller effective population sizes than a cosmopolitan, generalist species such as *D. melanogaster*, we cannot at this time evaluate the correlation of these factors with length variation in species. As a result, it is still unclear whether these explanations can fully account for length variation among Drosophila or among larger groups.

Are Size Frequency Distributions Similar Across Taxa?

We found extensive variation in frequency distribution shape across motif classes, but only occasional differences across species. Differences in slippage and mutation rates may explain much but not all of the variation among size classes in distribution shapes (Kruglyak et al. 2000). It is interesting that many repeat size histograms from database searches fit exponential distributions (Kruglyak et al. 1998; Schug et al. 1998*a*; Bachtrog et al. 1999; Kruglyak et al. 2000), but many histograms of allele sizes in natural populations are less skewed and have their modes shifted away from the smallest repeat size (see Pascual, Schug, and Aquadro 2000, for example). All of the distributions for our species, which resulted from isolating microsatellites from a cloned library, show a pattern similar to those from natural populations of the species, though the mean length for the clones is slightly less (data not shown). This similarity suggests that clones from libraries may be retrieving a representative sample of the size variation that exists in a population × genome space (i.e., alleles summed across individuals and loci) better than database searches.

The Tempo and Mode of Microsatellite Evolution

The differences among higher taxonomic groups in microsatellite composition and allele length may be explained by a number of processes. Microsatellite patterns may evolve very slowly and steadily through time, so that most or all of the species within a particular taxonomic group will have similar microsatellite patterns. Alternatively, microsatellite patterns may evolve fairly rapidly over short evolutionary time scales but be subject to certain taxon-specific mutational or selective constraints that keep variation in these patterns within certain bounds. This would be similar to variation in GC content among major groups of organisms: whereas individual nucleotide sites experience ongoing mutational changes, there is a strong phylogenetic component to genome-wide GC content (Mooers and Holmes 2000). For microsatellites, Kruglyak et al. (1998) showed that differences in average allele length may be explained by differential mutation rates across species and motif size classes (dimers, trimers, etc.), but the question remains why mutation rates vary among loci and species. Likewise, other explanations for microsatellite variation suffer the same shortfall.

Here, we have shown that these mechanisms of microsatellite evolution may be quite labile over relatively short evolutionary periods, but still show patterns of consistency over large taxonomic groups and, presumably, long time periods. Studies of microsatellite variation in several other Drosophila species also support the latter possibility, i.e., that microsatellites evolve quickly but are subject to taxon-specific constraints. For instance, species in the *obscura*, *melanogaster*, and *virilis* groups have been found to differ in the average length of microsatellite repeats (Hutter, Schug, and Aquadro 1998; Schug et al. 1998*a*; Pascual, Schug, and Aquadro 2000; Schlötterer and Harr, 2000; Noor, Kliman, and Machado 2001). Yet, the entire genus Drosophila has shorter microsatellites than other taxa. These mechanisms of microsatellite variation being dependent on ''phylogenetic scale'' are apparent for both mutation rates and birth/death of microsatellites.

The Utility of Comparative Studies and the Issues of Comparing Data Across Studies

As previously mentioned, we have used a comparative approach to understanding microsatellite evolution. With our protocol, we believe we gain a powerful genome-wide view of microsatellite variation within and among our five species. Comparisons of our results with those from other studies, however, require more scrutiny. Despite this caution, our data clearly show consistencies with other Drosophila and with phylogenetically more distant groups. The abundance of AC (high) and CG (low), for example, is generally a consistent rule across most animals, even though notable exceptions exist. Similarly, the greater average length of dimers over trimers appears to have broad phylogenetic consistency. In contrast to these observations, some attributes of microsatellites apparently have a very low phylogenetic signal. The proportion of dinucleotide and trinucleotide loci, for example, is not conserved even between the closely related sister taxa *D. arizonae* and *D. mojavensis*. Similarly, the average length of microsatellites within motif classes is not conserved across species at certain phylogenetic scales (within Drosophila, for example), but does follow general patterns across larger groups (all fish have long microsatellites).

It is noteworthy that some aspects of microsatellite variation are conserved over larger taxonomic scales. As microsatellites have mutation rates ranging between 10^{-2} and 10^{-6} (Dallas 1992; Ellegren 1995; Schug et al. 1998*b*),

some attributes, such as allele length, should quickly lose any phylogenetic signal. Consequently, unless the mutational process itself is phylogenetically constrained, there should be no real long-term historical pattern to this aspect of microsatellite evolution. The fact that large taxonomic groups (such as Drosophila or fish, see also Zhu, Queller, and Strassmann 2000; Wilder and Hollocher 2001; Noor, Kliman, and Machado 2001) exhibit trends in microsatellite characteristics suggests that this variation must reflect something more than simple slippage events at microsatellite sequences. Kruglyak et al. (2000) suggest that differences in equilibrium microsatellite lengths among species are due to different species-specific slippage rates.

The searching and scoring protocol we used in this study, namely our UniqFlank program and Tandem Repeats Finder (Benson 1999), yielded two unusual results compared to other microsatellite surveys. First, we identified a relatively high number of large motifs (pentamers and hexamers), and a large abundance of hexamer microsatellite sequences in our species. These results seem counterintuitive based on the trend toward decreasing frequencies of microsatellites as class size increases. However, this result is not unprecedented in Arthropods (Toth, Gáspári, and Jurka 2000) and may represent the higher acceptance of these microsatellites in coding regions. For our data, the high frequency of hexamers is not an artifact of the upper boundary limit of size classes in our search. When our search is reset to look for repeated patterns up to 7, 10, or 50 nucleotides long, hexamers consistently are found at relatively high frequencies, and heptamers are always very rare (data not shown). Presumably, many studies have not reported higher size classes of microsatellites because they have not been the focus of searches.

Second, we occasionally found patterns at some loci where point mutations or other imperfections in the microsatellite sequence are regularly repeated every few repeats. These observations represent a larger pattern embedded within the microsatellite sequence of a smaller motif. These ''nested'' patterns of variation may be useful in understanding the evolution of microsatellites in general and reconstructing the ''phylogenetic'' history of the microsatellite sequence in particular (Benson and Dong 1999). Nested patterns such as these suggest a number of possible explanations: (1) slippage events greater than ± 1; (2) other modes of changing allele size such as uneven crossing-over, mini-duplications, or indels; or (3) a complex evolutionary/populational history of the microsatellite locus.

As the numbers of microsatellite studies increase, the relationship between microsatellite length and the means by which they become isolated is becoming apparent. Admittedly, data sets from different sources (cloned libraries, database searches, natural populations) actually represent different samples of microsatellite variation in the genome. Cloned libraries, for example, sample on average one allele from numerous loci, whereas surveys of natural populations sample many alleles from one or a few loci. Additionally, database searches are usually biased in favor of coding regions or at least euchromatin. Differences in allele lengths among data sets are documented in numerous species (for Drosophila, see examples in Pascual, Schug, and Aquadro 2000; Schlötterer and Harr 2000), and data from database searches and cloned libraries consistently have lower and higher allele lengths, respectively, than surveys of natural populations. Preliminary data from natural populations of our five species suggest mean allele lengths slightly lower than what we report here based on library clones (data not shown). These trends suggest that results from different sources are comparable if these differences are taken into account.

To test the consistency of the results from our approach with other studies, we gathered 590 previously isolated *D. melanogaster* sequences (Schug et al. 1998*b*; sequences downloaded from C. Aquadro's Web site [http://www.mbg.cornell.edu/aquadro/microsatellite.html]). We then analyzed this ''database search'' using our identification and scoring protocol (data not shown). Our analysis of these sequences gives microsatellite lengths very similar to those in other studies (Kruglyak et al. 1998; Schug et al. 1998*b*; Bachtrog et al. 2000). Accordingly, we interpret microsatellite lengths in this survey as reflecting true similarities and differences among species in other studies.

In summary, we have found extensive differences and also striking similarities among microsatellites within our five species in the subgenus Drosophila, as well as within the genus Drosophila as a whole. Thus, the strength of the phylogenetic signal evident in patterns of microsatellite variation depends on the features (e.g., relative abundance of size classes and motifs, frequency distribution of microsatellite lengths) under consideration. Such patterns of microsatellite variability may provide a window into the forces that mold genome evolution as a whole in these and other organisms.

Acknowledgments

We thank M. Goodisman for providing productive comments on the manuscript; A. Holyoake for valuable technical advice and assistance; M. Ruth and Cori Routh for laboratory help; and G. Benson for providing helpful assistance with TRF and a new version of the program. This work was supported by a National Science Foundation (NSF) fellowship to C.L.R. under the Interdisciplinary Research Training Group on Plant-Insect Interactions (DBI-9602249). Additionally, three NSF grants to T.A.M. (INT-9402161, DEB-9510645, and DEB-0075312) provided support for this project. Both J.J. and K.A.D. were supported by NSF grant DEB-0074141.

Literature Cited

Bachtrog, D., M. Agis, M. Imhof, and C. Schlötterer. 2000. Microsatellite variability differs between dinucleotide repeat motifs: evidence from *Drosophila melanogaster*. Mol. Biol. Evol. **17**:1277–1285.

Bachtrog, D., S. Weiss, B. Zangerl, G. Brem, and C. Schlötterer. 1999. Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. Mol. Biol. Evol. **16**: 602–610.

Beckmann, J. S., and J. L. Weber. 1992. Survey of human and rat microsatellites. Genomics **12**:627–631.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27**:573–580.

Benson, G., and L. Dong. 1999. Reconstructing the duplication history of a tandem repeat. Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology **ISMB-99**:44–53.

Beverley, S. M., and A. C. Wilson. 1984. Molecular evolution in Drosophila and the higher Diptera. II: a time scale for fly evolution. J. Mol. Evol. **21**:1013.

Brooker, A. L., D. Cook, P. Bentzen, J. M. Wright, and R. W. Doyle. 1994. Organization of microsatellites differs between mammals and cold-water teleost fishes. Can. J. Fish. Aquat. Sci. **51**:1959–1966.

Calabrese, P. P., R. T. Durrett, and C. F. Aquadro. 2001. Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. Genetics **159**:839–852.

Dallas, J. F. 1992. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. Mamm. Genome **3**: 452–456.

Degtyareva, N. P., P. Greenwell, E. R. Hofmann, M. O. Hengartner, L. Zhang, J. G. Culotti, and T. D. Petes. 2002. *Caenorhabditis elegans* DNA mismatch repair gene msh-2 is required for microsatellite stability and maintenance of genome integrity. Proc. Natl. Acad. Sci. USA **99**:2158–2163.

Ellegren, H. 1995. Mutation rates at porcine microsatellite loci. Mamm. Genome **6**:376–377.

Estoup, A., P. Presa, F. Kreig, D. Vaiman, and R. Guymard. 1993*a*. $(CT)_n$ and $(GT)_n$ microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). Heredity **71**: 488–496.

Estoup, A., M. Solignac, M. Harry, and J. M. Cornuet. 1993*b*. Characterization of $(GT)_n$ and $(CT)_n$ microsatellites in two insect species: *Apis melifera* and *Bombus terrestris*. Nucleic Acids Res. **21**:1427–1431.

Ewing, B., L. Hillier, M. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8**:175–185.

Feldman, M. W., A. Bergman, D. D. Pollock, and D. B. Goldstein. 1997. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. Genetics **145**:207–216.

Fellows, D. P., and W. B. Heed. 1972. Factors affecting host plant selection in desert-adapted cactophilic Drosophila. Ecology **53**:850–858.

Field, D., and C. Wills. 1996. Long, polymorphic microsatellites in simple organisms. Proc. R. Soc. Lond. Ser. B. **263**: 209–215.

———. 1998. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. Proc. Natl. Acad. Sci. USA **95**:1647–1652.

Garza, J. C., M. Slatkin, and N. B. Freimer. 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraint on allele size. Mol. Biol. Evol. **12**: 594–603.

Goldstein, D., and C. Schlötterer. 1999. Microsatellites: evolution and applications. Oxford University Press, New York.

Grimaldi, D., A. C. James, and J. Jaenike. 1992. Systematics and modes of reproductive isolation in the holarctic *Drosophila testacea* species group (Diptera: Drosophilidae). Ann. Entomol. Soc. Am. **85**:671–685.

Gryfe, R., C. Swallow, B. Bapat, M. Redston, S. Gallinger, and J. Couture. 1997. Molecular biology of colorectal cancer. Curr. Probl. Cancer **21**:238.

Hamilton, M. B., E. L. Pincus, A. DiFiore, and R. C. Fleischer. 1999. BioTechniques **27**:500–507. Protocol may be found at: http://www.georgetown.edu/faculty/hamiltm1/lab.html.

Hancock, J. M. 1995. The contribution of slippage-like processes to genome evolution. J. Mol. Evol. **41**:1038–1047.

———. 1996. Simple sequences and the expanding genome. BioEssays **18**:421–425.

Harr, B., and C. Schlötterer. 2000. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. Genetics **155**:1213–1220.

Heed, W. B., and R. L. Mangan. 1986. Community ecology of the Sonoran Desert USA Mexico Drosophila. Pp. 311–346 *in* M. Ashburner, H. L. Carson, and J. N. Thompson, Jr., eds. The genetics and biology of Drosophila, Vol. 3E. Academic Press, Orlando, Fla.

Hutter, C. M., M. D. Schug, and C. F. Aquadro. 1998. Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: a reciprocal test of the ascertainment bias hypothesis. Mol. Biol. Evol. **15**:1620–1638.

Jaenike J. 1978. Resource predictability and niche breadth in Drosophila Quinaria species group. Evolution **32**:676–678.

Jurka, J., and C. Pethiyagoda. 1995. Simple receptive DNA sequences from primates: compilation and analysis. J. Mol. Evol. **40**:120–126.

Katti, M. V., P. K. Ranjekar, and V. S. Gupta. 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Biol. Evol. **18**:1161–1167.

Kruglyak, S., R. T. Durrett, M. D. Schug, and C. F. Aquadro. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc. Natl. Acad. Sci. USA **95**:10774–10778.

———. 2000. Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. Mol. Biol. Evol. **17**:1210–1219.

Lagercrantz, U., H. Ellegren, and L. Andersson. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. **21**:1111–1115.

Markow, T. A., and G. D. Hocutt. 1998. Reproductive isolation in Sonoran Desert Drosophila: testing the limits of the rules. Pp. 234–244 *in* D. J. Howard and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford University Press, New York.

Miller, S., and E. Myers. 1999. The FAKtory DNA sequence fragment assembly system. Technical Report 99–03, Department of Computer Science, University of Arizona, Tucson. 85721. Available at http://www.cs.arizona.edu/research/reports.html.

Mooers A. O. and E. C. Holmes. 2000. The evolution of base composition and phylogenetic inference. Trends Ecol. Evol. **15**:365–369.

Neff, B. D., and M. R. Gross. 2001. Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. Evolution **55**:1717–1733.

Noor, M. A. F., R. M. Kliman, and C. A. Machado. 2001. Evolutionary history of microsatellites in the Obscura group of Drosophila. Mol. Biol. Evol. **18**:551–556.

Noor, M. A. F., M. D. Schug, and C. F. Aquadro. 1999. Microsatellite variation in populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. Genet. Res. **75**:25–35.

Ortí, G., D. E. Pearse, and J. C. Avise. 1997. Phylogenetic assessment of length variation at a microsatellite locus. Proc. Natl. Acad. Sci. USA **94**:10745–10749.

Pascual, M., M. D. Schug, and C. F. Aquadro. 2000. High density of long dinucleotide microsatellites in *Drosophila subobscura*. Mol. Biol. Evol. **17**:1259–1267.

Pitnick, S. 1993. Operational sex ratios and sperm limitation in populations of *Drosophila pachea*. Behav. Ecol. Sociobiol. **33**:383–391.

Pitnick, S., G. S. Spicer, and T. Markow. 1997. Phylogenetic examination of female incorporation of ejaculate in Drosphila. Evolution **51**:833–845.

Powell, J. R. 1997. Progress and prospects in evolutionary biology: the Drosophila model. Oxford University Press, New York.

Powell, J. R., and E. N. Moriyama. 1997. Evolution of codon usage bias in Drosophila. Proc. Natl. Acad. Sci. USA **94**: 7784–7790.

Primmer, C. R., T. Raudsepp, G. P. Chowdhary, A. P. Møller, and H. Ellegren. 1997. Low frequency of microsatellites in the avian genome. Genome Res. **7**:471–482.

Ruiz, A., and W. B. Heed. 1978. Host plant specificity in the cactophilic *Drosophila mulleri* species complex. J. Anim. Ecol. **57**:237–249.

Ruiz, A., W. B. Heed, and M. Wasserman. 1990. Evolution of the *mojavensis* species cluster of cactophilic Drosophila, with descriptions of two new species. J. Hered. **81**:30–42.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schlötterer, C., and B. Harr. 2000. *Drosophila virilis* has long and highly polymorphic microsatellites. Mol. Biol. Evol. **17**:1641–1646.

Schug, M. D., C. M. Hutter, K. A. Wetterstrand, M. S. Gaudette, T. F. C. Mackay, and C. F. Aquadro. 1998*a*. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. Mol. Biol. Evol. **15**:1751–1760.

Schug, M. D., K. A. Wetterstrand, M. S. Gaudette, R. H. Lim, C. M. Hutter, M. S. Gaudette, and C. F. Aquadro. 1998*b*. The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. Mol. Ecol. **7**:57–70.

Shoemaker D. D. W., and J. Jaenike. 1997. Habitat continuity and the genetic structure of Drosophila populations. Evolution **51**:1326–1332.

Sia, E. A., S. Jinks-Robertson, and T. D. Petes. 1997. Genetic control of microsatellite stability. Mutat. Res.—DNA Repair **383**:61–70.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics **139**: 457–462.

Stallings, R. L., A. F. Ford, D. Nelson, D. C. Torney, C. E. Hildebrand, and R. K. Moyzis. 1994. Evolution and distribution of $(AC)_n$ repetitive sequences in mammalian genomes. Genomics **10**:807–815.

Thoren, P. A., R. J. Paxton, and A. Estoup. 1995. Unusually high frequency of $(CT)_n$ and $(GT)_n$ microsatellite loci in a yellow-jacket wasp, *Vespula rufa* (L.) (Hymenoptera: Vespidae). Insect Mol. Biol. **4**:141–148.

Toth, G., Z. Gáspári, and J. Jurka. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. **10**:967–981.

Van Den Bussche, R. A., J. L. Longmire, and R. J. Baker. 1995. How bats achieve a small C-value: frequency of repetitive DNA in Macrotus. Genome **6**:521–525.

Wilder, J., and H. Hollocher. 2001. Mobile elements and the genesis of microsatellites in dipterans. Mol. Biol. Evol. **18**: 384–392.

Zhu, Y., D. C. Queller, and J. E. Strassmann. 2000. A phylogenetic perspective on sequence evolution in microsatellite loci. J. Mol. Evol. **50**:324–338.

Stephen Palumbi, Associate Editor

Accepted March 31, 2003