Polymorphism and Locus-Specific Effects on Polymorphism at Microsatellite Loci in Natural *Drosophila melanogaster* Populations

Christian Schlötterer,*,†,‡ Claus Vogl‡ and Diethard Tautz*

**Zoologisches Institut der Universität München, 80333 München, Germany, †Department of Ecology and Evolution, University of Chicago, 60637 Chicago, Illinois and ‡Veterinärmedizinische Universität Wien, Institut für Tierzucht und Genetik, Josef Baumann Gasse 1, 1210 Wien, Austria*

Manuscript received August 13, 1996
Accepted for publication February 11, 1997

ABSTRACT

We have studied the natural variation at microsatellite loci in two African and five non-African populations of *Drosophila melanogaster*. Ten dinucleotide simple sequence loci were cloned from chromosomally mapped P1 clones and typed for single individuals from isofemale lines of the respective populations. We find that the African populations harbor the largest degree of diversity, while the non-African populations show a lower diversity. This supports previous results that *D. melanogaster* originated in Africa and spread across the rest of the world in historic times. Using genetic distance measures, we find also a distinct population subdivision between the non-African populations. Most interestingly, we find for some loci in some populations a strongly reduced variability, which cannot be explained by bottleneck effects. Employing a conservative test based on the variance in repeat number, we find that at least one locus in one population deviates significantly from the expectations of mutation-drift equilibrium. We suggest that this may be due to a recent selective sweep in this chromosomal region that may have been caused by a linked locus that was involved in local adaptation of the population.

NATURAL variation is the result of the interaction between selection and random drift. Correct interpretation of observed variability depends on a profound understanding of both parameters. The elaborate genetic and physical map of *Drosophila melanogaster* has provided an important step toward understanding genetic variability. It allows one to systematically asses the consequences of recombination rate variation on observed variability. Several studies demonstrated a positive correlation between recombination rate and variability. This phenomenon can be explained by either positive (hitchhiking) or negative (background selection) selection on a locus near the chromosomal region under study (MAYNARD SMITH and HAIGH 1974; CHARLESWORTH *et al.* 1993). With decreasing levels of recombination the selected and studied locus will be in linkage disequilibrium over larger physical distances. Consequently, loci located in low recombining regions will have a higher chance to be affected by selective sweeps and background selection, explaining their reduced variability.

In contrast to these "long range" effects in regions of low recombination, selection has a more local effect in regions of normal to high recombination. The extent of the genomic region in linkage disequilibrium with the selected site depends on the interplay of recombination rates and selection coefficients. Only a limited number of studies measuring genetic variability in correlation to known selection regimes has been described so far. Recently, TAYLOR *et al.* (1995) have used Pyrethroid insecticide resistance to compare patterns of variability at this locus to a genomic region not involved in insecticide resistance. The authors showed that directional selection has significantly reduced allelic diversity at the selected locus by removing uncommon alleles, with the unlinked control locus not being affected. While this study illuminates how selection may affect natural variation under a known selection regime, it has to be emphasized that similar approaches may not always be applicable. More complex selection regimes, such as natural selection involved in adaptation to a local habitat, preclude the *a priori* knowledge of appropriate candidate loci to study the consequences on natural variation.

With the progress of several genome projects, it has become clear that there is a significant fraction of genes with no homology and no known function. Hence, studies that are based on known candidate loci in organisms without a fully sequenced genome will miss these genes. An alternative approach may be to take many randomly chosen genomic regions and to measure their genetic variation in populations exposed to different environments. As a sequencing-based approach is still too expensive and time consuming, we employed hypervariable microsatellite markers, which are distributed across the genome and relatively easy to score.

In the past years microsatellites have become the marker of choice for many applications. Their abun-

Corresponding author: Christian Schlötterer, Veterinärmedizinische Universität Wien, Institut für Tierzucht und Genetik, Josef Baumann Gasse 1, 1210 Wien, Austria. E-mail: schlotc@hp01.boku.ac.at

dance, high polymorphism and codominant inheritance has facilitated their extensive use in behavioral ecology (QUELLER *et al.* 1993; SCHLÖTTERER and PEMBERTON 1994), genomic mapping (DIB *et al.* 1996; DIETRICH *et al.* 1996), phylogenetic inference (GOLDSTEIN *et al.* 1995) and population genetics (BOWCOCK *et al.* 1994; ESTOUP *et al.* 1995; VIARD *et al.* 1996).

The major mutation mechanism of microsatellites is presumably DNA slippage (TAUTZ and SCHLÖTTERER 1994). *In vitro* studies showed that slippage is more efficient with short repeat motifs (mono- or dinucleotide motifs) than with longer ones (SCHLÖTTERER and TAUTZ 1992). However, most slippage events occurring in microsatellite regions are apparently repaired by the mismatch repair system (STRAND *et al.* 1993). This, in turn, may be less efficient on longer repeat motifs than on shorter ones, thus counteracting the effects of the primary slippage rate. Accordingly, tri- and tetranucleotide-bearing microsatellites may be similarly polymorphic as those carrying dinucleotide repeats. Another effect that might play a role is the average length of the repeat unit. In a first survey of human microsatellite loci WEBER (1990) found that longer motifs tend to more variability than shorter ones. However, other studies in humans could not confirm this inference (BECKMANN and WEBER 1992; VALDES *et al.* 1993). A survey on *D. melanogaster* microsatellites found again no correlation between average repeat length and polymorphism, but suggested that there is a correlation between the variance of the repeat count and the maximal length (GOLDSTEIN and CLARK 1995). Since the variance may be a more sensitive indicator of mutation rate, they suggest that the mutation rate may indeed increase with repeat count, though possibly not smoothly, which might explain the differences in the findings between different studies.

Despite still existing uncertainties of the exact mutational forces acting on microsatellites, their mutation process can be described to a very good approximation with the stepwise mutation model (OHTA and KIMURA 1973; VALDES *et al.* 1993). Within the framework of the stepwise mutation model, test statistics can be derived to detect deviations from the neutral model.

As most microsatellites are probably neutral, no deviation from the expectations under the neutral model are expected by selection acting on the microsatellite itself. However, if a microsatellite locus is linked to a genomic region, which is the target for natural selection, then this microsatellite will show deviation from the neutral expectation. Given that microsatellite loci appear to be distributed across the whole genome in eukaryotes (DIB *et al.* 1996; DIETRICH *et al.* 1996), it seems likely that at least some of them are occasionally affected by hitchhiking events. In general, hitchhiking effects will increase the frequency of a linked microsatellite allele in the population. However, as microsatellites tend to have high mutation rates, new mutations will be accumulated after a hitchhiking event relatively quickly. As soon as the microsatellite variance has been restored by new mutation at the respective microsatellite, it will not be possible to detect this hitchhiking event by microsatellite analysis. Therefore, microsatellites should be sensitive indicators of very recent hitchhiking events only.

We have set out to test this prediction by analyzing microsatellite polymorphisms in natural *D. melanogaster* populations. *D. melanogaster* originated in Africa and colonized other regions of the world only ~10,000–15,000 years ago (DAVID and CAPY 1988; BÉNASSI and VEUILLE 1995). This would imply that *D. melanogaster* was exposed to very different selection regimes while colonizing climatically and ecologically diverse regions. Thus, we would expect to be able to demonstrate traces of local adaptation events, *i.e.,* hitchhiking of microsatellite alleles, in the genome of *D. melanogaster.*

MATERIALS AND METHODS

Population samples: Flies were collected in Munich, Germany (collected 1994, 19 lines), Sengwa Wildlife Preserve, Zimbabwe (collected 1990, 12 lines, obtained from C. AQUADRO and C.-I WU), Nairobi, Kenya (collected 1988, 14 lines, obtained from Bowling Green Stock Center), Peking, China (collected 1993, 10 lines, obtained from J. DAVID), New Delhi, India (collected 1994, 18 lines, obtained from J. DAVID), St. Marteen, Antilles (collected 1994, 20 lines, obtained from J. DAVID), Gosier, Antilles (collected 1994, 15 lines, obtained from J. DAVID). Genomic DNA was prepared from single female flies by the high salt extraction method (MILLER *et al.* 1988).

Isolation, amplification and detection of microsatellites: The mapped P1 clones (SMOLLER *et al.* 1991) used in this study are listed in Table 1. DNA was extracted by the use of Qiagen (Hilden, Germany) columns according to the instructions of the supplier. Microsatellites were isolated from single P1 clones following essentially the protocol of RASSMANN *et al.* (1991) with slight modifications. Instead of a simultaneous digest with *Alu*I, *Hae*III, and *Rsa*I, different combinations of the enzymes were used and digested separately. After pooling the digests they were precipitated, cloned into M13 mtvh (SCHLÖTTERER and WOLFF 1996), screened with dinucleotide simple sequence probes and sequenced. Primers were derived from the flanking sequences of the microsatellite using the Oligo-4 (National Biosciences, Inc.) program. PCR amplification conditions were 30 cycles of 1 min at 94°, 2 min at 44–53° (depending on the primer combination, see Table 1) and 1 min at 72°. Following the protocol of RASSMANN *et al.* (1991), ^{32}P-end-labeled PCR products were separated on a 6% denaturing polyacrylamide gel. Allele sizes were determined by running a sequence ladder adjacent to the PCR products.

Data analysis: Since isofemale lines were used for typing that had been propagated in the laboratory at small population sizes, most individuals were homozygous. For heterozygous individuals one allele was randomly discarded. Gene diversities and their standard deviations were calculated according to NEI (1987, p. 177) and were corrected for small sample sizes (NEI and ROYCHOUDHURY 1974).

Genetic distances were calculated by using either the Microsat 1.4 (MINCH *et al.* 1995) or Phylip (FELSENSTEIN 1991) software package. UPGMA and neighbor joining dendrograms were obtained by the Phylip software package (FELSENSTEIN 1991). Statistical analysis was carried out with Statview software.

TABLE 1

Microsatellites in *D. melanogaster*

P1 clone	Chromosomal location	No. of alleles	Variance in repeat number	Size range	Repeat type	Coefficient of exchange	Annealing temperature	Primer sequences
DS01001	3A1-3A4	10	4.9	217–240 (2.5–14)	GT	0.0222	53°	ccaacgcaacgcaaccag ctcccacccaaaatggaaata
DS06335a	3C1-3C6	12	32.3	84–113 (7.5–22)	GT	0.1400	53°[c]	actgtaattgctgttctatgt cgcacactgggacacaaaa
DS06335b	3C1-3C5	6	2.5	134–147 (8–14.5)	GT	0.1400	50°	gcacaatcacatcgtattcact attgttgttgctgcgattt
DS06577	3C7-3C7	8	67.0	122–145 (4.5–16)	GT	0.1212	54°	tttgcggcttcgtttgttta gatccgcccacatacacact
DS06238	35B1-35B3	8	1.8	131–138 (18.5–22)[a]	GA	0.0647	52°	gggcattaatctgcattttcta ttcattagtaattgcaggcact
DS08513	37B8-37C2	5	4.2	192–205 (6.5–13)	GT	0.0184	54°[c]	ctacaccatgccctgaaaag gacggcgtgaccatgtctg
DS00062	54A1-54B2	12	20.6	158–174 (6.5–14.5)	GT	0.0435	53°[b]	acgggaacgccatctaac agaagagaccctgcaacaca
DS00361	54B1-54B2	10	72.6	132–159 (5–18.5)	GT	0.0435	52°	caaccacccacaagcacac cctctccggttgggctac
DS08687a	57C5-57D1	8	4.8	178–187 (8–12.5)	GT	0.0718	54°[c]	tgatttggactgagatcagg gcccaacgaatcatttcac
DS08687b	57C5-57D1	13	8.7	155–174 (8–14.5)	GT	0.0718	52°	tgtcgagagcagcagcagt ttgccttcccctgttacag

List of P1 clones and derived microsatellites analyzed in this study. All loci harbor a GT/AC dinucleotide repeat apart of DS06238 that has a GA/TC repeat. The number repeats is calculated on the basis of the cloned microsatellite. Coefficients of exchange are taken from AQUADRO and BEGUN (1993).

[a] Clones falling outside the range of observed allele sizes.

[b,c] Loci may be multiplexed in PCR reaction.

A C++ code was written to test whether the observed low gene diversities are due to a sampling effect. Alleles at a single locus were randomly drawn with replacement from the entire population or non-African populations only until the same number of chromosomes was sampled as in the original sample. The gene diversity in the population generated by resampling was calculated and the fraction of populations yielding the same or lower gene diversity were taken as an estimator of the probability that the same or lower level of heterozygosity may have occurred by chance in a panmictic population. The C++ code and executables are available on request.

We tested for deviations from mutation-drift equilibrium as follows. Assuming the stepwise mutation model and that each population is in mutation-drift equilibrium for each locus, independent from other populations and loci, then the variances in repeat number have an equilibrium distribution, whose mean and variance has been given by MORAN (1975) and ZHIVOTOVSKY and FELDMAN (1995), while the mean values do not arrive at an equilibrium distribution.

Let the logarithms of the variances in repeat number for each locus/population combination be given by v_{ij}. Let the log variances be ordered such that each row corresponds to one population and each column to one locus. Calculate the row means, $v_{i.}$, corresponding to different populations, the column means, $v_{.j}$, corresponding to different loci, and the grand mean, $v_{..}$. The difference between the row means and the grand mean then reflects the influences specific to a population, *e.g.*, different effective population sizes. The difference between the column means and the grand mean reflects influences specific to a locus, *e.g.*, differences in mutation rate, different effective population size due to different location on a sex-chromosome or autosome, or different levels of background selection.

Table 3 gives a comparison of the log-variances of the individual locus/population combinations, the corresponding linear combination of the row means, of the column means, and of the grand mean. The X statistics should reflect all variation not accounted for by differences in row means or column means, *i.e.*, deviations from mutation-drift equilibrium, such as hitchhiking to loci under locally different selection regimes. To test individual locus/population combinations for deviation from neutral expectation, approximate t values can be calculated. Details of the test statistics are provided in the APPENDIX A. If the assumption of independence of the populations is violated, *e.g.*, because of migration or recent historic relationships of the populations, the variance of the log variances between populations should be reduced resulting in a more conservative test. A serious problem, however, is posed by loci monomorphic in a certain population, because this results in an indefinite logarithm of the variance in repeat number. For our data, we countered this problem by adding 0.1 to the variances before taking the logarithm.

The algorithm for testing deviations from mutation-drift equilibrium (see APPENDIX A) was written in C++ code and an executable DOS program is available on request.

RESULTS

Characterization of the microsatellites: To isolate microsatellites from chromosomally defined regions, we have prepared microlibraries from 17 mapped P1 clones from the first, second and fourth chromosome. These were screened with the dinucleotide simple sequence probes GT/AC and GA/TC. None of the three

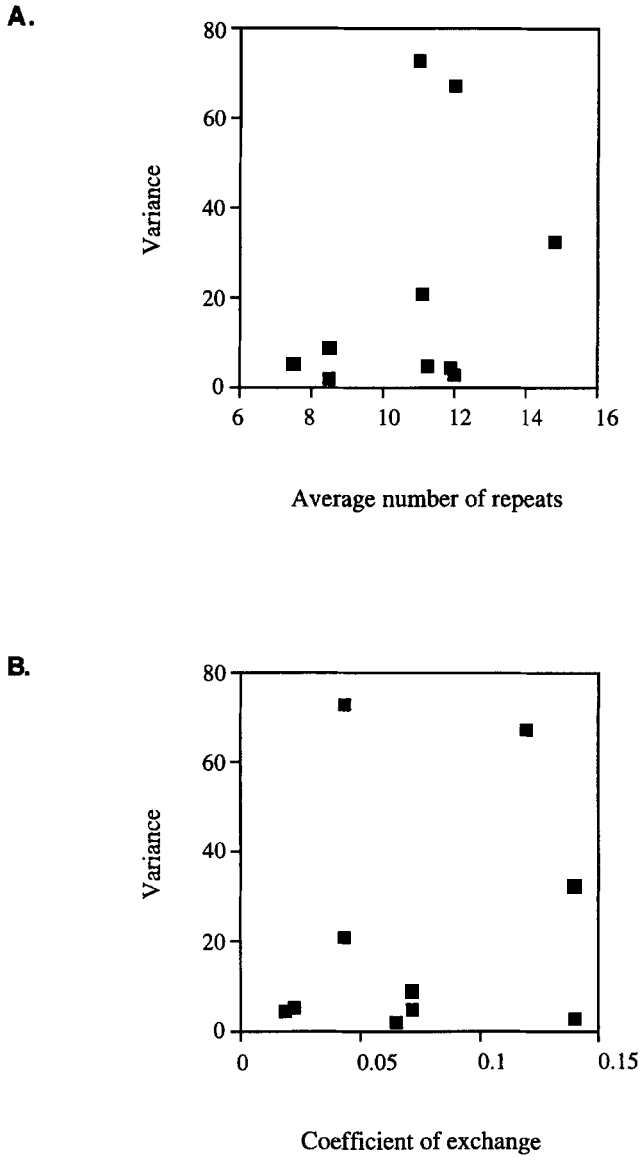

FIGURE 1.—(A) Variance plotted against average repeat number. No significant correlation was observed ($P = 0.31$). (B) Variance plotted against the coefficient of exchange. No significant correlation was observed ($P = 0.59$). Coefficients of exchange are taken from AQUADRO and BEGUN (1993).

clones from the fourth chromosome, nor of the four clones from the centromeric and telomeric regions yielded a positive signal. While this asymmetry of microsatellite distribution is still poorly understood (TAUTZ and SCHLÖTTERER 1994), it is in line with the *in situ* hybridization results on polytene chromosomes by PARDUE *et al.* (1987). The other P1 clones yielded up to five different dinucleotide repeat microsatellites.

Ten loci from the first and second chromosome were chosen for further analysis. All harbored eight to 12 uninterrupted repeat units and yielded between five and 13 alleles. The variance was not correlated with average length of the repeat ($r = 0.36$, d.f. = 8) (Figure 1A). We also did not find a correlation between the maximum number of repeats and variance ($r = 0.35$, d.f. = 8) (not shown). Using heterozygosity as an indicator of variation, we failed also to detect any correlation with the size of the repeat array (not shown). All but one locus showed alleles differing in length by single nucleotides only, suggesting that single nucleotide indels in the flanking regions contributed to the polymorphisms. Such indels are known to occur frequently in noncoding DNA of Drosophila. It should be noted that indels are affecting the calculated number of repeats, as this numbers are based on the sequence of the cloned microsatellites. Hence, an indel in the cloned sequence may introduce a strong bias in the calculated number of repeats.

Since the loci were isolated from genomic regions, which are exposed to different rates of recombination, we tested whether the variance in repeat number was correlated with the coefficient of exchange in the respective regions. While the nucleotide diversity of single copy genes in the Drosophila genome shows such a correlation (AQUADRO and BEGUN 1993), we found no correlation of microsatellite variance and the coefficient of exchange ($r = 0.19$, d.f. = 8) (Figure 1B). This result is in accordance with recently published trinucleotide polymorphisms in *D. melanogaster*, which did not find any correlation between any measure of variability and the recombination rate over a wider range than our study (MICHALAKIS and VEUILLE 1996).

Gene diversities in *D. melanogaster* populations: Between 10 and 20 isofemale lines from seven *D. melanogaster* populations were typed for 10 microsatellite loci. Most lines were homozygous as would be expected for lines kept in the laboratory at small population sizes. In cases where two alleles were found, one was chosen at random.

Gene diversities were calculated according to NEI (1987). The East African populations showed a higher average genetic diversity than the other populations (Table 2), as previously described for nucleotide variation (BEGUN and AQUADRO 1993; BEGUN and AQUADRO 1995; BÉNASSI and VEUILLE 1995). Locus DS06335b, with the lowest overall gene diversity, shows a higher variability in African populations, indicating that the reduced level of genetic diversity at this locus is not the consequence of mutational constraints on the locus itself.

For five loci we found a population specific pattern of reduction in variability. One population from the Antilles (St. Marteen) and one from Germany are monomorphic for locus DS06335b. Similarly, the Chinese population is monomorphic for locus DS06238. A strong reduction in variability was observed for locus DS01001 in the Indian population, for locus DS00361 in the German population and for locus DS08687 in the two populations from the Antilles. Bootstrap analysis based on the assumption of no population subdivision showed that the observed reduction of variability in these populations is not due to a sampling effect ($P < 0.05$, indicated in Table 2). If bootstrap resampling

TABLE 2

Gene diversities in *D. melanogaster*

	DS01001	DS06335a	DS06335b	DS06577	DS06238	DS08513	DS00062	DS00361	DS08687a	DS08687b	H
New Delhi, India (n = 18)	0.20** ± 0.082	0.65 ± 0.080	0.38 ± 0.093	0.44 ± 0.082	0.65 ± 0.068	0.56 ± 0.070	0.39 ± 0.097	0.56 ± 0.070	0.72 ± 0.051	0.63 ± 0.082	0.52
Beijing, China (n = 10)	0.55 ± 0.126	0.63 ± 0.074	0.37 ±0.112	0.50 ± 0.074	0.00**	0.63 ± 0.074	0.63 ± 0.074	0.52 ± 0.112	0.63 ± 0.074	0.52 ± 0.112	0.50
Munich, Germany (n = 19)	0.43 ± 0.087	0.71 ± 0.038	0.00*	ND	0.58 ± 0.057	0.51 0.060	0.83 ± 0.036	0.20* ± 0.084	0.68 ± 0.066	0.75 ± 0.029	0.52
Nairobi, Kenya (n = 14)	0.83 ± 0.046	0.80 ± 0.045	0.64 ± 0.048	ND	0.84 ± 0.033	0.37 ± 0.091	0.75 ± 0.056	0.72 ± 0.040	0.84 ± 0.033	0.83 ± 0.050	0.74
Sengwa Res., Zimbabwe (n = 12)	0.71 ± 0.053	0.90 ± 0.035	0.64 ± 0.097	0.78 ± 0.063	0.84 ± 0.034	0.42 ± 0.110	0.83 ± 0.050	0.81 ± 0.047	0.54 ± 0.107	0.87 ± 0.040	0.87
St. Marteen, Antilles (n = 20)	0.39 ± 0.097	0.81 ± 0.035	0.00*	0.70 ± 0.062	0.52 ± 0.072	0.43 ± 0.087	0.69 ± 0.069	0.56 ± 0.076	0.19** ± 0.079	0.80 ± 0.033	0.51
Gosier, Antilles (n = 15)	0.39 ± 0.107	0.74 ± 0.046	0.13 ± 0.079	ND	0.28 ± 0.109	0.65 ± 0.078	0.85 ± 0.031	0.73 ± 0.062	0.25* ± 0.100	0.65 ± 0.087	0.52
Average	0.50	0.75	0.31	0.61	0.53	0.51	0.71	0.59	0.55	0.72	0.58

Gene diversities of populations tested. Bootstrap replication probabilities were obtained for populations with a gene diversity of ≤0.25 for a given locus to test for deviation from random allele distributions. Probabilities for 1000 replications are as follows: $**P = <0.01$; $*P = <0.05$; ND, not determined. H indicates the gene diversity of a single population averaged over all loci.

was confined to non-African populations, probabilities were less significant, but still smaller than 0.05 for loci DS06238, DS00361 and DS08687.

These results cannot be ascribed to population bottlenecks, since the average gene diversity is not affected in these populations (Table 2). To obtain a rough estimate of the reduction of variability along the respective chromosome, we compared the behavior of loci from the same genomic region. Neighboring loci that were isolated from a P1 clone from the same genomic region showed average levels of gene diversities in the respective populations (Table 2). This indicates that only short chromosomal segments (not more than 80–100 kb) are affected by the loss of gene diversity.

Deviation from mutation-drift equilibrium: To test more specifically for deviation from mutation-drift equilibrium, we compared nine microsatellite loci in six populations (locus DS06238 was excluded because of missing data and the populations from the Antilles were pooled). The chi-square statistics and the approximate t values (in our study with 40 degrees of freedom) are given in Table 3. We find that the row means are not significantly different, *i.e.*, that the populations do not differ significantly. The column means, however, are significantly different, which is to be expected as no correction was made for different effective population sizes for autosomal *vs.* sex-chromosomal loci. Furthermore, mutation rates may vary between loci. The significant value of the X statistics indicates deviation from mutation-drift equilibrium. For the individual locus/population combinations, the locus DS01001 differed significantly from the predicted value in the Indian population, even if the conservative Bonferroni correction was applied. Two more loci were exactly on the border of significance for the one-sided test: DS00062 for the Zimbabwe population and DS06335 for the German population. The relatively high heterozyogosity of the former suggests that it might not be significantly genetically depauperate while homozygosity of the latter might lend more weight to the significance value. This reasoning receives more credibility as estimates of $4N_e\mu$ based on heterozygosity appear to have less variance than those based on the variance in repeat number (unpublished results from Monte-Carlo simulations).

Differentiation between *D. melanogaster* populations: Our data support the notion that *D. melanogaster* originated in Africa and spread from there over the rest of the world recently (DAVID and CAPY 1988; SINGH and LONG 1992). Figure 2 shows dendrograms based on Cavalli-Sforza's chord measure (CAVALLI-SFORZA and EDWARDS 1967). Despite slightly different branching patterns for two different tree building algorithms (UPGMA and NJ), the non-African populations are always grouped together. Other distance measures, like Reynolds distance and allele sharing (MINCH *et al.* 1995; REYNOLDS *et al.* 1983), yield identical tree topologies. NEI's genetic distance resulted in slightly different tree topologies for the non-African populations, however no different topology exceeded bootstrap values of >55%. Trees based on delta mu (GOLDSTEIN *et al.* 1995) were very different from the trees shown, but still the grouping of the African populations was consistent with all other trees. Population subdivision is further supported by bootstrap resampling under the assumption of a single, panmictic *D. melanogaster* population, which showed significant deviations for the populations-loci combina-

TABLE 3

Chi-square statistics and approximate *t* values

$v_{ij} = \ln[w_i N_j]$	DS01001	DS06335a	DS06335b	DS06238	DS08513	DS00062	DS00361	DS08687a	DS08687b	$v_{i.} = \ln[w_i N]$
A. Observed logarithms of the variances and means in repeat number (v_{ij}), and their row means ($v_{i.}$), column means ($v_{.i}$), and the grand mean ($v_{..}$)										
India	−1.585	3.435	1.916	−0.229	1.769	2.456	3.770	1.492	1.388	1.601
Zimbabwe	2.958	3.670	2.042	2.095	0.825	1.334	4.100	3.341	0.864	2.359
Kenya	3.404	4.119	0.790	1.454	0.803	1.981	3.347	2.198	2.368	2.274
China	1.470	2.661	−0.130	−2.303	1.394	3.211	1.713	1.131	0.839	1.110
Germany	0.172	2.386	−2.303	0.216	0.454	2.942	3.396	1.547	1.852	1.185
Antilles	−0.187	3.111	−1.523	−0.601	1.807	2.909	4.467	1.517	0.406	1.323
$v_{.j} = \ln[wN_j]$	1.039	3.230	0.132	0.105	1.175	2.472	3.465	1.871	1.286	$v = 1.642$
$d_{ij}*KL/((K-1)(L-1))$	DS01001	DS06335a	DS06335b	DS06238	DS08513	DS00062	DS00361	DS08687a	DS08687	
B. Normed deviations of particular population-locus combinations from their expectations										
India	−3.731*	0.358	2.581	−0.416	0.918	0.035	0.506	−0.493	0.207	
Zimbabwe	1.75	−0.406	1.719	1.833	−1.555	−2.716	−0.121	1.101	−1.661	
Kenya	2.522	0.376	0.037	1.031	−1.463	−1.644	−1.1	−0.445	0.655	
China	1.38	−0.054	0.374	−2.596	1.078	1.853	−1.786	−0.301	0.122	
Germany	−0.587	−0.657	−2.751	0.789	−0.38	1.352	0.567	0.193	1.474	
Antilles	−1.304	0.292	−1.871	−0.542	1.371	1.103	1.933	−0.051	−0.81	

The observed deviations of the row means ($\chi^2 = 14.021$) are not significantly different from those expected, while those for the column means ($\chi^2 = 69.328$) and for the test statistics χ are ($\chi^2 = 56.491$) (χ^2 (40;0.05) = 56). ($d_{ij}*KL/(K-1)(L-1)$)(t(40;0.05) = 3.5; Bonferroni-corrected, for two-sided test).

tions with reduced variation. If the resampling procedure was confined to non-African populations, three loci still showed significant reductions in variability (Table 2). Hence, corroborating population subdivision in *D. melanogaster.*

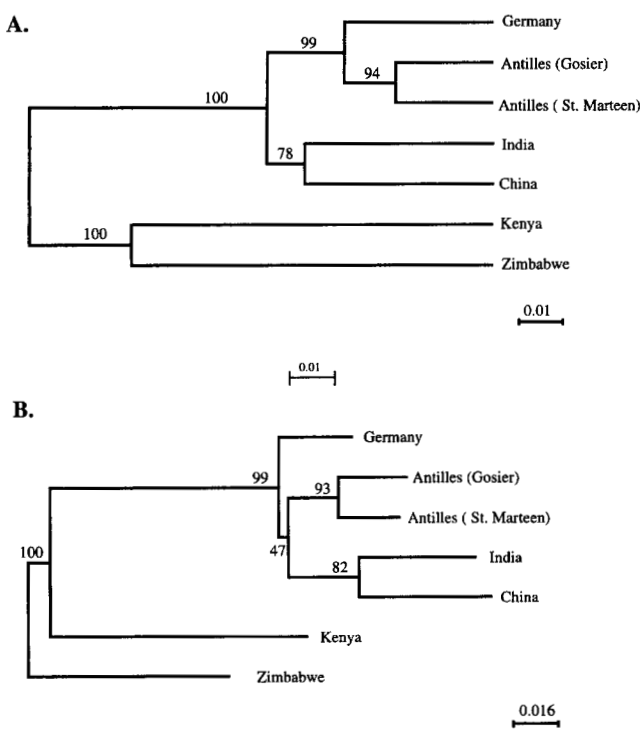

FIGURE 2.—Dendrograms based on Cavalli-Sforza's chord measure (CAVALLI-SFORZA and EDWARDS 1967). (A) UPGMA and (B) neighbor joining tree. Branch lengths are taken from the original data set and bootstrap values were determined separately. Bootstrap support from 100 replications are given for the respective branches. All trees shown are unrooted.

Interestingly, irrespective of the method used for determining genetic distances, the two East African populations are well separated from each other. This is in contrast to the results obtained for the *vermilion* data, where no genetic differentiation between the East African lines from Zimbabwe and Kenya was observed (BEGUN and AQUADRO 1995), hence microsatellites may thus be an even more sensitive indicator of population subdivision than nucleotide polymorphism.

Correlation of allele frequency distributions: Given the large effective population size of *D. melanogaster,* it can be assumed that allele frequencies between populations are correlated over many generations. In the non-African populations, which have been founded by a single or a few founder events (BÉNASSI and VEUILLE 1995), it has to be expected that allele frequency distributions outside Africa should be correlated (APPENDIX B). Figure 3 shows a graphic representation of the allele distributions in the four continents studied. While African populations have a wider allele spectrum, non-African populations often share a single predominant allele, suggesting a significant correlation between the allele frequencies in non-African populations. Given the observed population subdivision between *D. melanogaster* populations, standard tests for differences in allele distribution will detect differences between populations, despite the correlation of allele frequency distributions outside Africa. To quantify the correlation in allele frequency distribution between populations sharing a recent common ancestor, we are currently developing a new test statistic (C. VOGL, C. SCHLÖTTERER and R. BÜRGER, unpublished data).

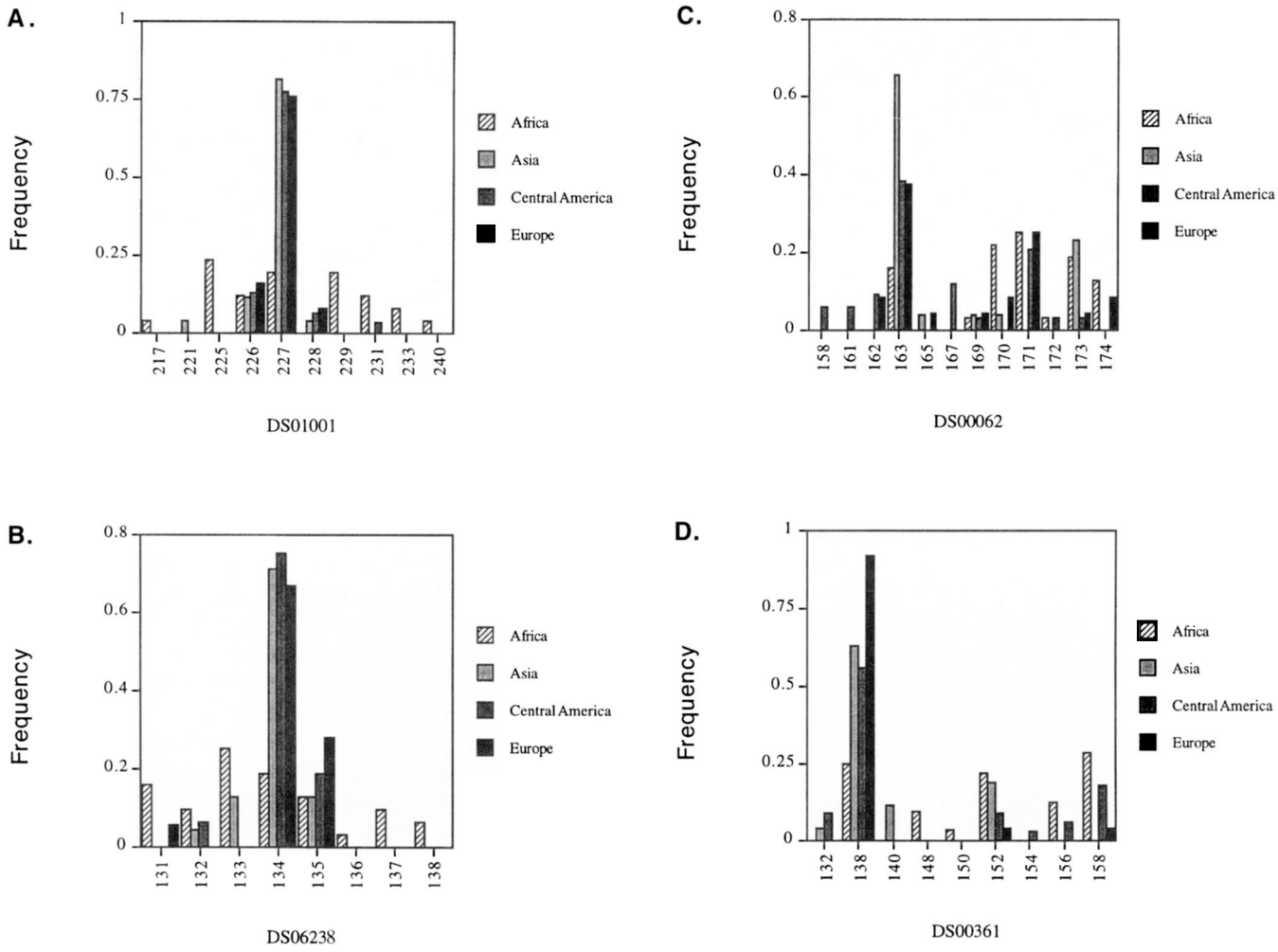

FIGURE 3.—Allele frequencies of four representative microsatellite loci (DS 00062, DS01001, DS 06238 and DS 00361) with the populations grouped into four different continents.

DISCUSSION

The aim of our study was to show the potential role of hitchhiking on the genetic diversity at microsatellite loci. We have indeed found evidence for such an effect. There is a population specific reduction in variability at some loci that would be difficult to explain by other means, like background selection.

D. melanogaster populations originated in Africa (LACHAISE *et al.* 1988) and colonized the Eurasian continent during the past 10,000–15,000 years (DAVID and CAPY 1988; SINGH and LONG 1992). The colonization of a new continent and of new habitats must have led to many local adaptations (DAVID and CAPY 1988). Such adaptations would have left their traces in the genome and it is therefore likely that the observed loss of gene diversity in short chromosomal segments is due to positive selection on loci that are linked to the microsatellite loci studied. Though it is difficult to assess how many genes would have been under positive selection for any given population, it does not seem unreasonable to assume that several hundred genes were involved. Under the stringent statistical test, only one locus in one population shows a highly significant deviation from random expectations in our data set. However, for five loci one may suspect that the detected loss of gene diversity may be due to a similar effect. The effect of a given locus under selection on the gene diversity of a linked microsatellite would of course depend on multiple factors, among them the degree of selection and the degree of linkage, which may differ between loci. One further important factor is the mutation rate of microsatellites. Estimated mutation rates of up to 10^{-2} have been published for mammals. If the mutation rates in *D. melanogaster* were this high, it would be extremely difficult to detect selective sweeps. After a selective sweep, which reduces the observed variance, new mutations would be accumulated very quickly and the observed variance would approach the equilibrium value in a very short time interval. Therefore, only a rather short timespan would be available to detect deviations from the equilibrium variance in natural populations. A recent study measuring mutation rates of *D. melanogaster* microsatellites estimated an average mutation rate of 6.3×10^{-6} (SCHUG *et al.* 1997), which is several orders of magni-

tudes lower than estimates for other organisms. Therefore, reduced variation in specific locus/population combinations is consistent with the low mutation rates of *D. melanogaster* microsatellites.

In our test statistic we did not consider gene flow between populations, as this would be difficult to handle theoretically. However, assuming constant and equal gene flow between all populations, our test is quite conservative. Gene flow would decrease the variance of the variance in repeat numbers between populations. Therefore, we would expect the differences between populations to be even more significant if we would consider gene flow between populations.

Instead of effects on linked genes, one could conceivably also invoke the possibility that the expansion of populations into new habitats would have been accompanied by some population bottlenecks, which might have led to a random loss of allelic diversity at some loci, but not at others. However, our data argue against this possibility. The fact that the non-African populations show a correlation in allele distribution suggests that they have maintained a rather high effective population size after emerging from Africa.

Our results on the population subdivision are somewhat unexpected. The question of a population substructure in *D. melanogaster* has been subject of a long standing debate. Isozyme studies in *D. melanogaster* indicated that high rates of gene flow are counteracting genetic differentiation of local *D. melanogaster* populations (SINGH and RHOMBERG 1987). However, recent studies demonstrated higher levels of genetic variability in East African populations compared with North American populations, implying there is an unappreciated degree of population structure in *D. melanogaster* (BEGUN and AQUADRO 1993; BEGUN and AQUADRO 1995). West African flies were found to harbor less polymorphism than East African ones (BÉNASSI *et al.* 1993). This population differentiation within Africa is supported by inversion polymorphisms between five populations from West Africa and five populations from East Africa (LEMEUNIER and AULARD 1992). Interestingly, our genetic distance data suggest a significant population subdivision between the Kenya and Zimbabwe population, which are separated by roughly 2000 km. The observed genetic differentiation of the two East African populations is reflected by assortative mating behavior of the Zimbabwe population (WU *et al.* 1995), as the Zimbabwe population is the only *D. melanogaster* population for which assortative mating has been described.

More controversy exists about the population structure of non-African populations. Several studies supported the idea of a relatively panmictic world-wide population (KREITMAN 1983; KREITMAN and AGUADÉ 1986; BERRY and KREITMAN 1992; SINGH and LONG 1992). Still, some experiments suggested population subdivision in non-African populations: the distribution of mitochondrial haplotypes is not even among *D. melanogaster* populations around the world (HALE and SINGH 1987). Data on the *vermilion* locus also show some indications of genetic differentiation between non-African populations (BEGUN and AQUADRO 1995). Similarly, studies of the molecular variation in the *y-ac-sc* region showed reduced genetic variation in some populations, while there is apparently no reduction in other populations (AGUADÉ *et al.* 1989; BEECH and LEIGH-BROWN 1989; EANES *et al.* 1989; BEGUN and AQUADRO 1991; MARTIN-CAMPOS *et al.* 1992). Finally, there is a well documented ecological differentiation among populations (DAVID *et al.* 1983). Hence, our results of a significant population subdivision in *D. melanogaster* is corroborated by several other studies.

An identification of hitchhiking effects is not unique to our data set. The microsatellite survey on *D. melanogaster* populations from North America has revealed one monomorphic locus among 18 loci (GOLDSTEIN and CLARK 1995). Similarly, a study of 10 microsatellite loci revealed two monomorphic microsatellites in an Australian population (ENGLAND *et al.* 1996). Though both studies did not show that the monomorphic loci are polymorphic in other populations, the size of the microsatellite motif of both loci was too long to be causative for the observed lack of polymorphism.

If hitchhiking of microsatellite loci indeed turns out to be more widespread, one could specifically screen for microsatellite loci with reduced variability in specific populations. These microsatellites could then serve as a marker to find linked genes that may have caused the effect. Thus, a systematic screen for genes involved in local adaptations could be devised.

We are grateful to C. AQUADRO, P. CAPY, J. DAVID, D. WOODRUFF and C.-I WU for providing *D. melanogaster* lines. P1 clones were provided by S. HERRMANN and M. ASHBURNER. We are grateful to R. BÜRGER for sharing unpublished results and many helpful comments on the manuscript. We would like to thank the members of the Department of Ecology and Evolution, University of Chicago for helpful discussions. J. PEMBERTON and B. CHARLESWORTH provided helpful comments on earlier versions of the manuscript. Special thanks to the reviewers and A. CLARK for their comments on the manuscript. This work was supported by the European Union program "Conservation of genetic resources III.2-Development of universal techniques for rapid screening of genetic variation"; Contract No. BIO2-CT92-0476, a DFG fellowship to C.S. and the Fonds zur Förderung der wissenschaftlichen Forschung (P11628-GEN).

LITERATURE CITED

AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989 Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122:** 607–615.

AQUADRO, C. F., and D. J. BEGUN, 1993 Evidence for and implications of genetic hitchhiking in the Drosophila genome, pp. 159–178 in *Mechanisms of Molecular Evolution*, edited by N. TAKAHATA and A. G. CLARK. Japan Scientific Societies Press, Tokyo.

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

BEECH, R. N., and A. J. LEIGH-BROWN, 1989 Insertion-deletion variation at the *yellow-achaete-scute* region in two natural populations of *Drosophila melanogaster*. Genet. Res. **53:** 7–15.

BEGUN, D. J., and C. F. AQUADRO, 1991 Molecular population genetics of the distal portion of the X chromosome in Drosophila:

evidence for genetic hitchhiking of the *yellow-achete* region. Genetics **129:** 1147–1158.

BEGUN, D., and C. F. AQUADRO, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365:** 548–550.

BEGUN, D. J., and C. F. AQUADRO, 1995 Molecular variation at the *vermilion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. Genetics **140:** 1019–1032.

BÉNASSI, V., and M. VEUILLE, 1995 Comperative population structuring of molecular and allozyme variation of *Drosophila melanogaster Adh* between Europe, West Africa and East Africa. Genet. Res. **65:** 95–103.

BÉNASSI, V., S. AULARD, S. MAZEAU and M. VEUILLE, 1993 Molecular variation of *Adh* and *P6* genes in African populations of Drosophila melanogaster and its relation to chromosomal inversions. Genetics **134:** 789–799.

BERRY, A., and M. KREITMAN, 1992 Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogster* an the east coast of North America. Genetics **134:** 869–893.

BOWCOCK, A. M., A. RUIZ-LINEARES, J. TONFOHRDE, E. MINCH, J. R. KIDD *et al.*, 1994 High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368:** 455–457.

CAVALLI-SFORZA, L. L., and A. W. F. EDWARDS, 1967 Phylogenetic analysis: models and estimation procedures. Evolution **32:** 550–570.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

DAVID, J. R., and P. CAPY, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4:** 106–111.

DAVID, J. R., R. ALLEMAND, J. VAN HERREWEGE and Y. COHET, 1983 Ecophysiology: abiotic factors pp. 106–170 in *The Genetics and Biology of Drosophila*, edited by M. ASHBURNER, H. L. CARSON and J. N. THOMPSON. Academic Press Inc., London.

DIB, C., S. FAURE, C. FIZAMES, D. SAMSON, N. DROUOT *et al.*, 1996 A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature **380:** 152–154.

DIETRICH, W. F., J. MILLER, R. STEEN, M. A. MERCHANT, D. DAMRON-BOLES *et al.*, 1996 A comprehensive genetic map of the mouse genome. Nature **380:** 149–152.

EANES, W. F., J. LABATE and J. W. AJIOKA, 1989 Restriction map variation associated with the *yellow-achaete-scute* region in four populations of *Drosophila melanogaster*. Mol. Biol. Evol. **6:** 492–502.

ENGLAND, P. R., D. A. BRISCOE and R. FRANKHAM, 1996 Microsatellite polymorphisms in a wild population of *Drosophila melanogaster*. Genet. Res. **67:** 285–290.

ESTOUP, A., L. GARNERY, M. SOLIGNAC and J.-M. CORNUET, 1995 Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics **140:** 679–695.

FELSENSTEIN, J., 1991 PHYLIP, Version 3.57c, University of Washington, Seattle.

GOLDSTEIN, D. B., and A. G. CLARK, 1995 Microsatellite variation in North American populations of *Drosophila melanogaster*. Nucleic Acids Res. **23:** 3882–3886.

GOLDSTEIN, D. B., A. RUIZ LINEARES, L. L. CAVALLI-SFORZA and M. W. FELDMAN, 1995 Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Natl. Acad. Sci. USA **92:** 6723–6727.

HALE, L. R., and R. S. SINGH, 1987 Mitochondrial DNA variation and genetic structure in populations of *Drosophila melanogaster*. Mol. Biol. Evol. **4:** 622–637.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304:** 412–417.

KREITMAN, M., and M. AGUADÉ, 1986 Genetic uniformity in two populations of *Drosophila melanogaster* as revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. Proc. Natl. Acad. Sci. USA **83:** 3562–3566.

LACHAISE, D., M.-L. CARIOU, J. R. DAVID, F. LEMEUNIER, L. TSACAS *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22:** 159–225.

LEMEUNIER, D., and S. AULARD, 1992 Inversion polymorphism in *Drosophila melanogaster*, pp. 339–405 in *Drosophila Inversion Polymorphism*, edited by C. B. KRIMBAS and J. R. POWELL. CRC Press, Cleveland, OH.

MARTIN-CAMPOS, J. M., J. M. COMERON, N. MIYASHITA and M. AGUADÉ, 1992 Intraspecific and interspecific variation at the *y-ac-sc* region of *Drosophila simulans* and *Drosophila melanogaster*. Genetics **130:** 805–816.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favorable gene. Genet. Res. **23:** 23–35.

MICHALAKIS, Y., and M. VEUILLE, 1996 Length variation of CAG/CAA trinucleotide repeats in natural populations of *Drosophila melanogaster* and its relation to the recombination rate. Genetics **143:** 1713–1725.

MILLER, S. A., D. D. DYKES and H. F. POLESKY, 1988 A simple salting out procedure from human nucleated cells. Nucleic Acids Res. **16:** 1215.

MINCH, E., A. RUIZ-LINARES, D. GOLDSTEIN, M. FELDMAN and L. L. CAVALLI-SFORZA, 1995 Microsat (version 1.4d): a computer program for calculating various statistics on microsatellite allele data.

NEI, M., 1987 *Molecular Evolutionary genetics*. Columbia University Press, New York.

NEI, M., and A. K. ROYCHOUDHURY, 1974 Sampling variances of heterozygosity and genetic distance. Genetics **76:** 379–390.

OHTA, T., and M. KIMURA, 1973 A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. **22:** 201–204.

PARDUE, M. L., K. LOWENHAUPT, A. RICH and A. NORDHEIM, 1987 $(dC\text{-}dA)_n\cdot(dG\text{-}dT)_n$ sequences have evolutionarily conserved chromosomal locations in Drosophila with implications for roles in chromosome structure and function. EMBO J. **6:** 1781–1789.

QUELLER, D. C., J. E. STRASSMANN and C. R. HUGHES, 1993 Microsatellites and kinship. Trends Ecol. Evol. **8:** 285–288.

RASSMANN, K., C. SCHLÖTTERER and D. TAUTZ, 1991 Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting. Electrophoresis **12:** 113–118.

REYNOLDS, J. B., B. S. WEIR and C. C. COCKERHAM, 1983 Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics **105:** 767–779.

SCHLÖTTERER, C., and J. PEMBERTON, 1994 The use of microsatellites for genetic analysis of natural populations pp. 203–214 in *Molecular Ecology and Evolution: Approaches and Applications*, edited by B. SCHIERWATER, B. STREIT, G. P. WAGNER and R. DESALLE. Birkhäuser Verlag, Basel.

SCHLÖTTERER, C., and D. TAUTZ, 1992 Slippage synthesis of simple sequence DNA. Nucleic Acids Res. **20:** 211–215.

SCHLÖTTERER, C., and C. WOLFF, 1996 M13mtvh-an improved M13 vector for rapid and simple cloning of PCR products. Trends Genet. **12:** 286–287.

SCHUG, M. D., T. F. C. MACKAY and C. F. AQUADRO, 1997 Low mutation rates of microsatellite loci in *Drosophila melanogaster*. Nature Genet. **15:** 99–102.

SINGH, R. S., and A. LONG, 1992 Geographic variation in Drosopila: from molecules to morphology and back. Trends Ecol. Evol. **7:** 340–345.

SINGH, R. S., and L. R. RHOMBERG, 1987 A comprehensive study of genic variation in natural popuations of Drosophila melanogaster. I. Estimates of gene flow from rare alleles. Genetics **115:** 313–322.

SMOLLER, D. A., D. PEROV and D. L. HARTL, 1991 Characterization of bacteriophage P1 libary containing inserts of *Drosophila* DNA of 75–100 kilobase pairs. Chromosoma **100:** 487–494.

STRAND, M., T. A. PROLLA, R. M. LISKAY and T. D. PETES, 1993 Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature **365:** 274–276.

TAUTZ, D., and C. SCHLÖTTERER, 1994 Simple sequences. Curr. Op. Gen. Dev. **4:** 832–837.

TAYLOR, M. F. J., Y. SHEN and M. E. KREITMAN, 1995 A population genetic test of selection at the molecular level. Science **270:** 1497–1499.

VALDES, A. M., M. SLATKIN and N. B. FREIMER, 1993 Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics **133:** 737–749.

VIARD, F., P. BREMOND, R. LABBO, F. JUSTY, B. DELAY *et al.*, 1996 Microsatellites and the genetics of highly selfing populations in the freshwater snail *Bulinus truncatus*. Genetics **142:** 1237–1247.

WEBER, J. L., 1990 Informativeness of human $(dC\text{-}dA)_n\cdot(dG\text{-}dT)_n$ polymorphisms. Genomics **7:** 524–530.

WU, C-I., H. HOLLOCHER, D. J. BEGUN, C. F. AQUADRO, Y. XU *et al.*, 1995 Sexual isolation in *Drosophila melanogaster*. A possible case

of incipient speciation. Proc. Natl. Acad. Sci. USA **92:** 2519–2523.

ZHIVOTOVSKY, L. A., and M. W. FELDMAN, 1995 Microsatellite variability and genetic distances. Proc. Natl. Acad. Sci. USA **92:** 11549–11552.

Communicating editor: A. G. CLARK

APPENDIX A

A test for deviations from mutation-drift equilibrium: Data on variances in repeat number must be available for L microsatellite loci from K populations from large samples. Throughout, the single-step mutation model with mutation rate w_l at the lth locus and population size N_k of the kth population is assumed. Thus, all even moments of the mutant distribution are equal to w_l and all odd moments are equal to 0. Furthermore, it is assumed that all locus/population combinations are in mutation-drift equilibrium independently from each other. Let $W_{kl} = N_k w_l$. Define V_{kl} as the observed variance of repeat number in the kth population at the lth locus and v_{kl} as the natural logarithm of V_{kl}, *i.e.*, $v_{kl} = \log V_{kl}$. Further, let $v_{k.}$ denote the mean of the log variances over L loci for population k, $v_{.l}$ as the mean of the log variances over K populations for locus l, and $v_{..}$ as the grand mean. Under assumption of mutation-drift equilibrium the expectation of the variance of repeat number in the kth population at the lth locus is $\mathcal{E}\{V_{kl}\} = (N_k - 1)\, w_l \approx W_{kl}$. The equilibrium variance of the variance in repeat number is approximately $\widehat{\mathrm{Var}}\{V_{kl}\} \approx (8W_{kl}^2 + W_{kl})/6$ (ZHIVOTOVSKY and FELDMAN, 1995).

Monte-Carlo simulations employing a coalescence model showed that the log variances are approximately normally distributed, but that first-order Taylor series approximations to the means and variances of the log variances are considerably biased. A better approximation to equilibrium mean and variance of the log variances is

$$\hat{v}_{kl} \approx \log W_{kl} - {}^1\!/_2 \log {}^7\!/_3 \approx \log [0.65\ W_{kl}]$$

and

$$\widehat{\mathrm{Var}}\,[v_{kl}] \approx 0.85$$

(R. BÜRGER personal communication). This approximation was checked by numerical simulations and is used throughout this manuscript (details on the simulation results and the derivation of the approximation are given in C. VOGL, C. SCHLÖTTERER, R. BÜRGER, unpublished data).

Under the assumption of constant mutation rates per locus, the quantity $u_{ij} = v_{i.} + v_{.j} - v_{..}$ has the same expectation as v_{ij}. The variance of u_{ij} can be calculated as follows: note that

$$\begin{aligned} u_{ij} &= v_{i.} + v_{.j} - v_{..} \\ &= \frac{1}{K}\sum_{k=1}^{K} v_{kj} + \frac{1}{L}\sum_{l=1}^{L} v_{il} - \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L} v_{kl} \\ &= \frac{K+L-1}{KL}\, v_{ij} + \frac{L-1}{KL}\sum_{k\neq i}^{K} v_{kj} \\ &\qquad + \frac{K-1}{KL}\sum_{l\neq j}^{L} v_{il} - \frac{1}{KL}\sum_{k\neq i}^{K}\sum_{l\neq j}^{L} v_{kl}. \end{aligned}$$

Now use the independence of the different v_{ij} and the approximate equality of variances of the different v_{ij} for realistic parameter values. It is easy to show that the variance of u_{ij} is equal to $(1/L + 1/K + 1/LK)$ times the variance of v_{ij}, *i.e.*, converges to zero as the number of populations and loci increases.

Define $d_{kl} = (v_{kl} - u_{kl})^2 / \mathrm{Var}\{u_{kl}\} \approx 0.85\,(v_{kl} - u_{kl})^2$, $d_{k.}$ as the means of the d_{kl} over loci, and $d_{.j}$ as the means over populations. Under the assumptions of independent mutation-drift equilibrium among populations and equal mutation rates at each across populations, the following three quantities have a limiting chi-square distribution with $(N-1)(K-1)$ degrees of freedom: the weighted differences of the row-means from the grand mean, under the assumption of equal population sizes of all populations,

$$(K-1)\sum_{l=1}^{L} d_{.l},$$

the weighted differences of the column-means from the grand mean, under the assumption of equal mutation rates at all loci,

$$(L-1)\sum_{k=1}^{K} d_{k.},$$

and

$$X^2 = \sum_{k=1}^{K}\sum_{l=1}^{L} d_{kl}.$$

Significant deviations from predicted values must be attributed to violations of assumptions, in particular those due to selection. For large K and L, deviation might be attributed to a particular locus-population combination by multiplying $v_{kl} - u_{kl}$ with

$$KL/[(K-1)(L-1)]$$

and comparing this approximate t with a table. For family error rates, the Bonferroni correction should be employed with $\alpha_B = 1 - (1-\alpha)^{(K-1)(L-1)}$.

As discussed in the text, evolutionary forces different from mutation or drift might influence row- and column-means, such that deviation from homogeneity of row- or column-means might occur even if mutation and drift are homogenous. These forces, however, should not influence the X^2 statistics.

APPENDIX B

Allele frequencies

Size	Total	India	Zimbabwe	Kenya	China	Germany	Antilles, St. M	Antilles, Gosier
				DS01001				
217	0.01	0	0	0.125	0	0	0	0
221	0.01	0	0	0	0.111	0	0	0
225	0.062	0	0.333	0.25	0	0	0	0
226	0.113	0.111	0	0.125	0.111	0.158	0.111	0.154
227	0.639	0.889	0.167	0	0.667	0.737	0.777	0.769
228	0.052	0	0	0	0.111	0.105	0.056	0.077
229	0.052	0	0.417	0	0	0	0	0
231	0.031	0	0	0.25	0	0	0.056	0
233	0.021	0	0	0.25	0	0	0	0
240	0.01	0	0.083	0	0	0	0	0
				DS06335A				
84	0.02	0	0.083	0	0	0	0.062	0
86	0.051	0.125	0	0.214	0	0	0	0
92	0.031	0	0.083	0.143	0	0	0	0
94	0.071	0	0.083	0	0	0	0.188	0.2
96	0.163	0.125	0.083	0	0.375	0.294	0.25	0.067
98	0.031	0	0.083	0.071	0	0.059	0	0
100	0.122	0.125	0.25	0.358	0	0	0.062	0.067
102	0.357	0.563	0.169	0.143	0.5	0.412	0.313	0.4
104	0.122	0.062	0.083	0	0	0.235	0.125	0.266
106	0.01	0	0	0	0.125	0	0	0
108	0.01	0	0.083	0	0	0	0	0
113	0.01	0	0	0.071	0	0	0	0
				DS06335B				
134	0.019	0.111	0	0	0	0	0	0
136	0.009524	0	0.083	0	0	0	0	0
140	0.095	0.111	0.083	0.385	0.222	0	0	0
142	0.819	0.778	0.584	0.461	0.778	1	1	0.933
144	0.048	0	0.167	0.154	0	0	0	0.067
147	0.009524	0	0.083	0	0	0	0	0
				DS06577				
122	0.047	0	0.038	0	0	0.1		
128	0.297	0.722	0	0.625	0	0.05		
132	0.016	0	0.083	0	0	0		
138	0.031	0	0.167	0	0	0		
140	0.141	0	0.417	0	0	0.2		
142	0.047	0	0.167	0	0	0.05		
144	0.281	0.056	0.083	0.375	0.5	0.5		
145	0.141	0.222	0	0	0.5	0.1		
				DS06238				
131	0.063	0	0.25	0.143	0	0.077	0	0
132	0.053	0.067	0	0.143	0	0	0.053	0.077
133	0.095	0.2	0.167	0.286	0	0	0	0
134	0.537	0.533	0.25	0	1	0.538	0.684	0.846
135	0.189	0.2	0.083	0.214	0	0.385	0.263	0.077
136	0.011	0	0	0.071	0	0	0	0
137	0.032	0	0.083	0.143	0	0	0	0
138	0.021	0	0.167	0	0	0	0	0
				DS08513				
192	0.009524	0	0	0	0	0	0	0.067
198	0.029	0.056	0.083	0.071	0	0	0	0
202	0.6	0.278	0.75	0.715	0.5	0.631	0.737	0.6

APPENDIX B

Continued

Size	Total	India	Zimbabwe	Kenya	China	Germany	Antilles, St. M	Antilles, Gosier
				DS06238				
204	0.162	0.056	0.167	0.214	0.125	0.316	0.105	0.133
205	0.2	0.610	0	0	0.375	0.053	0.158	0.2
				DS00062				
158	0.019	0	0	0	0	0	0.105	0
161	0.019	0	0	0	0	0	0	0.133
162	0.038	0	0	0	0	0.056	0.105	0.067
163	0.346	0.777	0	0.071	0.375	0.277	0.526	0.2
165	0.019	0.056	0	0	0	0.056	0	0
167	0.038	0	0	0	0	0	0.053	0.2
169	0.038	0.056	0.083	0	0	0.056	0	0.067
170	0.096	0	0.333	0.214	0.125	0.111	0	0
171	0.183	0	0.083	0.429	0	0.277	0.158	0.266
172	0.019	0	0.083	0	0	0	0.053	0
173	0.125	0.111	0.25	0.143	0.5	0.056	0	0.067
174	0.058	0	0.168	0.143	0	0.111	0	0
				DS00361				
132	0.038	0	0	0	0.111	0	0	0.2
138	0.528	0.610	0.167	0.071	0.667	0.894	0.631	0.467
140	0.028	0.056	0	0	0.222	0	0	0
148	0.028	0	0.25	0	0	0	0	0
150	0.009434	0.056	0	0	0	0	0	0
152	0.142	0.278	0.083	0.357	0	0.053	0.105	0.067
154	0.009434	0	0	0	0	0	0.053	0
156	0.057	0	0.083	0.215	0	0	0	0.133
158	0.151	0	0.334	0.357	0	0.053	0.211	0.133
159	0.009434	0	0.083	0	0	0	0	0
				DS08687A				
178	0.038	0	0	0.214	0	0.053	0	0
180	0.038	0.056	0	0.143	0	0.053	0	0
181	0.009524	0	0	0	0	0.053	0	0
182	0.171	0.444	0.083	0.143	0.125	0.158	0.105	0.067
184	0.114	0.111	0.167	0.071	0.375	0.158	0	0.067
185	0.038	0.111	0	143	0	0	0	0
186	0.581	0.278	0.667	0.286	0.5	0.525	0.895	0.866
187	0.009524	0	0.083	0	0	0	0	0
				DS08687B				
155	0.01	0	0.083	0	0	0	0	0
163	0.01	0	0.083	0	0	0	0	0
164	0.01	0	0	0.083	0	0	0	0
165	0.051	0.118	0.251	0	0	0	0	0
166	0.121	0	0.083	0.25	0	0.312	0.105	0.071
167	0.141	0.059	0	0.083	0.222	0.312	0.158	0.143
168	0.02	0	0	0	0	0	0.053	0.071
169	0.051	0.118	0.251	0	0	0	0	0
170	0.081	0.118	0	0.083	0.111	0.126	0.105	0
171	0.384	0.587	0	0.335	0.667	0.25	0.316	0.572
172	0.091	0	0.083	0.083	0	0	0.263	0.143
173	0.02	0	0.083	0.083	0	0	0	0
174	0.01	0	0.083	0	0	0	0	0