Genetic Differentiation Between American and European *Drosophila melanogaster* Populations Could Be Attributed to Admixture of African Alleles

G. Caracristi and C. Schlötterer

Institut für Tierzucht und Genetik, Wien, Austria

A total of 48 polymorphic microsatellite loci were characterized in 13 *Drosophila melanogaster* populations originating from Europe, America, and Africa. Consistent with previous results, the African *D. melanogaster* populations were the most differentiated populations and harbored most variation. Despite an overall similarity, American and European populations were significantly differentiated. Interestingly, genetic distances based on the proportion of shared alleles as well as F_{ST} values suggested that the American *D. melanogaster* populations are more closely related to the African populations than European ones are. We also detected a higher proportion of putative African alleles in the American populations, indicating recent admixture of African alleles on the American continent.

Introduction

Population genetics has a long-standing history of research aiming to understand the evolutionary forces shaping natural variation. With the availability of completed genomic sequences, population genetics is about to shift its emphasis from single locus studies toward a genomic approach eventually focusing on entire genomes. Through an assessment of the functional importance of natural polymorphisms, population genetics is becoming an important research discipline contributing to a functional annotation of sequenced genomes. Nevertheless, for a full exploitation of the potential of population genetics, a good understanding of the demographic history is required.

Originally, Drosophila species other than *D. melanogaster* were the focus of population genetic studies (Lewontin and Hubby 1966; Jones et al. 1981). The progress of molecular biology using *D. melanogaster* as a genetic model has shifted the research emphasis, even though other species would have been better suited because their ecology and demography were better characterized. Initial studies using phenotypic variation (Capy et al. 1986), chromosomal inversion polymorphisms (Voelker, and Mukai 1977; Mettler, Singh 1989; Lemeunier and Aulard 1992), restriction fragment length polymorphisms (Hale and Singh 1991; Begun and Aquadro 1993), and allozyme polymorphisms (David 1982; Oakeshott et al. 1982; Singh, Hickey, and David 1982; Singh and Rhomberg 1987) indicated that *D. melanogaster* populations are differentiated. Later surveys using DNA sequencing to study sequence variation focused either on worldwide samples of single individuals collected at different localities or on a small number of local populations (e.g., Kreitman 1983). The rationale for this experimental design was the high cost of DNA sequencing and the low levels of divergence among non-African populations. Although these sequence polymorphism analyses provided enormous benefit to the entire field of population genetics, the drawback of this experimental design is that the current knowledge about the demography

of *D. melanogaster* still rests on a number of studies using markers which are suspected to be influenced by natural selection: allozymes, mtDNA, and chromosomal inversions (Mettler, Voelker, and Mukai 1977; Voelker et al. 1978; Hickey 1979; Hale and Singh 1987; Singh and Rhomberg 1987; Singh, Hickey, and David 1982). Given the importance of *D. melanogaster* in population genomics and the anticipated investments in a molecular characterization of population polymorphisms, a characterization of the demographic past of *D. melanogaster* using neutral markers is needed.

Our current understanding is that *D. melanogaster* originated in Sub-Saharan Africa (Lachaise et al. 1988). The first out-of-Africa habitat expansion of *D. melanogaster* occurred between 10,000 and 15,000 years ago and involved the Eurasian continent (David and Capy 1988). In the more recent past North America and Australia were colonized (David and Capy 1988). Based on the similarities between Caribbean and African flies, an additional colonization event from Africa to Central/South America has been assumed (David and Capy 1988).

In this study, we used 48 polymorphic microsatellite loci to investigate the genetic differentiation between American and European *D. melanogaster* populations. Our results indicate that, despite an overall similarity, the *D. melanogaster* populations are well separated between the two continents. Interestingly, much of the genetic differentiation between the two continents could be attributed to admixture of African alleles to the American *D. melanogaster* populations.

Materials and Methods

Population Samples

We used the following *D. melanogaster* population samples: 19 lines from Harjavalta (Finland, 1996), 30 lines from Katowice (Poland, 2000), 30 lines from Weil am Rhein (Germany, 2000), 30 lines from Copenhagen (Denmark, 1998), 30 lines from Texel (Netherlands, 1997), 32 lines from Naples (Italy, 2000), 31 lines from Rockaway (New Jersey, 1999), 30 lines from Pennsylvania (1998), 30 lines from West End (North Carolina, 2000), 30 lines from Groth Winery (Napa Valley, California, 1996), 19 lines from ''La Milpa'' archeological site (north-western Belize, Central America, 1999), 15

Table 1
List of the 48 Microsatellite Loci Used in This Study

Locus	Chromosome	Cytological Position	Number of Alleles[a]	Heterozygosity[a]				Variance in Repeat Number[a]			
				Total	Europe	America	Africa	Total	Europe	America	Africa
AE002566_gt	X	3A	8.0	0.37	0.16	0.43	0.82	2.08	1.35	1.53	5.64
X3439769	X	3E	12.7	0.63	0.53	0.67	0.80	20.23	19.77	23.03	14.59
X3306698	X	3E	17.9	0.64	0.50	0.71	0.91	26.99	16.76	31.96	45.27
X3343263	X	3E	10.1	0.61	0.47	0.68	0.83	11.36	7.84	13.16	17.40
X3516772	X	3F	12.3	0.63	0.57	0.60	0.86	9.14	1.45	15.15	17.19
X3655941	X	3F	11.8	0.69	0.63	0.72	0.81	4.63	3.43	4.57	8.38
X3829513	X	4A	13.4	0.71	0.66	0.73	0.83	34.17	28.72	42.22	30.41
X4944599	X	4E	12.3	0.58	0.53	0.51	0.88	4.89	3.22	1.57	18.17
X5179712	X	4F	16.4	0.58	0.59	0.44	0.87	11.08	11.77	9.69	12.44
X5326452	X	5A	15.5	0.83	0.80	0.86	0.84	24.65	26.80	26.98	12.37
X5973753	X	5D	19.5	0.52	0.45	0.45	0.94	17.44	3.18	7.46	85.15
X7028104	X	7A	10.1	0.72	0.66	0.76	0.78	4.68	3.40	5.75	5.82
X8022709	X	7D	9.0	0.43	0.44	0.33	0.68	2.41	1.53	2.19	5.57
X13039889	X	11E	23.6	0.79	0.74	0.81	0.91	21.60	11.21	23.95	46.87
X13203739	X	11F	16.4	0.70	0.68	0.66	0.89	42.75	43.80	47.84	26.88
X14425888	X	12F	12.0	0.57	0.56	0.46	0.86	9.81	8.10	6.31	23.73
X15146508	X	13C	12.5	0.38	0.25	0.35	0.86	1.79	0.17	0.37	10.16
X15149564	X	13C	6.9	0.52	0.46	0.52	0.72	3.04	1.81	4.13	4.00
X15279912	X	13E	11.9	0.69	0.64	0.68	0.86	49.08	49.78	60.29	18.95
X15854539	X	14A	9.7	0.52	0.45	0.47	0.81	0.87	0.25	0.55	3.51
DS09020	X	15A	16.0	0.46	0.41	0.33	0.92	3.32	1.29	1.02	15.14
X17869774	X	17A	16.8	0.61	0.50	0.62	0.87	17.87	18.07	18.06	16.82
X19942741	X	19C	16.2	0.71	0.68	0.68	0.90	32.14	31.59	34.18	28.71
Pkg-TC	2	23A	9.0	0.73	0.68	0.76	0.79	4.52	3.59	4.68	6.88
Pkg-GT	2	23A	12.0	0.65	0.59	0.68	0.77	18.69	21.12	15.26	20.01
Dm0600-TC	2	24C3-D1	10.9	0.53	0.50	0.43	0.87	7.00	5.29	6.18	14.18
ft-CA	2	24E	5.0	0.48	0.42	0.53	0.56	10.15	7.38	12.21	13.31
AC005270	2	24E1-F1	16.4	0.80	0.81	0.77	0.81	26.75	28.65	28.72	16.13
Dm0332-TC	2	29F	4.4	0.48	0.33	0.65	0.49	1.47	0.39	2.78	1.41
2L/10056972	2	31A	15.3	0.73	0.69	0.71	0.89	6.97	5.08	5.93	15.23
Adh-TC	2	35B	8.8	0.64	0.65	0.69	0.50	11.23	12.32	12.05	5.88
cact-TC	2	35F	4.5	0.17	0.06	0.22	0.42	0.29	0.06	0.26	1.08
cact-TG	2	35F	4.5	0.49	0.52	0.53	0.32	0.66	0.63	0.84	0.32
Cad-GA	2	38D4-E1	8.9	0.47	0.39	0.46	0.70	2.31	2.29	1.99	3.14
tor-TA	2	43B3-C5	4.5	0.31	0.22	0.31	0.60	0.97	0.36	0.74	3.40
2R/5196790	2	46F	10.0	0.65	0.57	0.67	0.85	8.16	8.71	7.42	8.34
Drogpad	2	47A	12.5	0.52	0.39	0.55	0.81	10.63	2.62	16.96	18.84
2R5377736	2	47A	9.4	0.72	0.70	0.71	0.80	4.37	4.61	3.79	5.09
2R 5394848	2	47A	7.6	0.63	0.61	0.63	0.65	8.17	9.20	7.20	7.48
2R 5442209	2	47A	7.0	0.56	0.59	0.53	0.53	1.76	2.39	1.22	1.24
2R 5481770	2	47A	12.6	0.62	0.54	0.65	0.78	23.70	22.71	20.53	34.60
2R 5491247	2	47A	6.0	0.57	0.51	0.62	0.62	1.18	0.60	1.96	0.97
2R 5510443	2	47A	9.0	0.44	0.39	0.38	0.76	2.53	1.88	1.54	6.94
2R 5514150	2	47A	5.0	0.48	0.46	0.43	0.70	0.66	0.50	0.49	1.56
Dm0620	2	51E	6.0	0.54	0.51	0.56	0.62	1.22	1.19	1.09	1.64
Pkc53E-GA	2	53D	11.9	0.51	0.41	0.55	0.74	2.13	1.30	1.91	5.14
Ote-GA	2	55A2-B1	7.5	0.43	0.39	0.40	0.65	2.52	1.06	1.07	10.55
Dm0600-TA	2	55F	10.1	0.50	0.46	0.52	0.55	1.02	0.67	0.51	3.33

[a] Estimators of variability were calculated for each population separately and subsequently averaged.

lines from Harare (Zimbabwe, Africa), and 24 lines from Kenya (Africa), caught at various locations.

A single female fly was randomly chosen from each line. For most populations this fly was the daughter of the female which founded the line. As the founder female had been inseminated before collection, the flies used in this study can be regarded as a random sample from the wild. For those populations which did not consist of F_1 individuals (Finland, North Carolina, California, Kenya, and Zimbabwe), we randomly selected one allele for our analyses to account for the inbreeding during the propagation of the lines. The non-discarded data set, however, was used for F_{ST} calculations and the counts of alleles

occurring in Europe but absent in America (and vice versa).

Microsatellites

Genomic DNA was isolated for each line using a single female fly by the high salt extraction method (Miller, Dykes, and Polesky 1988). We typed 48 microsatellite loci (25 on the 2nd chromosome and 23 on the X chromosome; table 1). Further details are available from the authors' Web site: http://i122server.vu-wien.ac.at/. 10 μl polymerase chain reactions (PCR) were carried out with 100 ng of genomic DNA, ^{32}P-labeled forward primer,

Table 2
Mean Expected Heterozygosity, Variance in Repeat Number

	Europe		America		Africa	
	$\overline{H}$	$\overline{V}$	$\overline{H}$	$\overline{V}$	$\overline{H}$	$\overline{V}$
	0.52	9.16	0.57	11.23	0.75	14.16
	($\pm$0.16)	($\pm$11.86)	($\pm$0.15)	($\pm$13.82)	($\pm$0.14)	($\pm$15.16)
Europe[a]			0.0003	0.0089	0.0000	0.0029
America[a]					0.0000	0.0029

[a] P values for differences in variability among continents were determined by the Wilcoxon test for the mean variability at each locus in the two continents compared.

1.5 mM MgCl$_2$, 200 μM dNTPs, 1 μM of each primer, and 0.5 U *Taq* polymerase. A typical cycling profile consisted of 30 cycles for 50 s at 94°C, 50 s at 50° to 57°C (depending on the primer pair), and 50 s at 72°C. All polymerase chain reactions (PCR) were run with an initial denaturing step of 3 min at 94°C and a final extension of 45 min at 72°C for quantitative terminal transferase activity of the *Taq* polymerase. The PCR products were separated on 7% denaturing polyacrylamide gels (32% formamide, 5.6 M urea) and visualized by autoradiography. The PCR products were sized by loading a ''slippage ladder'' next to the amplified microsatellites (Schlötterer and Zangerl 1999).

Data Analysis

Measures of genetic variation, such as heterozygosity, variance in allele size, and number of alleles, were calculated using MS-Analyzer software, version 2.32 (Dieringer and Schlötterer 2003). When more than a single population was typed for one continent, estimates of variability were calculated for each population separately and subsequently averaged. This treatment was chosen to avoid the Wahlund effect (Hartl and Clark 1989).

The proportion of shared alleles was calculated by MSA software, version 2.32 (Dieringer and Schlötterer 2003). The obtained distance matrix was converted into a dendrogram using the Neighbor-Joining algorithm (Saitou and Nei 1987) provided with the PHYLIP software package (Felsenstein 1991) and graphically displayed with TreeView (Page 1996). The statistical significance of the nodes of the dendrogram were evaluated by bootstrapping (Efron and Gong 1983).

To estimate population differentiation, pairwise Θ values were determined as an unbiased estimate of F_{ST} (Weir and Cockerham 1984) using the MS-Analyzer program. In the following discussion we will refer to Θ values as F_{ST} values. The significance of pairwise F_{ST} values was tested by permuting genotypes among populations (10,000 times), as this method does not rely on Hardy-Weinberg assumptions (Goudet et al. 1996). To account for multiple testing, we used the Bonferroni method (Sokal and Rohlf 1995).

To determine the proportion of ancestry of individuals in African and non-African populations, we used the admixture model in the Structure program (Pritchard, Stephens, and Donnelly 2000), choosing a burn-in length of 50,000 steps, with 10^6 MCMC iterations. No prior information about the population structure was provided.

The program assumes the populations to be in Hardy-Weinberg equilibrium and the loci to be in linkage equilibrium. On the basis of the difference in allele frequencies, it assigns the individuals probabilistically to the assessed clusters, regardless of their geographical origin. For inbred individuals, only the randomly selected allele was used, and the second allele was entered as missing data (J. Pritchard, personal communication). To verify the results, we performed a second run of Structure with a differently discarded data set (see above) and obtained qualitatively similar results. The number of populations (groups) in our data set was estimated following the outline given by Pritchard, Stephens, and Donnelly (2002). In brief, the MCMC scheme was run for different values of MAXPOPS (K). For each K, Structure provides ln $P(X|K)$, from which we calculated the posterior probabilities of K assuming a uniform prior of K ($K \in \{1, 2, 3, 4, 5\}$), as described in the Structure manual.

A hierarchical analysis of molecular variance (AMOVA) using the ARLEQUIN program, version 2.000 (Schneider and Excoffier 1999), was employed to partition total variance components into those derived within and among three a priori defined groups: America, Europe, and Africa.

We used the Mantel test (10,000 permutations) as implemented in the GENEPOP program (Raymond and Rousset 1995) to test for a correlation between genetic and geographical distances (Rousset 1997). Geographical distances (in km) between sampling sites were obtained using a Web-based distance calculator (http://williams.best.vwh.net/gccalc.htm).

Results

A total of 48 microsatellite loci were analyzed in populations originating from three continents, Africa, Europe, and America; 23 loci mapped to the X chromosome and 25 to the 2nd chromosome. As already noted (Harr et al. 1998), we also found substantial differences in levels of variability among loci (table 1; see online Supplementary Material for a breakdown by populations). On average, European populations were less variable than American ones, independent of whether variation was measured by expected heterozygosity or the variance in repeat number ($P < 0.009$, Wilcoxon test; table 2). The same results were obtained when X-chromosomal and 2nd chromosomal microsatellites were analyzed separately (data not shown). Consistent with previous studies (David 1982; Begun and Aquadro 1993; Begun

Table 3
Genetic Differentiation Measured by Pairwise F_{ST} Values

	Finland	Poland	Germany	Denmark	Netherlands	Italy	New Jersey	California	Pennsylvania	North Carolina	Belize	Kenya	Zimbabwe
Finland		0.057	0.067	0.083	0.098	0.054	0.072	0.088	0.066	0.071	0.142	0.144	0.169
Poland	**		0.048	0.056	0.067	0.031	0.079	0.079	0.073	0.077	0.152	0.17	0.201
Germany	**	**		0.039	0.037	0.026	0.038	0.026	0.034	0.032	0.105	0.148	0.178
Denmark	**	**	**		0.045	0.04	0.06	0.055	0.051	0.059	0.123	0.154	0.18
Netherlands	**	**	**	**		0.052	0.056	0.04	0.048	0.054	0.124	0.166	0.205
Italy	**	**	**	**	**		0.048	0.048	0.053	0.05	0.122	0.159	0.187
New Jersey	**	**	**	**	**	**		0.023	0.01	0.004	0.044	0.113	0.135
California	**	**	**	**	**	**	**		0.027	0.032	0.085	0.165	0.191
Pennsylvania	**	**	**	**	**	**	NS	**		0.006	0.065	0.111	0.131
North Carolina	**	**	**	**	**	**	NS	**	NS		0.048	0.119	0.141
Belize	**	**	**	**	**	**	**	**	**	**		0.116	0.141
Kenya	**	**	**	**	**	**	**	**	**	**	**		0.014
Zimbabwe	**	**	**	**	**	**	**	**	**	**	**	NS	

NOTE.—Upper triangle: pairwise F_{ST} values between populations.

** $P < 0.05$ (after Bonferroni correction).

NS: not significant.

and Aquadro 1995; Aguadé 1998, 1999; Kauer et al. 2002), the African populations were the most variable ones (table 2). Although only a moderate number of loci for each chromosome were analyzed, we observed the same trend described by Kauer et al. (2002) for the X chromosome and third chromosome. Gene diversity in non-African populations was more reduced on X chromosomes than on autosomes (X chromosomes: 36% reduction, autosomes 26% reduction). This result suggests that similar evolutionary forces affect second and third chromosomes.

We used the analysis of molecular variance (AMOVA; Excoffier, Smouse, and Quattro 1992) to determine the partitioning of variation among populations and continents. Although most variation (86.88%) was contained within populations, the between-population (4.75%) and among-continent (8.37%) component of variation was found to be significant ($P < 0.0001$).

Pairwise F_{ST} analyses, however, indicated that the amount of genetic differentiation varied among populations. Within continents average F_{ST} values were low and all four non-significant pairwise F_{ST} values are observed in within-continent comparisons (table 3). Pairwise F_{ST} values were higher among European populations (0.053) than among American populations (0.034). Even though this difference was only marginally significant ($P = 0.0522$, Mann-Whitney U test) it suggests less differentiation among the American *D. melanogaster* populations.

All pairwise comparisons between African and non-African populations resulted in high F_{ST} values ranging from 0.11 to 0.20 (table 3). Between European and American populations average pairwise F_{ST} values were lower, averaging 0.071, but each of the comparisons was statistically significant (table 3). Interestingly, mean pairwise F_{ST} values between African and American *D. melanogaster* populations were significantly lower than mean pairwise F_{ST} values between African and European populations ($P = 0.0024$, Mann-Whitney U test). This observation is surprising, given that North American *D. melanogaster* populations were presumably colonized from temperate European populations (David and Capy 1988).

For further verification of the pairwise F_{ST} results, we calculated the proportion of shared alleles among all populations. The corresponding cladogram (fig. 1) confirms the F_{ST}-based results. European and African flies are the most distant populations. Although American and European flies are separated from the African flies by the same long branch, American flies form a separate clade, grouping closer to the African flies (fig. 1).

Given that *D. melanogaster* microsatellites have low mutation rates (Schlötterer et al. 1998; Schug et al. 1998), most microsatellite alleles are expected to have retained their original state since the recent out-of Africa habitat expansion approximately 10,000 years ago. Thus, a comparison of the allele distributions among continents should reflect largely demographic events, rather than mutational events. We determined the counts for those alleles which were confined to either European (289) or American (361) *D. melanogaster* populations. As these counts are highly dependent on the number of chromosomes analyzed in each group, we compared them to the counts for alleles shared between both continents [15,347 (European alleles shared with America) and 10,851 (American alleles shared with Europe)]. Note that this measurement counts the number of occurrences of certain alleles (e.g., alleles confined to Europe); for example, allele 124 at locus X3439769 occurs four times in American populations but is absent in European ones; the count would thus be four). Using a 2×2 contingency table, we found that the American *D. melanogaster* populations had significantly more continent-specific alleles than the European ones ($P < 0.0001$, Fisher's exact test).

A recently introduced model-based clustering approach for multilocus genotype data can be used to infer the probability of individuals having ancestry in any of the specified populations. We first determined the most likely number of clusters in our data set. The highest posterior probability was found for three clusters (table 4) corresponding to the three continents. Next, we determined the proportion of shared ancestral genotypes for the African cluster (defined by the highest proportion of ancestral alleles in Africa). The mean proportion of shared
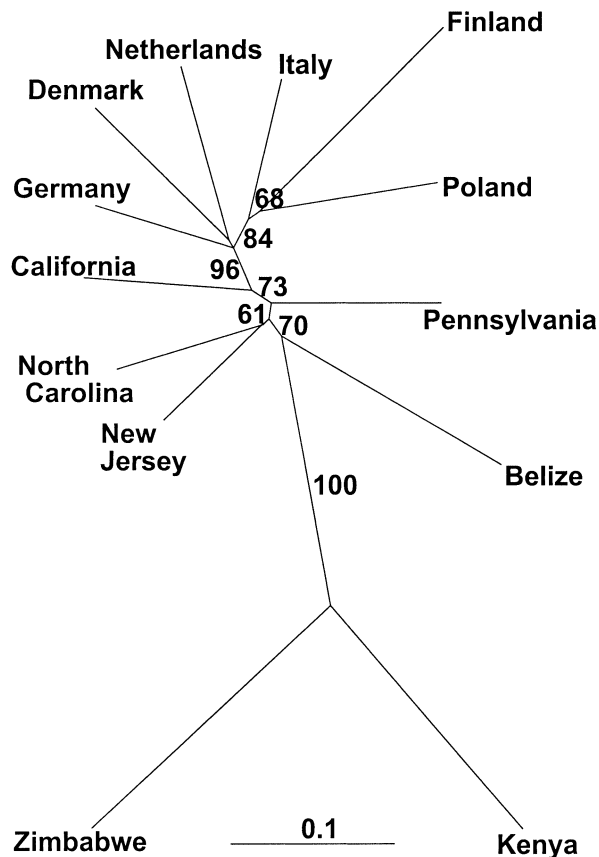
Fig. 1.—Neighbor-Joining tree based on the proportion of shared alleles. Bootstrap values indicate the statistical support for the corresponding node. Only bootstrap values above 50 are shown.

Table 4
Estimated Posterior Probabilities of *K*, the Number of *D. melanogaster* Populations

K	lnP(X|K) 1,000,000 iterations	P(K|X)[a]
1	−32871.4	<0.0001
2	−31692.8	<0.0001
3	−31123.8	≈1
4	−31142.3	<0.0001
5	−31158.3	<0.0001

[a] Assuming a uniform prior for K ($K \in \{1, 2, 3, 4, 5\}$).

As expected for the very recent colonization of America, the differentiation among American and European populations was low, but statistically significant. In contrast to expectations, American *D. melanogaster* populations were more polymorphic than European populations. Because of the low mutation rate of *D. melanogaster* microsatellites and the recent colonization history, a larger effective population size of the American populations cannot account for the higher variability. A further discrepancy in the prevailing demographic model was revealed by an analysis of genetic distances. Rather than grouping European populations between American and African populations, a cladogram based on the proportion of shared alleles placed American populations between the European and African ones (fig. 1).

Despite these apparent discrepancies, the overall similarity between European and American populations strongly suggests that American flies were derived from European ones, but the demographic history of the American populations is more complex. One possibility is that the American *D. melanogaster* populations may have gained additional variation by admixture from other populations. Our study did not include South American and Caribbean populations, but they provide good potential source populations. South American *D. melanogaster* populations are still very poorly studied, but more information is available for *D. melanogaster* from the Caribbean Islands. Based on body size, allozyme frequencies, hydrocarbon composition, and sequence variation at the *desat* locus, Caribbean *D. melanogaster* clearly display African traits (Capy et al. 1986; Takahashi et al. 2001). Assuming that those African traits originate from a different wave of colonization than the European colonization (David and Capy 1988; Takahashi et al. 2001), Caribbean populations are a potential source of African alleles. Migration from Caribbean/South American populations toward Northern America would have introduced a different set of alleles ultimately derived from Africa. Hence, admixture from Caribbean (or South American) populations could explain why American populations are more variable than European ones. Furthermore, as admixture provides other African alleles not contained in the European populations, it also explains the higher proportion of continent specific alleles in the American populations.

Finally, the phylogenetic position of the American populations between European and African populations (fig. 1) also suggests that the American populations have some additional African alleles.

ancestral African alleles averaged 0.28% in European flies, whereas American flies had on average 0.41% shared ancestral genotypes. Given that Structure detects only recent admixture, the small values of shared African ancestry inferred by Structure indicate that the differences between American and European populations cannot be attributed to recent admixture.

Discussion

The first wave of out-of-Africa habitat expansion of *D. melanogaster* occurred about 10,000 years ago and included Europe (David and Capy 1988). Around the middle of the 19th century, North American *D. melanogaster* populations were colonized from Europe (David and Capy 1988). Given that colonization events are often associated with a loss of variability, North American flies are expected to harbor only a subset of the European allelic variation, which in turn is a subset of the African diversity. This effect is expected to be particularly pronounced if the number of founders was small, as suggested for the colonization of American *D. subobscura* (Pascual et al. 2001).

Consistent with previous reports, our study also confirmed the pronounced differences between African and non-African *D. melanogaster* populations, with non-African populations being less variable than African ones.

Origin of Central American Populations

Given that Caribbean flies were shown to harbor several African traits (David and Capy 1988; Takahashi et al. 2001), their relationship to continental Central American flies may provide further insight into the colonization history of *D. melanogaster* in America. Our study indicates that flies from the Central American mainland (Belize) share more similarity with the European and North American flies than with African *D. melanogaster*. Nevertheless, figure 1 clearly shows that Belize is the non-African population grouping closest to the African ones. As our study did not include flies from the Caribbean, and the Belize population was not analyzed for the presence of phenotypic African traits, the comparison to a recent microsatellite survey may be informative. Schlötterer, Vogl, and Tautz (1997) characterized microsatellite variation in two African, one European, and two Caribbean populations. Despite the well-described African traits, Caribbean flies grouped with the European flies, and not with the African populations (Schlötterer, Vogel, and Tautz 1997). Within the limits of a study based on 10 microsatellites only, the data suggest that Caribbean populations are genetically very close to the populations on the North American continent, similar to the results of this study for the Belize population.

Population Substructure Within Continents

Mark recapture experiments suggested a large dispersal capacity of North American *D. melanogaster* (Coyne and Milstead 1987). Molecular studies based on restriction fragment length polymorphism (RFLP) data of the *Adh* region (Kreitman and Aguadé 1986) also support the hypothesis of high gene flow preventing population differentiation in North America. In contrast, a RFLP study of the *Pgd* locus detected differences among American populations (Begun and Aquadro 1994). Several allozyme studies also inferred population differentiation among American populations (Johnson and Schaffer 1973; Singh and Long 1992). The interpretation of these results, however, is significantly complicated by the frequent non-neutral behavior of allozyme polymorphism. Our study relied on a large number of microsatellite loci, which are largely evolving neutrally (Michalakis and Veuille 1996; Schlötterer 2000). Despite the possibility that some microsatellite loci may be affected by linkage to a selected gene (Slatkin 1995; Schlötterer and Wiehe 1999), the analysis of a large number of microsatellites should reflect demography rather than selection.

In addition to the significant differentiation among continents, we also found small but significant differences among populations within continents. While all pairwise F_{ST} values among European populations were significant, not all American populations were significantly differentiated. Also, mean pairwise F_{ST} values were higher in Europe than in America. This difference among European and American populations may be a reflection of higher rates of gene flow in America. Using the Mantel test, however, no significant correlation between genetic differentiation (F_{ST}) and geographic distance could be detected for either continent ($P > 0.1$). An alternative explanation accounts for the more recent colonization of the American populations (David and Capy 1988). Assuming that a large number of individuals spread over the continents within a very short time after the colonization event, then no differentiation between populations is expected. Subsequently, local populations are formed which adapt to their local environment. Depending on the effective population size of local populations, they diverge from others mainly by genetic drift. This hypothesis has recently been used to explain the fact that in Australia adjacent populations were found to be more differentiated than more distantly located ones (Agis and Schlötterer 2001). Given that Europe was colonized earlier than America (David and Capy 1988), genetic drift may have operated for a longer time in Europe than in America, leading to a higher genetic differentiation.

Nevertheless, a closer inspection of the pairwise F_{ST} values among the American populations indicates that only those comparisons, which included either the Central American population from Belize or the population from California, resulted in significant F_{ST} values. No significant differentiation was detected among the three populations collected at the North American East Coast (table 3). Interestingly, together with Belize, these populations had the highest proportion of putative African alleles (data not shown). The Californian population, however, appeared more like the European populations. This is also reflected in the phylogenetic tree, which groups the Californian population closest to the European cluster. These data suggest that the admixture of African alleles was mainly restricted to the East Coast, but they need to be confirmed with a larger sample including more populations from the West Coast. Additional support for genetic differentiation between East and West Coast *D. melanogaster* populations, is provided by an RFLP survey which revealed differences at the *Pgd* locus between populations collected in California and North Carolina (Begun and Aquadro 1994).

Absence of Clinal Microsatellite Variation on the American East Coast

Both allozyme and sequence polymorphism studies indicated the presence of clinal variation on the east American coast (Berry and Kreitman 1993; Long and Singh 1995; Sawyer et al. 1997; Schmidt, Duvernell, and Eanes 2000). Based on the evidence for admixture of African alleles, these clines may be explained by a continuous gene flow from South to North along the East Coast. Under this hypothesis, a correlation between geographic distance and genetic differentiation would be expected. Our microsatellite analysis, however, did not indicate such a correlation whether the Central American population was included or not ($P > 0.1$, Mantel test). While we note that our analysis of four populations may not have enough power to detect clinal variation, this observation is consistent with a study of sequence variation at the *Adh* locus, which described clinal variation only for selected sites, but not for silent variation (Berry and Kreitman 1993). While the authors attributed their results to high rates of gene flow in combination with

selection, it is also consistent with a recent colonization event leading to similar allele frequencies at not selected sites (Agis and Schlötterer 2001). Selection operating after the colonization could have resulted in the observed cline, even in the absence of a high continuous gene flow among populations.

Implications for Standard Neutrality Tests

Given that allele frequency spectrum and haplotype-based neutrality tests are extremely sensitive to demographic events (Nielsen 2001), our finding that North American *D. melanogaster* populations show admixture with African alleles emphasizes that the interpretation of allele frequency spectra in the North American *D. melanogaster* population requires special caution. Preferably, true multilocus tests (Kim and Stephan 2002; Schlötterer 2002; Wall, Andolfatto, and Przeworski 2002) should be used, as these tests do not only rely on single loci but also attempt to capture demography by accounting for variation among loci. With the progress of the DNA sequencing technology, true multilocus sequencing studies in *D. melanogaster* have become feasible and are expected to provide a more accurate answer on the extent to which demography and selection have shaped patterns of variability in *D. melanogaster*.

Acknowledgments

Literature Cited

Agis, M., and C. Schlötterer. 2001. Microsatellite variation in natural *Drosophila melanogaster* populations from New South Wales (Australia) and Tasmania. Mol. Ecol. **10**:1197–1205.

Aguadé, M. 1998. Different forces drive the evolution of the *Acp26Aa* and *Acp26Ab* accessory gland genes in the *Drosophila melanogaster* species complex. Genetics **150**:1079–1089.

———. 1999. Positive selection drives the evolution of the *Acp29AB* accessory gland protein in *Drosophila*. Genetics **152**:543–551.

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.

———. 1994. Evolutionary inferences from DNA variation at the 6-phosphogluconate dehydrogenase locus in natural populations of *Drosophila*: selection and geographic differentiation. Genetics **136**:155–171.

———. 1995. Molecular variation at the *vermilion* locus in

geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. Genetics **140**:1019–1032.

Berry, A., and M. Kreitman. 1993. Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* an the east coast of North America. Genetics **134**:869–893.

Capy, P., J. David, R. Allemand, Y. Carton, G. Febvay, and A. Kermarec. 1986. Genetic analysis of *Drosophila melanogaster* in the French West Indies and comparison with pops from other parts of the world. Genetica **69**:167–176.

Coyne, J. A., and B. Milstead. 1987. Long-distance migration of *Drosophila*: dispersal of *D. melanogaster* alleles from a Maryland orchard. Am. Nat. **130**:70–82.

David, J. R. 1982. Latitudinal variability of *Drosophila melanogaster*: allozyme frequencies divergence between European and Afrotropical populations. Biochem. Genet. **20**:747–762.

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

Dieringer, D., and C. Schlötterer. 2003. Microsatellite analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. Mol. Ecol. Notes. **3**:167–169.

Efron, B., and G. Gong. 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. Am. Stat. **37**:36–48.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131**:479–491.

Felsenstein, J. 1991. PHYLIP (phylogeny inference package). Version 3.57c. Distributed by the author, Department of Genetics, University of Washington, Seattle.

Goudet, J., M. Raymond, T. de Meeüs, and F. Rousset. 1996. Testing differentiation in diploid populations. Genetics **144**:1933–1940.

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.

———. 1991. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history vs. selection in the genetic structure of geographic populations. Genetics **1991**:103–117.

Harr, B., B. Zangerl, G. Brem, and C. Schlötterer. 1998. Conservation of locus specific microsatellite variability across species: a comparison of two Drosophila sibling species *D. melanogaster* and *D. simulans*. Mol. Biol. Evol. **15**:176–184.

Hartl, D. L., and A. G. Clark. 1989. Principles of population genetics, second edition. Sinauer Associates, Sunderland, Mass.

Hickey, D. A. 1979. The geographical pattern of an enzyme polymorphism in *D. melanogaster*. Genetica **51.1**:1–4.

Johnson, F. M., and H. E. Schaffer. 1973. Isozyme variability in species of the genus *Drosophila*. VII. Genotype-environment relationships in populations of *D.melanogaster* from the eastern United States. Biochem. Genet. **10**:149–163.

Jones, J. S., S. H. Bryant, R. C. Lewontin, J. A. Moore, and T. Prout. 1981. Gene flow and the geographical distribution of a molecular polymorphism in *Drosophila pseudoobscura*. Genetics **98**:157–178.

Kauer, M., B. Zangerl, D. Dieringer, and C. Schlötterer. 2002. Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *Drosophila melanogaster*. Genetics **160**:247–256.

Kim, Y., and W. Stephan. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics **160**:765–777.

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, and M. Ashburner. 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* C. B. Krimbas and J. R. Powell, eds. Drosophila inversion polymorphism. CRC Press, Cleveland.

Lewontin, R. C., and J. L. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics **54**: 595–609.

Long, A. D., and R. S. Singh. 1995. Molecules versus morphology: the detection of selection acting on morphological characters along a cline in *Drosophila melanogaster*. Heredity **74**:569–589.

Mettler, L. E., R. A. Voelker, and T. Mukai. 1977. Inversion clines in populations of *Drosophila melanogaster*. Genetics **87**:169–176.

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 for extracting DNA from human nucleated cells. Nucleic Acids Res. **16**:1215.

Nielsen, R. 2001. Statistical tests of selective neutrality in the age of genomics. Heredity **86**:641–647.

Oakeshott, J. G., J. B. Gibson, P. R. Anderson, W. R. Knibb, D. G. Anderson, and G. K. Chambers. 1982. Alcohol dehydrogenase and glycerold-3-phosphate dehydrogenase clines in *Drosophila melanogaster* of different continents. Evolution **36**:86–96.

Page, R. D. M. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12**:357–358.

Pascual, M., C. F. Aquadro, V. Soto, and L. Serra. 2001. Microsatellite variation in colonizing and palearctic populations of *Drosophila subobscura*. Mol. Biol. Evol. **18**: 731–740.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945–959.

Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenism. J. Hered. **86**:248–249.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. Genetics **145**:1219–1228.

Saitou, R. K., and M. Nei. 1987. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

Sawyer, L. A., J. M. Hennessy, A. A. Peixoto, E. Rosato, H. Parkinson, R. Costa, and C. P. Kyriacou. 1997. Natural variation in a Drosophila clock gene and temperature compensation. Science **278**:2117–2120.

Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma **109**:365–371.

———. 2002. A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics **160**: 753–763.

Schlötterer, C., R. Ritter, B. Harr, and G. Brem. 1998. High mutation rates of a long microsatellite allele in *Drosophila melanogaster* provide evidence for allele-specific mutation rates. Mol. Biol. Evol. **15**:1269–1274.

Schlötterer, C., C. Vogl, and D. Tautz. 1997. Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. Genetics **146**:309–320.

Schlötterer, C., and T. Wiehe. 1999. Microsatellites, a neutral marker to infer selective sweeps. Pp. 238–248 *in* D. Goldstein and C. Schlötterer, eds. Microsatellites—evolution and applications. Oxford University Press, Oxford.

Schlötterer, C., and B. Zangerl. 1999. The use of imperfect microsatellites for DNA fingerprinting and population genetics. Pp. 153–165 *in* J. T. Epplen and T. Lubjuhn, eds. DNA profiling and DNA fingerprinting. Birkhäuser, Basel.

Schmidt, P. S., D. D. Duvernell, and W. F. Eanes. 2000. Adaptive evolution of a candidate gene for aging in Drosophila. Proc. Natl. Acad. Sci. USA **97**:10861–10865.

Schneider, S., and L. Excoffier. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when mutation rates vary among sites: application to human mitochondrial DNA. Genetics **152**:1079–1089.

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

Singh, R. S. 1989. Population genetics and evolution of species related to *Drosophila melanogaster*. Annu. Rev. Genet. **23**: 425–453.

Singh, R. S., D. A. Hickey, and J. David. 1982. Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. Genetics **101**:235–256.

Singh, R. S., and A. Long. 1992. Geographic variation in Drosophila: 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 populations of *Drosophila melanogaster*. I. Estimates of gene flow from rare alleles. Genetics **115**:313–322.

Slatkin, M. 1995. Hitchhiking and associative overdominance at a microsatellite locus. Mol. Biol. Evol. **12**:473–480.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. W. H. Freeman and Company, New York.

Takahashi, A., S. C. Tsaur, J. A. Coyne, and C. I. Wu. 2001. The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **98**:3920–3925.

Voelker, R. A., C. C. Cockerham, F. M. Johnson, H. E. Schaffer, T. Mukai, and L. E. Mettler. 1978. Inversions fail to account for allozyme clines. Genetics **88**:515–527.

Wall, J. D., P. Andolfatto, and M. Przeworski. 2002. Testing models of selection and demography in *Drosophila simulans*. Genetics **162**:203–216.

Weir, B. S., and C. C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution **38**:1358–1370.