

# History and Structure of Sub-Saharan Populations of *Drosophila melanogaster*

John E. Pool<sup>1</sup> and Charles F. Aquadro

Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853

Manuscript received March 29, 2006

Accepted for publication August 10, 2006

## ABSTRACT

*Drosophila melanogaster* is an important model organism in evolutionary genetics, yet little is known about the population structure and the demographic history of this species within sub-Saharan Africa, which is thought to contain its ancestral range. We surveyed nucleotide variation at four 1-kb fragments in 240 individual lines representing 21 sub-Saharan and 4 Palearctic population samples of *D. melanogaster*. In agreement with recent studies, we find a small but significant level of genetic differentiation within sub-Saharan Africa. A clear geographic pattern is observed, with eastern and western African populations composing two genetically distinct groups. This pattern may have resulted from a relatively recent establishment of *D. melanogaster* in western Africa. Eastern populations show greater evidence for long-term stability, consistent with the hypothesis that eastern Africa contains the ancestral range of the species. Three sub-Saharan populations show evidence for cosmopolitan introgression. Apart from those cases, the closest relationships between Palearctic and sub-Saharan populations involve a sample from the rift zone (Uganda), suggesting that the progenitors of Palearctic *D. melanogaster* might have come from this region. Finally, we find a large excess of singleton polymorphisms in the full data set, which is best explained by a combination of population growth and purifying selection.

**D**ROSOPHILA *melanogaster*, the common pomace fly or vinegar fly, is an important model organism in many areas of evolutionary genetics. In particular, considerable research has focused on detecting the action of positive selection from patterns of nucleotide variation in this species (e.g., HARR *et al.* 2002; KAUER *et al.* 2002; ORENKO and AGUADÉ 2004; BAUER DUMONT and AQUADRO 2005; OMETTO *et al.* 2005; POOL *et al.* 2006). However, a deviation from the null hypothesis of neutral evolution in an equilibrium population might be caused either by a departure from selective neutrality or by a nonequilibrium population history. For example, negative values of the Tajima's *D* statistic are commonly interpreted as evidence of a recent selective sweep, yet demographic forces such as population growth, migration, and severe bottlenecks are equally capable of generating such a result (TAJIMA 1989). Other proposed tests of neutral evolution, such as Fay and Wu's *H* (FAY and WU 2000) and the likelihood method of KIM and STEPHAN (2002), are also strongly influenced by demography (e.g., DEPAULIS *et al.* 2003; JENSEN *et al.* 2005). Therefore, the inference of selection from

population genetic data relies critically on an accurate understanding of demographic history.

In the case of *D. melanogaster*, our demographic knowledge is surprisingly limited. It is generally accepted that *D. melanogaster* had an ancestral range in Africa and has reached some parts of the world quite recently. On the basis of current species distributions, polymorphic characters, and ecological information, LACHAISE *et al.* (1988) suggest an Afrotropical origin for *D. melanogaster*, specifically in western or central equatorial Africa. Studies of molecular polymorphism have supported the African origin of *D. melanogaster*. For example, BEGUN and AQUADRO (1993) found that DNA sequence variation at seven X-linked loci was more than twice as high in a Zimbabwe sample compared to a sample from the United States. Additional studies involving microsatellites (e.g., KAUER *et al.* 2002) and DNA sequence variation (e.g., OMETTO *et al.* 2005) have also found higher genetic variation in sub-Saharan than in cosmopolitan populations.

The expansion of *D. melanogaster* from sub-Saharan Africa into northern Africa and Eurasia is thought to have occurred on the order of 10,000 years ago. LACHAISE *et al.* (1988) hypothesized that *D. melanogaster* might have crossed the central Sahara between 6500 and 9500 years ago, a period when this region may have been less arid. On the basis of the number of apparently new mutations in cosmopolitan samples, BAUDRY *et al.* (2004) estimate a time of 6400 years since the out-of-Africa bottleneck, although this estimate may represent

**This article is dedicated to the memory of Daniel Lachaise, whose contributions to the field of *Drosophila* evolutionary biology helped to lay the groundwork for numerous studies, including this one.**

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries with accession nos. DQ842636–DQ843599.

<sup>1</sup>Corresponding author: Department of Molecular Biology and Genetics, Biotechnology Bldg., Cornell University, Ithaca, NY 14853.  
E-mail: jep36@cornell.edu

the time since population size recovery. THORNTON and ANDOLFATTO (2006) recently applied a Bayesian inference method to infer bottleneck parameters from multilocus nucleotide variation in Zimbabwe and Netherlands populations, obtaining a point estimate of 16,000 years for the time since the bottleneck began (assuming 10 generations/year and an inferred initial effective population size of  $2.4 \times 10^6$ ). However, we do not know where within sub-Saharan Africa the progenitors of Palearctic *D. melanogaster* originated, and therefore the suitability of Zimbabwe as a proxy for the cosmopolitan source population is uncertain.

Populations of *D. melanogaster* from Europe, northern Africa, and the Middle East typically show low levels of genetic differentiation (e.g., BAUDRY *et al.* 2004; DIERINGER *et al.* 2005; SCHLÖTTERER *et al.* 2006). In contrast, Asian populations appear to be more structured (BAUDRY *et al.* 2004; SCHLÖTTERER *et al.* 2006), but are still thought to share a single, common origin with other Palearctic populations. The subsequent expansion of *D. melanogaster* to other parts of the world (such as the Americas and Australia) is thought to have occurred within the past few hundred years (DAVID and CAPY 1988), primarily because *D. melanogaster* is absent from the earliest insect collections in these regions. The primary source for New World populations of *D. melanogaster* appears to be Europe, but varying levels of direct African ancestry have also been inferred (CARACRISTI and SCHLÖTTERER 2003).

Rather little is known about the genetic structure and population history of *D. melanogaster* within sub-Saharan Africa. Although the divergence time between *D. melanogaster* and *D. simulans* has been estimated at 2–3 million years (MY) (LACHAISE *et al.* 1988; RUSSO *et al.* 1995), TAMURA *et al.* (2004) recently suggested a divergence time of 5.4 MY (after accounting for codon bias at 62 genes). According to the biogeographic hypothesis of LACHAISE *et al.* (1988), this speciation event left *D. melanogaster* occupying a central African range west of the African Rift, with *D. simulans* inhabiting the eastern African coast and/or islands such as Madagascar. Both *D. melanogaster* and *D. simulans* eventually shifted from wild-living to human commensal species (although *D. melanogaster* is more strictly commensal), but the timing of these transitions is unknown (LACHAISE and SILVAIN 2004).

Today, *D. melanogaster* can be found across the majority of sub-Saharan Africa, but relatively few studies of molecular variation have sampled more than one or two sub-Saharan populations. Early studies of sequence variation, using Kenya and Zimbabwe samples, found a very low level of differentiation between these populations (BEGUN and AQUADRO 1995), a finding supported by subsequent work (e.g., CARACRISTI and SCHLÖTTERER 2003; HADDRILL *et al.* 2005).

In contrast, studies of silent variation among *Adh* slow haplotypes (BÉNASSI and VEUILLE 1995) and length

variation in trinucleotide coding repeats (MICHALAKIS and VEUILLE 1996) found substantial differentiation between an eastern (Malawi) and a western African (Ivory Coast) population. In both of these studies, the greatest divergence was found between Malawi, on the one hand, and Ivory Coast and the cosmopolitan samples on the other hand. The loci sampled in these two studies were located on the second chromosome, and the high level of African differentiation observed might have been influenced by the *In(2L)t* inversion. This inversion, like several others, shows strong differences in frequency between sub-Saharan populations (LEMEUNIER and AULARD 1992; AULARD *et al.* 2002), which might be related to the action of selection.

Three recent studies of molecular polymorphism have included more than two sub-Saharan population samples of *D. melanogaster*. First, a study of sequence polymorphism at four X-linked protein-coding genes included two eastern (Kenya, Zimbabwe) and two western (Ivory Coast, Niger) populations, along with several cosmopolitan samples (BAUDRY *et al.* 2004). The authors reported significant  $F_{ST}$  values (0.050–0.072) between most pairs of sub-Saharan samples, with the exception of Kenya–Zimbabwe (0.026) and Niger–Zimbabwe (0.016), concluding that small but significant differentiation exists within sub-Saharan Africa. The Kenya sample was found to possess many of the common cosmopolitan haplotypes, and the authors proposed that cosmopolitan *D. melanogaster* may have originated in eastern Africa. DIERINGER *et al.* (2005) conducted a microsatellite analysis that included 10 sub-Saharan population samples from five different countries. These sub-Saharan samples were found to be significantly different from northern African samples (which were more similar to European populations). A low level of subdivision within sub-Saharan Africa was suggested, but a clear geographic pattern did not emerge, and genetic distance was more highly correlated with collection date than with geographic distance. HADDRILL *et al.* (2005) surveyed nucleotide variation at 10 X-linked noncoding fragments in Gabon, Kenya, and Zimbabwe samples from sub-Saharan Africa, along with The Netherlands and Pennsylvania cosmopolitan samples. These populations were found to group in three genetic clusters: Gabon, Kenya–Zimbabwe, and Netherlands–Pennsylvania. Compared with the eastern African populations (Kenya and Zimbabwe), the western African population (Gabon) had lower variation and greater linkage disequilibrium, which might suggest a relatively recent founding for this population. However, demographic inferences based on a few population samples are necessarily tentative, and the limited geographic sampling of molecular variation from sub-Saharan *D. melanogaster* has not provided a clear understanding of the history of these populations or of the spatial pattern of genetic structure.

**TABLE 1**  
**Population samples included in this study**

Abbreviation	Country	Locality	<i>n</i> <sup>a</sup>	Date collected	Collector
Ch	China	Beijing	10	9/1992	X. Huang
Fr	France	Bordeaux	10	7/1998	E. Gravot and M. Hamblin
Is	Israel		10		
Tu	Tunisia		9		
Gu	Guinea	Dondé	10	6/2005	B. B. Sow
Nr	Niger	Kareygorou	10	12/2004	D. Ali
Bn <sup>b</sup>	Benin	Cotonou	9	5/2004	G. Goergen
Ng	Nigeria	Maiduguri	10	9/2004	D. Gwary and B. Sastawa
CM, Cam-M	Cameroon	Maroua	10	3/2004	J. Pool
CD, Cam-D	Cameroon	Mbalang-Djalingo	10	3/2004	P. Konje and J. Pool
CW, Cam-W	Cameroon	Mbengwi	6	4/2004	J. Pool and E. Ajeah
CN, Cam-N	Cameroon	Nkouondja	10	4/2004	J. Pool
CO, Cam-O	Cameroon	Oku	10	4/2004	J. Pool
CY, Cam-Y	Cameroon	Yokadouma	8	4/2004	P. Konje and J. Pool
Ga	Gabon	N'Toum	10	3/2002	B. Ballard and S. Charlat
CB <sup>c</sup>	Congo	Brazzaville	10	1999	J. David
Er	Eritrea	Asmara	10	7/2004	S. Ghebremeskel Hailu
Ug	Uganda	Namulonge	10	4/2005	J. Ogwang
Rw	Rwanda	Gisakura	8	6/2005	S. Margles
KN, <sup>d</sup> Kenya-N	Kenya	Naivasha	10	7/1988	R. Woodruff
KM, Kenya-M	Kenya	Malindi	10	7/2002	B. Ballard
Mw	Malawi	Mwanza	10	7/2001	B. Ballard
ZK, Zim-K	Zimbabwe	Lake Kariba	10	5/1994	T. Mutangadura
ZS, Zim-S	Zimbabwe	Sengwa	10	9/1990	R. Ramey
SA	South Africa	Stellenbosch	10	2/1999	J. Lowe

<sup>a</sup> Number of individuals sequenced (from separate isofemale lines).

<sup>b</sup> Sample consists of seven isofemale lines from Benin and two isofemale lines from Kade, Ghana (F. Ansah and E. Okai, 9/2004).

<sup>c</sup> Mass cultured for several generations before separation into isofemale lines.

<sup>d</sup> Sample consists of six isofemale lines from Naivasha and four isofemale lines from North Kinangop, Kenya (sites <50 km apart).

Here we report an analysis of nucleotide variation from 25 population samples of *D. melanogaster* (Table 1), 21 of which are sub-Saharan (Figure 1). Data were obtained from a total of 240 isofemale lines at four 1-kb X-linked fragments, each located in a highly recombining, long intergenic region. The broad geographic sampling allowed us to identify a distinct geographic pattern of genetic differentiation and to make inferences regarding the history of *D. melanogaster* in sub-Saharan Africa and the founding of Palearctic populations.

## MATERIALS AND METHODS

DNA sequence data were obtained from up to 10 isofemale lines from each of the population samples of *D. melanogaster* listed in Table 1 and from a single male *D. simulans* (from an isofemale line collected in Gisakura, Rwanda, in 2005). DNA was isolated from a homozygous X chromosome line or from a single male fly representing each isofemale line, using a method adapted from a protocol from G. Rubin (<http://www.bio.com/protocolstools/protocol.jhtml?id=p82>).

Four X-linked loci were amplified and sequenced; each was at least 10 kb from any annotated gene and located in a region of high recombination (KINDAHL 1994; HEY and KLIMAN 2002). The loci studied are designated by their cytological locations: 4F2, 8A4, 11A5, and 12F1. These locations corre-

spond approximately to genetic map positions 11, 24, 38, and 49 cM, respectively. Further information concerning the precise location, neighboring genes, and PCR and sequencing primers for these loci can be found in supplemental Table S1 at <http://www.genetics.org/supplemental/>.

PCR products were purified using shrimp alkaline phosphatase and exonuclease I enzymes. Sequencing reactions using ABI BigDye ddNTPs were read by an ABI 3730 capillary sequencer at the Biotechnology Resource Center of Cornell University (<http://www.brc.cornell.edu>). Sequence reads were aligned using Sequencher 4.2. Chromatograms were visually inspected at all putatively polymorphic sites. Data consisted of both single- and double-stranded sequence coverage, and as expected, no evidence of heterozygosity was observed. Final alignments were made using the CLUSTALW algorithm as implemented in MegAlign 5.08 and manually corrected.

HKA tests (HUDSON *et al.* 1987) were conducted using an online, multilocus implementation (<http://lifesci.rutgers.edu/~hey/lab/HeylabSoftware.htm>). Most DNA summary statistics, including Tajima's *D* (TAJIMA 1989) and the  $F_{ST}$  of HUDSON *et al.* (1992), were calculated using DnaSP 4.0 (ROZAS *et al.* 2003).  $F_{ST}$  distance trees were constructed using the neighbor-joining algorithm implemented in MEGA 3.1 (KUMAR *et al.* 2004). The nearest neighbor statistic ( $S_{nn}$ ) of HUDSON (2000) tests how often a sequence's nearest neighbor (most similar sequence) is from the same geographic sample.  $S_{nn}$  matrices and corresponding *P*-values (from 10,000 permutations) for each locus were obtained using libsequence (THORNTON 2003). The four locus-specific  $S_{nn}$  *P*-values

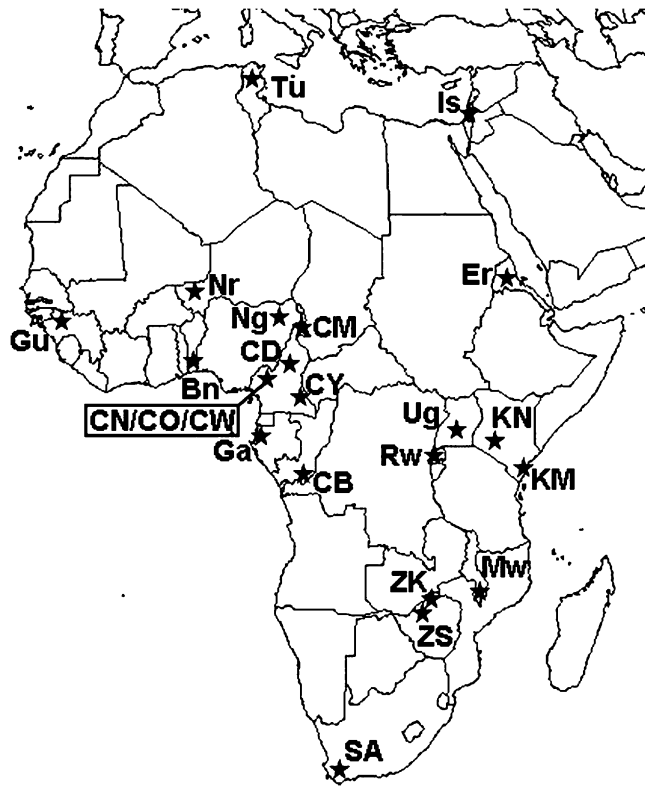


FIGURE 1.—Geographic location of population samples used in this study. Population samples are identified by the two-letter abbreviations given in Table 1.

obtained for each population comparison were combined using the method described by BAILEY and GRIBSKOV (1998). Analysis of molecular variance (AMOVA; EXCOFFIER *et al.* 1992) and Mantel matrix correspondence tests (MANTEL 1967) for the relationship between genetic and geographic distances were performed using Arlequin 3.0 (EXCOFFIER *et al.* 2005).

In addition to the summary statistics listed above, genetic differentiation was also examined using the Bayesian clustering method Structure 2.1 (PRITCHARD *et al.* 2000; FALUSH *et al.* 2003), which estimates the proportion of each individual's ancestry that derives from each of  $K$  genetic subpopulations. The analysis was run using the linkage model, which takes into account the genetic distances between sites. Since the relative positions for the linkage model only need to be proportional to genetic distances, the physical distances between sites were used as a proxy. Structure was run for a burn-in length of 1,000,000 replicates and a run length of 1,000,000 replicates. Results were graphically depicted using Distruct (ROSENBERG 2004).

Unlike most of the analyses performed, Structure does not rely on a specific mutation model and can therefore accommodate both substitution and indel polymorphism. We included indel regions and sites with more than two segregating nucleotides in the Structure data, according to the following rules, which were intended to maximize the information retained while not counting an indel more than once: (1) Sites with singleton insertions were omitted; (2) singleton deletions were coded as missing data; (3) for nonsingleton indels of 1 bp, the deletion was coded as a fifth character state; and (4) for nonsingleton indels  $>1$  bp, the indel region was collapsed to represent a single variable site, and each unique haplotype within the indel region was considered a separate allele.

Linkage disequilibrium was assessed by estimating the population recombination rate ( $\rho$ ). For each population at each locus,  $\rho$  was estimated using the method of HUDSON (2001), as implemented in LDhat 2.0 (<http://www.stats.ox.ac.uk/~mcvean/LDhat/>). Log likelihoods for values of  $\rho$  from 0 to 1000 (per 1-kb locus, estimated with an interval size of 1) were summed across loci to obtain a composite likelihood surface and point estimate of  $\rho$  for each population. Because the estimation of  $\rho$  depends on the presence of variation, higher values of  $\theta$  may lead to higher estimates of  $\rho$ . To remove this effect, estimates of  $\rho$  were divided by  $\pi$  (an estimator of  $\theta$ ), as suggested by HADDRILL *et al.* (2005).

A simulation approach was used to test whether population growth could account for the number of singleton polymorphisms observed in the full data. The coalescent simulator ms (HUDSON 2002) was utilized to simulate sets of four genetically unlinked loci resembling those examined in this study, each with a sample size of 240, a length of 768 bp, 110 segregating sites, and a population recombination rate ( $\rho$ ) of 150/locus. The above length and number of segregating sites reflect the portion of our data for which ancestral states could be inferred from the *D. simulans* sequence, averaged across the four loci.  $\rho = 150$  is a conservative estimate for the level of recombination, since the recombination rate for the loci studied is likely to be  $>3 \times 10^{-8}$  recombination events/bp/generation (KINDAHL 1994; HEY and KLIMAN 2002), the effective population size is  $\sim 2.4 \times 10^6$  (THORNTON and ANDOLFATTO 2006), the full length of each locus is actually  $>1000$  bp, and therefore  $\rho (3N_e\mu)$  should be at least  $3 \times (2.4 \times 10^6) \times (3 \times 10^{-8}) \times 1000 = 216$ .

## RESULTS

**Variability:** A total of 4202 consensus sites were sequenced, of which 3222 were deletion free among the 240 *D. melanogaster* chromosomes sampled. These deletion-free sites contained a total of 506 variable sites, of which 471 were biallelic. Consistent with previous studies, we find that sub-Saharan populations are consistently more variable than cosmopolitan populations. As shown in Table 2, the sub-Saharan sample with the lowest nucleotide diversity (Niger: 0.0072) is substantially more variable than the most diverse Palearctic sample (Tunisia: 0.0039). Within sub-Saharan Africa, highly variable populations exist both in the east [*e.g.*, Kenya–Malindi (KM), Malawi (Mw), Uganda (Ug), and both Zimbabwe samples] and in the west [*e.g.*, Nigeria, (Ng) Cameroon–Oku].

Most sub-Saharan samples (17 of 21) gave negative values for Tajima's  $D$  (Table 2), indicating an excess of rare alleles compared to the neutral equilibrium expectation (TAJIMA 1989). On the lower end, average  $D$ -values between  $-0.8$  and  $-0.9$  were observed for both eastern (Kenya–Malindi and Uganda) and western (Gabon and Nigeria) populations. Excluding the Congo–Brazzaville (CB) sample (discussed below), a total of 17 of 80 single-locus  $D$ -values from sub-Saharan populations were significantly negative (falling below the 2.5% cutoff of  $D$ -values from neutral simulations). The negative  $D$ -values observed in these populations could reflect demographic histories involving population growth and/or purifying selection at the sequenced loci. Cosmopolitan populations gave

TABLE 2

## Nucleotide-variation summary statistics (average across loci)

Population <sup>a</sup>	<i>n</i>	$\theta_w^b$	$\pi^b$	TajD <sup>b</sup>	Var ( <i>D</i> ) <sup>c</sup>
China	10	0.0036	0.0034	-0.58	3.01
France	10	0.0030	0.0034	0.70	1.43
Israel	10	0.0046	0.0035	-1.01	1.11
Tunisia	9	0.0050	0.0039	-0.64	1.05
Guinea	10	0.0099	0.0089	-0.57	0.25
Niger	10	0.0083	0.0072	-0.55	0.25
Benin	9	0.0099	0.0091	-0.39	0.37
Nigeria	10	0.0130	0.0104	-0.84	0.39
Cam-M	10	0.0090	0.0075	-0.78	0.24
Cam-D	10	0.0107	0.0093	-0.50	0.09
Cam-W	6	0.0106	0.0101	-0.33	0.20
Cam-N	10	0.0106	0.0094	-0.55	0.27
Cam-O	10	0.0115	0.0097	-0.77	0.14
Cam-Y	8	0.0096	0.0084	-0.66	0.08
Gabon	10	0.0095	0.0078	-0.85	0.53
Congo	10	0.0078	0.0086	0.33	2.49
Eritrea	10	0.0086	0.0092	0.33	0.25
Uganda	10	0.0121	0.0102	-0.80	0.03
Rwanda	8	0.0101	0.0103	0.20	0.15
Kenya-N	10	0.0085	0.0086	0.06	0.05
Kenya-M	10	0.0123	0.0101	-0.89	0.15
Malawi	10	0.0114	0.0098	-0.69	0.10
Zim-K	10	0.0122	0.0105	-0.62	0.11
Zim-S	10	0.0112	0.0104	-0.36	0.14
South Africa	10	0.0096	0.0088	-0.42	0.12

<sup>a</sup>Population abbreviations are listed in Table 1. The four cosmopolitan populations are listed first; then the sub-Saharan populations are listed geographically from northwest to southeast.

<sup>b</sup>Watterson's  $\theta$  ( $\theta_w$ , from the number of segregating sites), nucleotide diversity ( $\pi$ ), and Tajima's *D* (TajD) are averaged across the four loci for each population.

<sup>c</sup>Variance of the four locus-specific values of Tajima's *D*.

both positive and negative *D*-values, with a high variance among loci (Table 2; supplemental Table S2 at <http://www.genetics.org/supplemental/>), consistent with expectations for bottlenecked populations (*e.g.*, HADDRILL *et al.* 2005). With only four loci, the information concerning variance in Tajima's *D* is clearly limited, but it was notable that all eastern African populations except Eritrea had  $V[D] < 0.15$ , while 9 of 12 western African populations had  $V[D] > 0.15$ .

The anomalously high variance in *D* observed in the CB sample is most likely a product of nonrandom sampling from this population. This sample did not consist of true isofemale lines and was included only to assess whether this population contained an appreciable quantity of unique variation. It did not, and was excluded from most subsequent analyses. Also excluded from most analyses was the Cameroon-Mbengwi (CW) sample, which was collected within 100 km of both the Cameroon-Oku (CO) and the Cameroon-Nkoundja (CN) samples. The CW sample was included primarily to test for fine-scale genetic structure, but none was observed, and levels of variation were similar (Table 2).

Since CN and CO had larger sample sizes, most analyses were restricted to these samples.

None of the sequenced loci showed evidence for recent positive selection, and in general patterns of variability were consistent among loci (locus-specific summary statistics are presented in supplemental Table S2 at <http://www.genetics.org/supplemental/>). Locus 8A4 showed moderately lower variability (and divergence) overall, but no sub-Saharan population showed dramatically reduced variability at any locus. Multilocus HKA tests (HUDSON *et al.* 1987) performed for each population sample resulted in no rejections of the neutral model (in fact, there were no *P*-values  $< 0.25$ ). Although some single-locus values of Tajima's *D* were observed to reject the null model (supplemental Table S2 at <http://www.genetics.org/supplemental/>), none of the cosmopolitan *D*-values fell outside the 95% confidence interval of empirical *D*-values obtained by OMETTO *et al.* (2005) for 241 variable loci in a European population of *D. melanogaster*. For sub-Saharan populations (aside from the CB sample), all of the moderately negative *D*-values that had fallen outside the neutrally simulated distribution (supplemental Table S2 at <http://www.genetics.org/supplemental/>) fell within the 95% confidence interval of 253 empirical *D*-values from a Zimbabwe population (OMETTO *et al.* 2005).

**Genetic differentiation:** High pairwise  $F_{ST}$  values (generally  $> 0.2$ ) were observed between cosmopolitan and sub-Saharan populations (Table 3). High  $F_{ST}$  values were also observed between China (Ch) and other cosmopolitan samples, which is consistent with recent findings (BAUDRY *et al.* 2004; SCHLÖTTERER *et al.* 2006) of genetic structure among Asian populations of *D. melanogaster*. Comparisons among the other cosmopolitan samples (France, Israel, and Tunisia) gave no evidence for differentiation.

Within sub-Saharan Africa, most pairwise  $F_{ST}$  values were  $< 0.1$ , but many of these were statistically significant (Table 3). In general,  $F_{ST}$  values  $> 0.03$  often had permutation test *P*-values  $< 0.05$  (depending on the sample sizes and the configuration of variation in the populations compared).  $S_{nn}$  *P*-values were significant in about the same number of pairwise comparisons as those for  $F_{ST}$ , with mostly the same comparisons identified by both statistics (supplemental Table S3 at <http://www.genetics.org/supplemental/>).  $F_{ST}$  and  $S_{nn}$  values were included primarily to provide a quantitative estimate of the differentiation observed between any given pair of populations, and the *P*-values listed in Table 3 and supplemental Table S3 (<http://www.genetics.org/supplemental/>) do not reflect corrections for multiple testing. However, given that 174 of 253 pairwise  $F_{ST}$  comparisons had *P*-values  $< 0.05$  (Table 3), while an average of 13 false positives would be expected in 253 tests at the 0.05 significance level, it seems that a large majority of significant results in these tables are likely to indicate genuine population structure.

TABLE 3  
Pairwise  $F_{ST}$  (above diagonal) and  $P$ -values (below diagonal)

Population	Ch	Fr	Is	Tu	Gu	Nr	Bn	Ng	CM	CD	CN	CO	CY	Ga	Er	Ug	Rw	KN	KM	Mw	ZK	ZS	SA
China	—	0.532	0.524	0.480	0.377	0.396	0.375	0.317	0.387	0.299	0.284	0.310	0.303	0.313	0.346	0.282	0.301	0.346	0.318	0.346	0.321	0.342	0.374
France	0.000	—	0.009	0.023	0.277	0.250	0.262	0.226	0.302	0.259	0.265	0.228	0.275	0.275	0.238	0.214	0.301	0.294	0.337	0.319	0.288	0.337	0.086
Israel	0.000	0.331	—	0.031	0.298	0.259	0.260	0.243	0.324	0.256	0.264	0.231	0.264	0.268	0.247	0.210	0.315	0.292	0.349	0.340	0.290	0.342	0.105
Tunisia	0.000	0.246	0.801	—	0.277	0.254	0.251	0.240	0.306	0.243	0.258	0.219	0.254	0.260	0.239	0.195	0.290	0.276	0.315	0.313	0.260	0.321	0.087
Guinea	0.000	0.000	0.000	0.000	—	0.029	0.020	0.005	0.014	0.030	0.026	0.011	0.003	0.083	0.116	0.019	0.048	0.076	0.047	0.044	0.027	0.065	0.085
Niger	0.000	0.000	0.000	0.000	0.089	—	0.016	0.013	0.018	0.015	0.022	0.011	0.021	0.030	0.102	0.010	0.083	0.104	0.114	0.114	0.081	0.114	0.080
Benin	0.000	0.000	0.000	0.000	0.169	0.188	—	0.007	0.025	0.042	0.027	0.015	0.021	0.055	0.113	0.002	0.052	0.059	0.069	0.059	0.030	0.071	0.064
Nigeria	0.000	0.000	0.000	0.000	0.372	0.225	0.314	—	-0.015	-0.001	0.001	-0.005	0.008	0.044	0.077	-0.006	0.042	0.048	0.044	0.031	0.052	0.037	0.066
Cam-M	0.000	0.000	0.000	0.000	0.221	0.161	0.098	0.809	—	-0.006	-0.015	0.008	0.026	0.027	0.132	-0.004	0.053	0.063	0.052	0.041	0.052	0.063	0.086
Cam-D	0.000	0.000	0.000	0.000	0.080	0.196	0.027	0.487	0.578	—	-0.025	0.002	-0.024	-0.004	0.080	-0.017	0.035	0.074	0.055	0.064	0.038	0.071	0.086
Cam-N	0.000	0.000	0.000	0.000	0.095	0.133	0.097	0.428	0.745	0.895	—	0.003	-0.010	-0.006	0.120	-0.021	0.030	0.069	0.058	0.064	0.051	0.067	0.092
Cam-O	0.000	0.000	0.000	0.000	0.271	0.236	0.199	0.598	0.307	0.413	0.396	—	0.003	0.027	0.090	-0.008	0.049	0.063	0.060	0.055	0.040	0.061	0.076
Cam-Y	0.000	0.000	0.000	0.000	0.384	0.167	0.167	0.322	0.110	0.855	0.637	0.387	—	0.013	0.111	-0.009	0.054	0.066	0.057	0.070	0.047	0.068	0.104
Gabon	0.000	0.000	0.000	0.000	0.001	0.061	0.011	0.010	0.088	0.559	0.572	0.100	0.245	—	0.122	0.017	0.091	0.123	0.131	0.139	0.102	0.131	0.116
Eritrea	0.000	0.000	0.000	0.000	0.003	0.002	0.004	0.012	0.001	0.009	0.000	0.003	0.009	0.001	—	0.087	0.135	0.161	0.174	0.165	0.141	0.180	0.134
Uganda	0.000	0.000	0.000	0.001	0.142	0.253	0.420	0.606	0.553	0.828	0.870	0.672	0.639	0.158	0.004	—	0.013	0.037	0.032	0.030	0.020	0.042	0.042
Rwanda	0.000	0.000	0.000	0.000	0.106	0.016	0.114	0.101	0.089	0.141	0.178	0.096	0.121	0.118	0.003	0.266	—	0.075	0.018	0.015	0.022	0.042	0.103
Kenya-N	0.000	0.000	0.000	0.000	0.011	0.001	0.031	0.041	0.031	0.012	0.013	0.013	0.034	0.001	0.000	0.084	0.048	—	0.029	0.040	0.056	0.068	0.101
Kenya-M	0.000	0.000	0.000	0.000	0.014	0.000	0.001	0.011	0.007	0.008	0.010	0.006	0.012	0.000	0.000	0.048	0.213	0.126	—	-0.023	-0.004	-0.007	0.102
Malawi	0.000	0.000	0.000	0.000	0.021	0.000	0.001	0.043	0.015	0.003	0.003	0.006	0.005	0.000	0.000	0.043	0.244	0.065	0.937	—	0.004	-0.012	0.104
Zim-K	0.000	0.000	0.000	0.000	0.094	0.001	0.065	0.004	0.005	0.037	0.009	0.024	0.013	0.000	0.000	0.136	0.204	0.023	0.549	0.381	—	0.009	0.063
Zim-S	0.000	0.000	0.000	0.000	0.004	0.000	0.002	0.001	0.003	0.000	0.003	0.009	0.005	0.000	0.000	0.019	0.104	0.007	0.637	0.751	0.291	—	0.097
South Africa	0.000	0.009	0.002	0.012	0.001	0.016	0.006	0.003	0.002	0.001	0.002	0.003	0.002	0.000	0.001	0.045	0.010	0.005	0.001	0.000	0.010	0.002	—

Population abbreviations are listed in Table 1.

Some sub-Saharan populations, particularly Eritrea (Er) and South Africa (SA), showed a consistent pattern of elevated differentiation when compared to other sub-Saharan samples (Table 3; supplemental Table S3 at <http://www.genetics.org/supplemental/>). An examination of locus-specific  $F_{ST}$  matrices and neighbor-joining population distance trees revealed particularly close relationships between each of these populations and cosmopolitan samples at certain loci (4F2 and 12F1 for Eritrea; 4F2, 8A4, and 11A5 for South Africa; locus-specific  $F_{ST}$  matrices are given in supplemental Table S4 at <http://www.genetics.org/supplemental/>), suggesting the possibility of a cosmopolitan admixture. More moderate elevations in  $F_{ST}$  were also observed for the Gabon (Ga), Kenya–Naivasha (KN), Niger (Nr), and Rwanda (Rw) populations. Gabon also showed potential evidence of cosmopolitan admixture, having a very close relationship to the cosmopolitan populations at locus 12F1 only (supplemental Table S4 at <http://www.genetics.org/supplemental/>). While the heterogeneous pattern of locus-specific cosmopolitan introgression that we infer seems contrary to the expectations of neutral gene flow, it is very similar to the pattern observed by KAUER *et al.* (2003) in *D. melanogaster* from Harare, Zimbabwe, and could indicate selection against immigrant genotypes at specific loci. While it can be difficult to decisively differentiate the hypotheses of cosmopolitan admixture *vs.* cosmopolitan co-ancestry, the high levels of linkage disequilibrium (low  $\rho/\pi$  estimates) inferred for these three populations (Table 4) are consistent with a history of recent admixture. The KN, Nr, and Rw samples showed no obvious evidence of large-scale cosmopolitan admixture at any locus, and their differentiation might be attributable to some combination of local population size reductions and restricted gene flow.

In contrast to the more differentiated sub-Saharan populations discussed above, Uganda showed particularly close relationships to many other sub-Saharan samples (Table 3; supplemental Table S2 and Figure S1 at <http://www.genetics.org/supplemental/>). While Uganda showed modest differentiation from most other eastern African populations, it showed essentially no differentiation from any of the Cameroon and Nigeria samples (despite a geographic separation of ~2000–2500 km) and fairly close relationships to the other western African samples as well. A second group of sub-Saharan populations with no detectable differentiation was found within eastern Africa, composed of the Kenya–Malindi, Malawi, Zimbabwe–Kariba (ZK) and Zimbabwe–Sengwa (ZS) samples (which span a range of ~2100 km).

When 17 sub-Saharan populations were analyzed (all except CB, CW, Er, and SA), a significant relationship was observed between genetic distance (measured as  $F_{ST}$ ) and geographic distance ( $P = 0.00023$ ; squared correlation coefficient = 0.26, Mantel test). However,

TABLE 4

Summary statistics indicating stable demographic history

Population <sup>a</sup>	$\pi$ (rank) <sup>b</sup>	Max. TajD (rank) <sup>c</sup>	$\rho/\pi$ (rank) <sup>b</sup>
Uganda	0.0102 (5)	−0.61 (1)	13.0 (5)
Kenya–M	0.0101 (6)	−0.40 (3)	15.1 (4)
Zim–K	0.0105 (1)	−0.25 (7)	10.0 (10)
Zim–S	0.0104 (2)	0.06 (14)	15.4 (3)
Malawi	0.0098 (7)	−0.33 (4)	10.2 (8)
Cam–O	0.0097 (8)	−0.40 (2)	9.8 (11)
Cam–D	0.0093 (10)	−0.12 (8)	12.1 (7)
Cam–M	0.0075 (18)	−0.27 (6)	17.2 (2)
Nigeria	0.0104 (3)	0.03 (12)	5.8 (13)
Benin	0.0091 (12)	0.42 (17)	17.5 (1)
Cam–Y	0.0084 (16)	−0.29 (5)	10.0 (9)
Cam–N	0.0094 (9)	0.00 (10)	4.6 (14)
Niger	0.0072 (19)	−0.01 (9)	12.7 (6)
Guinea	0.0089 (13)	0.05 (13)	8.6 (12)
S.Africa	0.0088 (14)	0.00 (11)	4.2 (15)
Rwanda	0.0103 (4)	0.71 (18)	0.9 (19)
Eritrea	0.0092 (11)	0.91 (19)	2.7 (17)
Gabon	0.0078 (17)	0.12 (15)	3.1 (16)
Kenya–N	0.0086 (15)	0.23 (16)	1.3 (18)

Population abbreviations are listed in Table 1.

<sup>a</sup>Populations are ordered according to the sum of their ranks for the three statistics presented.

<sup>b</sup>Populations are ranked from highest to lowest values of  $\pi$  (averaged across loci) and  $\rho/\pi$  (composite likelihood for all four loci).

<sup>c</sup>“Max. TajD” refers to the highest value of Tajima’s *D* observed at any locus for the given population. These values are ranked from lowest to highest.

the exceptions to this relationship mentioned above (closely related populations separated by large geographic distances) suggested that a simple isolation by distance model might not be a complete explanation for patterns of genetic differentiation in sub-Saharan *D. melanogaster*. A neighbor-joining population tree based on  $F_{ST}$  values also suggested a geographic pattern for differentiation among the sub-Saharan samples (Figure 2). Specifically, the eastern populations (KM, KN, Mw, Rw, ZK, and ZS) grouped separately from the western populations, with Uganda occupying an intermediate position. No relationship between genetic distance and geographic distance was detected within the eastern group of 6 populations ( $P = 0.40$ , Mantel test) or within the western group of 11 populations (including Uganda;  $P = 0.48$ ). Thus, the relationship between genetic and geographic distances detected for 17 sub-Saharan populations appears to have been driven primarily by differentiation between eastern and western populations. Correspondingly, we note that comparisons between eastern and western sub-Saharan populations showed greater differentiation (mean  $F_{ST} = 0.062$  for 66 pairwise comparisons) than comparisons among eastern populations (mean  $F_{ST} = 0.025$  for 15 pairwise comparisons) or among western populations (mean  $F_{ST} = 0.011$  for 55 pairwise comparisons).

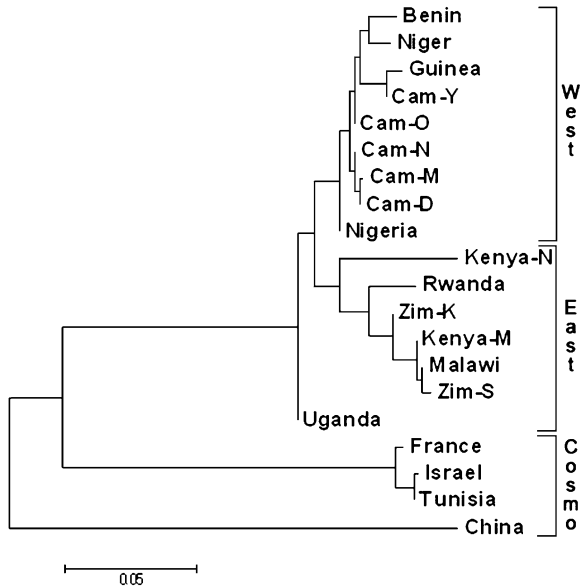


FIGURE 2.—Unrooted population distance tree generated by neighbor-joining analysis of  $F_{ST}$  values. Distance is indicated by the lengths of horizontal branches. Putatively admixed populations (Er, Ga, SA) are excluded.

The Mantel test was also applied to test for a correlation between genetic distance and collection date in sub-Saharan *D. melanogaster*, such as that reported by DIERINGER *et al.* (2005). No such correlation was found ( $P = 0.11$  for a simple Mantel test;  $P = 0.34$  after accounting for the correlation between genetic differentiation and geographic distance). Most of the samples analyzed in this study were collected within the past few years (Table 1), and this might have prevented such a correlation from being found. However, the close relationships observed between the newer Kenya-Malindi and Malawi samples and both of the older Zimbabwe samples would argue against any recent and rapid genetic change at the loci examined.

Structure analysis was applied first to the entire data set. For  $K = 2$  subpopulations, the cosmopolitan populations were placed largely in the cluster shaded in Figure 3A. Most sub-Saharan populations had ancestry

in both clusters: western populations showed approximately equal ancestry in both clusters, while most eastern populations had a majority of ancestry in the unshaded clusters (Figure 3A). When the number of subpopulations was set to  $K = 3$ , an increase in the log likelihood of the data from  $-20,188$  to  $-18,894$  was observed. Although cosmopolitan samples were placed primarily in one cluster, the geographic pattern involving the other two clusters was not clear (supplemental Figure S2 at <http://www.genetics.org/supplemental/>).

To clarify the pattern of genetic differentiation within sub-Saharan Africa, Structure was run using 17 sub-Saharan samples (excluding CB, CW, Er, and SA). With  $K = 2$  subpopulations, eastern samples tended to have similar levels of ancestry in both clusters, while western samples had a larger portion of ancestry in the unshaded cluster in Figure 3B. A significant correlation was observed between longitude and cluster membership ( $R^2 = 0.53$ ;  $P = 0.0009$ ; supplemental Figure S3 at <http://www.genetics.org/supplemental/>), reinforcing the geographic basis of this differentiation.

Although runs with up to 10 subpopulations were performed in both cases, Structure gave little support for differentiation beyond  $K = 3$  for the full data set and  $K = 2$  for the sub-Saharan data set. While log-likelihood scores generally continued to rise for larger values of  $K$ , every run with  $K$ -values greater than those specified above produced one or more nearly empty clusters (having  $<0.1$  of each population's ancestry), and further geographic patterns were not evident. Thus, there seemed to be no justification for inferring additional genetic clusters on the basis of the Structure results.

AMOVA (EXCOFFIER *et al.* 1992) was performed using cosmopolitan (4 populations) and sub-Saharan (17 populations) groupings. These groupings were found to account for 9.9% of variation, with another 9.6% of variation found among populations within these groups, and 80.5% of variation occurring within populations. A second AMOVA was performed on the sub-Saharan populations only, using eastern (6 populations) and western (11 populations) groupings. In this case, 8.6% of variation in sub-Saharan Africa was

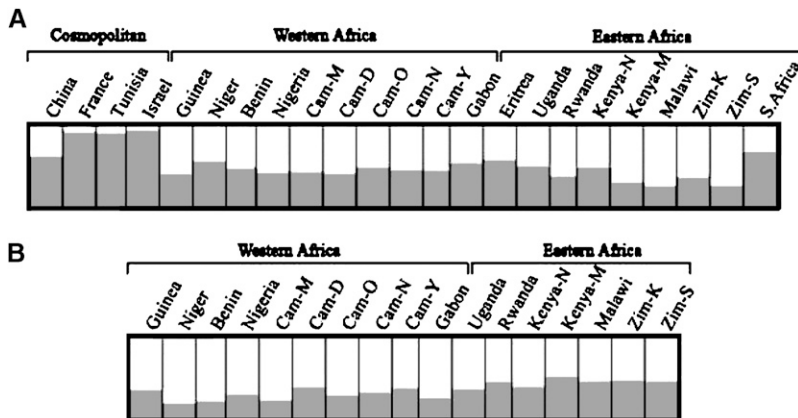


FIGURE 3.—Population membership in the genetic clusters identified by Structure. In each case, one of the two clusters is arbitrarily shown as shaded, the other as open. (A) Population ancestry for the  $K = 2$  Structure analysis of cosmopolitan and sub-Saharan samples. (B) Population ancestry for the  $K = 2$  Structure analysis of non-admixed sub-Saharan samples only.



between the eastern and western groups, compared to just 1.9% of variation between populations within those groups, while 89.5% of variation was found within populations.

**Relative antiquity of eastern and western sub-Saharan populations:** The ancestral range of *D. melanogaster* is thought to be within sub-Saharan Africa (LACHAISE *et al.* 1988). However, we do not know whether *D. melanogaster* has occupied a broad geographic range within sub-Saharan Africa for a long period of time, or if this species might have recently expanded within sub-Saharan Africa from a smaller ancestral range. Therefore, we consider the patterns of polymorphism expected for ancient *vs.* recently founded populations. First, we expect recently established populations to be less variable than ancient populations, due to the loss of variation during a founder event. Second, we expect the population bottlenecks associated with founder events to alter the allele-frequency spectra of recently established populations (*e.g.*, WALL *et al.* 2002; DEPAULIS *et al.* 2003; HADDRILL *et al.* 2005). Specifically, recently founded populations may have higher average Tajima's *D*-values (due to the loss of rare alleles during a moderate bottleneck). Also, we expect recently founded populations to have a higher variance among loci in Tajima's *D*, due to the stochastic effects of a founder-event-associated bottleneck. An exception to these criteria for Tajima's *D* would be if a recently founded population had suffered a very extreme bottleneck (such that nearly all variation was lost at most loci), since this might also produce consistently negative Tajima's *D*-values. However, all of our sub-Saharan samples are sufficiently variable that a severe bottleneck does not appear to be a plausible demographic history for any of them. Finally, we expect recently founded populations to have increased linkage disequilibrium (*e.g.*, REICH *et al.* 2001; WALL *et al.* 2002; THORNTON and ANDOLFATTO 2006), because not every allelic combination present in the ancestral population will be transferred to the new population.

These expectations were quantified using three summary statistics: nucleotide diversity ( $\pi$ ), maximum Tajima's *D* (the highest *D*-value observed at any locus for a given population, thus incorporating the expected increase both in *D* and its variance for recently founded populations), and  $\rho/\pi$  (the population recombination rate, estimated as described in MATERIALS AND METHODS and divided by  $\pi$  to account for the effect of variability on its estimation). Thus, we expect ancient populations with comparatively stable demographic histories to have high  $\pi$ , low maximum *D*, and high  $\rho/\pi$ .

The pattern produced by any one of these three statistics would not have yielded a clear geographic hypothesis for the ancestral range of *D. melanogaster*. However, when sub-Saharan populations were ordered according to the sum of their ranks for the three statistics, a geographic pattern was evident (Table 4). The

**TABLE 5**  
Differentiation of sub-Saharan populations from cosmopolitan populations

Population <sup>a</sup>	$F_{ST}$ (rank) <sup>b</sup>	$S_{nn}$ (rank) <sup>b</sup>	Shared polymorphism (rank) <sup>c</sup>
Uganda	0.212 (1)	0.712 (1)	43 (1)
Cam-N	0.235 (3)	0.769 (3)	37 (5)
Nigeria	0.233 (2)	0.764 (2)	36 (7)
Cam-O	0.238 (6)	0.789 (5)	37 (5)
Cam-D	0.237 (5)	0.777 (4)	35 (8)
Zim-K	0.276 (9)	0.799 (7)	41 (2)
Guinea	0.302 (12)	0.811 (10)	39 (3)
Kenya-M	0.307 (14)	0.806 (8)	38 (4)
Malawi	0.306 (13)	0.799 (6)	34 (9)
Niger	0.269 (7)	0.810 (9)	29 (14)
Kenya-N	0.279 (10)	0.816 (12)	34 (9)
Cam-Y	0.236 (4)	0.822 (13)	29 (14)
Benin	0.274 (8)	0.832 (14)	33 (11)
Cam-M	0.312 (15)	0.812 (11)	32 (13)
Rwanda	0.288 (11)	0.865 (16)	29 (14)
Zim-S	0.318 (16)	0.863 (15)	33 (11)
Eritrea <sup>d</sup>	0.231	0.702	39
Gabon <sup>d</sup>	0.231	0.744	35
South Africa <sup>d</sup>	0.144	0.668	42

Population abbreviations are listed in Table 1.

<sup>a</sup>Populations are ordered according to the sum of their ranks for the three statistics presented.

<sup>b</sup> $F_{ST}$  and  $S_{nn}$  values are arithmetic means of the four pairwise values from comparisons of the given sub-Saharan population with each of the four cosmopolitan populations. The 17 sub-Saharan populations are ranked starting from the lowest value (closest relationship).

<sup>c</sup>Number of polymorphisms shared between the given sub-Saharan population and one or more of the cosmopolitan populations (total for all four loci).

<sup>d</sup>Not ranked due to apparent cosmopolitan admixture.

five populations that showed the best evidence for ancient and stable histories (Ug, KM, ZK, ZS, and Mw) were all from eastern Africa. Thus, in spite of evolutionary and sampling variance affecting the estimation of these statistics, a hypothesis for the ancestral range of this species could be generated.

**Source of Palearctic *D. melanogaster*:** Relationships between sub-Saharan and cosmopolitan populations were investigated with the goal of inferring which sub-Saharan populations might resemble the source (pre-bottleneck) population that gave rise to Palearctic *D. melanogaster*. As mentioned above, three sub-Saharan populations (Er, Ga, and SA) displayed evidence of recent cosmopolitan admixture. Aside from these samples, 16 sub-Saharan populations were ranked according to three statistics illustrating differentiation from the Palearctic populations (Table 5).  $F_{ST}$ ,  $S_{nn}$ , and the number of shared polymorphisms all revealed Uganda as the sub-Saharan population most closely related to the Palearctic samples.

Relationships between sub-Saharan and Palearctic populations were also depicted using neighbor-joining population distance trees on the basis of  $F_{ST}$ . When the nonadmixed sub-Saharan populations were analyzed along with the Palearctic samples, the sample branching off closest to the Palearctic samples was Uganda (Figure 2). Aside from Er and SA, Ug and KN were also the two sub-Saharan populations with the highest level of ancestry in the predominantly cosmopolitan cluster from the  $K = 3$  run of Structure.

**Excess of singleton polymorphisms and evaluation of potential causes:** For a neutral equilibrium population evolving according to the infinite-sites model, theory predicts that the number of singleton polymorphisms observed in a sample should be  $\theta$ , regardless of the sample size (*e.g.*, Fu 1995). For our full data set, the expected number of singletons (mutations present in 1 of 240 individuals) would be 83.6 if  $\theta$  is estimated from the number of segregating sites ( $S = 506$ ). In contrast to this prediction, we observed 279 singleton sites (55.1% of all segregating sites observed were singletons, compared to 16.5% expected on the basis of  $\theta_s$ ). Rechecking of chromatograms suggested that this singleton excess was not a result of base-calling errors. Many singletons were verified by high-quality sequence on both strands, and we do not expect to observe PCR misincorporations in sequence amplified from a large pool of genomic DNA, so we conclude that this singleton excess represents a real characteristic of polymorphism in *D. melanogaster*.

The  $D$  statistic of Fu and Li (1993) specifically tests for deviations from the expected number of singleton polymorphisms (external mutations). For our full data set, Fu and Li's  $D$  was  $-7.69$  (indicating an excess of singletons), and the null model was rejected ( $P < 0.001$  using simulations with recombination). The observed deviation could indicate a departure from selective neutrality or from demographic equilibrium. We investigated the ability of three potential causes to generate the observed singleton excess: population structure, population growth, and negative selection (constraint).

Population structure was observed in our data, and it is possible that such structure might have helped to generate the observed singleton excess. To account for this possibility, we assessed the proportion of singleton polymorphisms in groups of populations that contain no detectable structure. Two groups of four populations each were tested: one from eastern Africa (the "East4" sample: KM, Mw, ZK, ZS) and one from western Africa [the "West4" sample: Cameroon–Mbalang–Djalango (CD), CN, CO, Ng]. In these two groupings of 40 individuals each, the proportions of singleton polymorphisms were 0.527 for the East4 sample (119 singletons of 226 segregating sites) and 0.502 for the West4 sample (109 singletons of 217 segregating sites), compared to an expectation of 0.235 based on  $\theta_s$ . These results were compared to random subsamples of the

TABLE 6

Simulations showing the effect of population growth on the proportion of singleton polymorphisms observed in a sample of 240

Time of growth <sup>a</sup>	N ratio <sup>b</sup>	Proportion of singletons <sup>c</sup>	$P$ (sim > obs) <sup>d</sup>
—	—	0.1661	0
0.1333	0.1	0.2575	0
0.1333	0.01	0.2726	0
0.1333	0.001	0.2745	0
0.1333	0.0001	0.2748	0
0.1333	0.00001	0.2748	0
0.1333	0.000001	0.2748	0
0.01333	0.1	0.4147	0
0.01333	0.01	0.5159	0.2315
0.01333	0.001	0.5289	0.4084
0.01333	0.0001	0.5301	0.4204
0.01333	0.00001	0.5306	0.4318
0.01333	0.000001	0.5304	0.4327
0.001333	0.1	0.3750	0
0.001333	0.01	0.7466	1
0.001333	0.001	0.8581	1
0.001333	0.0001	0.8721	1
0.001333	0.00001	0.8732	1
0.001333	0.000001	0.8734	1
0.000133	0.1	0.2045	0
0.000133	0.01	0.4496	0.0098
0.000133	0.001	0.8529	1
0.000133	0.0001	0.9692	1
0.000133	0.00001	0.9829	1
0.000133	0.000001	0.9843	1

<sup>a</sup>The time, in coalescent units before the present, when an instantaneous population-size increase occurred.

<sup>b</sup>The ratio of the preincrease population size to the post-increase population size.

<sup>c</sup>The average proportion of singletons observed across 10,000 simulated replicates.

<sup>d</sup>The proportion of simulated replicates that contained a greater proportion of singletons than the observed data.

data in which 40 of 240 individuals were chosen without respect to their geographic origin (thus sampling across known population structure). The 10,000 random resamplings had a mean proportion of singletons of 0.532. Both the East4 and West4 values were well within the 95% confidence interval of resampled values: 42.0% of resampled values were below the East4 grouping's proportion of singletons, while 18.3% fell below the West4 value. Since no significant difference could be detected between structured and unstructured subsamples, we conclude that population structure has little effect on the proportion of singletons in our data.

To assess the potential of population growth to account for the observed singleton excess, we used the coalescent simulator ms (HUDSON 2002) to generate data sets resembling our own under a model of instantaneous growth (see MATERIALS AND METHODS). As shown in Table 6, a variety of population growth parameter combinations give a proportion of singletons

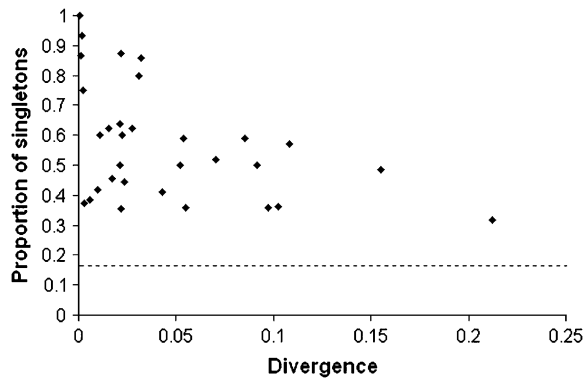


FIGURE 4.—The relationship between interspecific divergence and proportion of singleton polymorphisms for 100-bp sections of the loci studied. The dashed line indicates the proportion of singletons expected in a neutral equilibrium population.

that is similar to or much greater than that observed in our data. For example, growth occurring 0.01333 coalescent units ago ( $\sim 10,000$  years ago) produces slightly  $>50\%$  singletons for a wide range of growth rates (between 100- and 1,000,000-fold growth). More recent growth tends to produce greater singleton excess, with  $>98\%$  singletons in the most extreme cases. Thus, while the true history of any *D. melanogaster* population is likely to be more complex than the growth models simulated (involving, for example, population structure and population bottlenecks), a demographic history that involves recent growth appears to have considerable power to generate a singleton excess like the one observed.

Although population growth could be a sufficient explanation for the observed singleton excess, we also tested for evidence of negative selection. The four sequenced fragments are from long intergenic regions, but previous studies have reported evidence for constraint on noncoding DNA in *D. melanogaster* (BERGMAN and KREITMAN 2001; ANDOLFATTO 2005; KERN and BEGUN 2005), and the high rate of recombination in these regions could allow purifying selection to act efficiently. Deleterious mutations may thus be kept at low frequency and be observed as singletons in our sample. If this has occurred, one might expect regions of our sequenced loci that have a stronger singleton excess to also have reduced divergence between species (both patterns being generated by selective constraint). Upon breaking our loci into 100-bp windows, we found evidence for an inverse correlation between divergence (using *D. simulans* as an outgroup) and proportion of singleton polymorphisms (Figure 4;  $R^2 = 0.19$ ;  $P = 0.015$ ), thus providing evidence that purifying selection is at least partially responsible for the singleton excess in our data. However, the proportion of singletons shown in Figure 4 seems to “bottom out” between 0.3 and 0.4, which is still well above the expected value of 0.165. The window containing the lowest proportion of singletons

(0.316) also had the highest divergence (0.212), which is suggestive of a rather low level of selective constraint. Since even this window contains approximately twice as many singletons as expected on the basis of  $\theta$ , it seems that purifying selection alone may not be a sufficient explanation for the observed singleton excess. Instead, the most plausible explanation might be a combination of population growth and negative selection.

## DISCUSSION

An analysis of DNA sequence variation at four 1-kb X-linked fragments was performed for 25 population samples of *D. melanogaster* to explore genetic structure and historical relationships involving sub-Saharan populations. The primary structure detected within sub-Saharan Africa differentiated eastern from western populations (with Uganda having a closer relationship to the western samples). Evidence for east–west structure included pairwise  $F_{ST}$  values, Mantel test results indicating a significant correlation between genetic and geographic distance within all of sub-Saharan Africa but not within the eastern and western population groups, neighbor-joining population distance trees depicting the separation between eastern and western populations, Structure analysis of sub-Saharan populations showing a correlation between longitude and the cluster membership of populations, and AMOVA analysis indicating that considerably more variation exists between eastern and western population groupings than among populations within those groups. These findings are consistent with previous studies of inversion frequencies (LEMEUNIER and AULARD 1992; AULARD *et al.* 2002) and studies of molecular variation that had found significant differentiation between one (or two) eastern and one western African population sample (*e.g.*, BÉNASSI and VEUILLE 1995; MICHALAKIS and VEUILLE 1996; HADDRILL *et al.* 2005).

Additional differentiation within sub-Saharan Africa was observed for populations that showed close relationships to cosmopolitan populations at one or more loci. Specifically, low locus-specific  $F_{ST}$  values in cosmopolitan comparisons and high levels of linkage disequilibrium provided evidence for cosmopolitan admixture affecting South Africa at three loci, Eritrea at two loci, and Gabon at one locus. Thus, cosmopolitan admixture was inferred for a minority of sub-Saharan samples, but the disparate locations of these three samples, along with previous reports of cosmopolitan admixture in Brazzaville, Congo (VOUIDIBIO *et al.* 1989; CAPY *et al.* 2000), and Harare, Zimbabwe (KAUER *et al.* 2003), suggest that cosmopolitan genotypes have been introduced at multiple sub-Saharan sites.

Apart from cases of likely cosmopolitan admixture, we were interested in assessing which sub-Saharan population showed the closest relationship with the

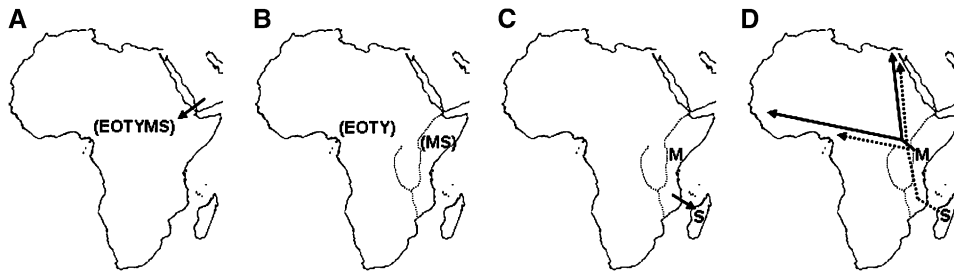


FIGURE 5.—Hypothesized history of *D. melanogaster* and related species. (A) EOTYMS refers to the common ancestor of *D. erecta*, *D. orena*, *D. teissieri*, *D. yakuba*, *D. melanogaster*, and *D. simulans*, (along with the three known island endemic species), which reached Africa <20 MYA. (B) The split between the ancestor of *D. melanogaster* and *D. simulans*

(MS) and the ancestor of *D. erecta*, *D. orena*, *D. teissieri*, and *D. yakuba* (EOTY) occurred ~10–15 MYA, possibly associated with emergence of the East African Rift. (C) The divergence of *D. melanogaster* and *D. simulans* 2–6 MYA may have been initiated by the dispersal of the latter species to Madagascar or other Indian Ocean islands. (D) More recently, *D. melanogaster* expanded from eastern Africa into western Africa and Eurasia, and *D. simulans* expanded into continental Africa and Eurasia.

Palaearctic populations. This population could represent the best contemporary proxy for the source population that gave rise to the original Palaearctic population of *D. melanogaster*, and its location might suggest the approximate geographic route by which *D. melanogaster* first left sub-Saharan Africa. On the basis of  $F_{ST}$ ,  $S_{nn}$ , shared polymorphisms, population distance trees, and Structure cluster membership, it was clear that Uganda had the closest genetic affinity with the Palaearctic samples. While such a result could be attributable to cosmopolitan admixture in Uganda, the high variation and low linkage disequilibrium observed in this sample argue against such an explanation.

LACHAISE *et al.* (1988) suggested that *D. melanogaster* might have expanded northward from western Africa between 6000 and 9500 years ago, crossing a Saharan region that was less arid at the time. Although we found that western African populations were very closely related to our Uganda sample, it was Uganda that showed the closest relationship to the Palaearctic populations. If we are correct in concluding that the source population that gave rise to Palaearctic *D. melanogaster* originated from the equatorial rift zone, it may not be necessary to invoke a crossing of the central Sahara. Instead, the northward expansion of *D. melanogaster* might have occurred via the Nile Valley and/or the Red Sea coastal area.

By analyzing some contrasting predictions for polymorphism in populations with relatively stable demographic histories *vs.* those that may have experienced recent founder events, we identified five populations from eastern Africa that showed the best evidence for ancient and stable histories. Specifically, Uganda, Kenya–Malindi, Zimbabwe–Kariba, Zimbabwe–Sengwa, and Malawi were found to have, on average, higher variation, lower and less variable Tajima’s *D*-values, and less linkage disequilibrium than the other sub-Saharan populations studied (Table 4). These results could suggest that the ancestral range of *D. melanogaster* was within eastern Africa (either east of the rift or within the rift zone), which is consistent with the hypotheses of VEUILLE *et al.* (2004) and HADDRILL *et al.* (2005). However, the geographic breadth of these five popula-

tions, along with the existence of unsampled regions within sub-Saharan Africa, prevents us from drawing precise conclusions regarding the geographic origin of *D. melanogaster*. Importantly, our data do not rule out an ancestral range that includes the western rift mountains, which could be compatible with an origin in the “East Central refugia” of tropical forest, a possibility suggested by LACHAISE *et al.* (1988) and LACHAISE and SILVAIN (2004). Alternatively, given that *D. melanogaster* has greater resistance to desiccation (VAN HERREWEGE and DAVID 1997) and to high and low temperatures (STANLEY *et al.* 1980) than any of its close relatives, this species may have evolved in a relatively less humid region of eastern Africa. It is also possible that the geographic range of this species has changed through time in response to climate change and other factors.

It is interesting that although the Uganda population shows its closest relationships to be with western African populations, it was identified along with four other eastern African populations in our analysis of putatively ancient population samples. If the hypothesis of an eastern African origin for *D. melanogaster* is correct, the above results might suggest that western populations were founded comparatively recently and were derived from a population very similar to our Uganda sample. Clearly, this proposed westward expansion from the equatorial rift zone did not result in a substantial loss of variability for most western populations. Therefore, this expansion may have involved a large number of founders and/or a high rate of subsequent migration between founder and source populations.

In light of the above discussion, our results suggest the following as a reasonable hypothesis for the history of *D. melanogaster* and related species. While most other subgroups within the *D. melanogaster* species group are endemic to East Asia, the nine described species of the *D. melanogaster* subgroup are native to sub-Saharan Africa and nearby islands (LACHAISE *et al.* 1988, 2000). Therefore, it is thought that the ancestor of the *melanogaster* subgroup reached Africa from Eurasia via the Arabian landmass sometime after the first land connection was established, <20 million years ago (MYA) (Figure 5A).

Regarding the earliest speciation event within the *D. melanogaster* subgroup, earlier studies (e.g., LACHAISE *et al.* 1988; Russo *et al.* 1995) suggested that this branching separated the ancestor of *D. erecta* and *D. oreana* from the remaining members of this clade (the *D. melanogaster*/*D. simulans* and the *D. teissieri*/*D. yakuba* species pairs thus forming a clade). However, recent DNA studies have supported a different topology in which the *D. erecta*/*D. oreana* and the *D. teissieri*/*D. yakuba* species pairs form a clade, while the *D. melanogaster*/*D. simulans* lineage would have diverged from these taxa earlier (KOPP and TRUE 2002; KASTANIS *et al.* 2003; KO *et al.* 2003). The estimated timing of this speciation event, ~10–15 MYA (TAMURA *et al.* 2004), suggests the possibility that the emergence of the East African Rift (GIRDLER 1991) might have played a role in the isolation of these taxa (Figure 5B).

Restricting our focus to the *D. melanogaster*/*D. simulans* lineage, aridification of the rift zone was suggested by LACHAISE *et al.* (1988) as a potential vicariant event separating populations that would become *D. melanogaster* (west of the rift) from those that would become *D. simulans* (east of the rift or on Indian Ocean islands). Recent studies of DNA sequence polymorphism in *D. simulans* have suggested that populations from Madagascar are more genetically diverse than populations from continental Africa (DEAN and BALLARD 2004; VEUILLE *et al.* 2004; BAUDRY *et al.* 2006). If, as these data imply, *D. simulans* originated on Madagascar or other Indian Ocean islands, there would be no need to invoke the rift valley as a barrier separating this species from *D. melanogaster*. Rather, the simplest explanation for the separation of these two species would be that the ancestor of *D. simulans* dispersed to Madagascar (or other islands), while populations that remained on the mainland became *D. melanogaster* (Figure 5C). Thus, it is not necessary to place *D. melanogaster* west of the rift to explain its divergence from *D. simulans*, and our analysis supports an ancestral range for *D. melanogaster* in eastern Africa.

More recently, we infer that *D. melanogaster* expanded in two directions from its ancestral range in eastern Africa, perhaps spreading from a point of origin near our Uganda sample. Expanding westward, *D. melanogaster* extended its range into central and western Africa (Figure 5D). The newly founded sub-Saharan populations incurred little loss of variation, and only mildly altered allele frequencies and linkage disequilibrium. Also, perhaps associated with this same wave of expansion, *D. melanogaster* moved north, eventually reaching northern Africa and Eurasia (Figure 5D). This expansion gave rise to the first Palearctic populations and was associated with a more severe loss of variation, as documented by several multilocus studies (e.g., KAUER *et al.* 2002; OMETTO *et al.* 2005; THORNTON and ANDOLFATTO 2006). The inferred geographic expansion of *D. melanogaster*, north and west from the equatorial rift zone,

might have been facilitated by the domestication of this species, which could be compatible with the hypothesis of a relatively recent domestication event in the Nilo-Saharan region (LACHAISE and SILVAIN 2004). *D. simulans* also expanded its range, founding mainland populations in eastern and central Africa and also spreading beyond sub-Saharan Africa (Figure 5D).

The results presented here should provide an improved framework for the testing of specific demographic hypotheses in this important model species. For a variety of evolutionary analyses, including studies of adaptive differences among populations, this demographic knowledge should help researchers to select the most appropriate sub-Saharan populations to test hypotheses of interest. For example, we have shown that our Uganda sample may be a better proxy for the Palearctic source population than Zimbabwe (which has commonly been used for this purpose), so this sample should be particularly useful in comparisons between cosmopolitan and sub-Saharan populations. Increased understanding of demography in sub-Saharan *D. melanogaster* should also aid in the interpretation of data from evolutionary studies by allowing researchers to place their results in a stronger historical context.

We are especially grateful to those individuals who collected flies on our behalf (listed in Table 1) and those who facilitated J.E.P.'s travel and collection in Cameroon (particularly the Wildlife Conservation Society). This work was supported by a National Science Foundation Doctoral Dissertation Improvement Grant (DEB-0411730) to J.E.P. and C.F.A. and by a National Institutes of Health grant (GM36431) to C.F.A. Additional support related to fly collection was provided by the Cornell Graduate School and Sigma Xi of Cornell.

#### LITERATURE CITED

- ANDOLFATTO, P., 2005 Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**: 1149–1152.
- AULARD, S., J. R. DAVID and F. LEMEUNIER, 2002 Chromosomal inversion polymorphism in Afrotropical populations of *Drosophila melanogaster*. *Genet. Res.* **79**: 49–63.
- BAILEY, T. L., and M. GRIBSKOV, 1998 Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**: 48–54.
- BAUDRY, E., B. VIGINIER and M. VEUILLE, 2004 Non-African populations of *Drosophila melanogaster* have a unique origin. *Mol. Biol. Evol.* **21**: 1482–1491.
- BAUDRY, E., N. DEROME, M. HUET and M. VEUILLE, 2006 Contrasted polymorphism patterns in a large sample of populations from the evolutionary genetics model *Drosophila simulans*. *Genetics* **173**: 759–767.
- BAUER DUMONT, V., and C. F. AQUADRO, 2005 Multiple signatures of positive selection downstream of *Notch* on the X chromosome in *Drosophila melanogaster*. *Genetics* **171**: 639–653.
- BEGUN, D. J., 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 *vermillion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. *Genetics* **140**: 1019–1032.
- BÉNASSI, V., and M. VEUILLE, 1995 Comparative population structuring of molecular and allozyme variation of *Drosophila melanogaster* *Adh* between Europe, West Africa, and East Africa. *Genet. Res.* **65**: 95–103.
- BERGMAN, C. M., and M. KREITMAN, 2001 Analysis of conserved non-coding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res.* **11**: 1335–1345.

- CAPY, P., M. VEUILLE, M. PAILLETTE, J.-M. JALLON, J. VOUIDIBIO *et al.*, 2000 Sexual isolation of genetically differentiated sympatric populations of *Drosophila melanogaster* in Brazzaville, Congo: The first step towards speciation? *Heredity* **84**: 468–475.
- CARACRISTI, G., and C. SCHLÖTTERER, 2003 Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. *Mol. Biol. Evol.* **20**: 792–799.
- DAVID, J. R., and P. CAPY, 1988 Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet.* **4**: 106–111.
- DEAN, M. D., and J. W. O. BALLARD, 2004 Linking phylogenetics with population genetics to reconstruct the geographic origin of a species. *Mol. Phylogenet. Evol.* **32**: 998–1009.
- DEPAULIS, F., S. MOUSSET and M. VEUILLE, 2003 Power of neutrality tests to detect bottlenecks and hitchhiking. *J. Mol. Evol.* **57**: S190–S200.
- DIERINGER, D., V. NOLTE and C. SCHLÖTTERER, 2005 Population structure in African *Drosophila melanogaster* revealed by microsatellite analysis. *Mol. Ecol.* **14**: 563–573.
- 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.
- EXCOFFIER, L., G. LAVAL and S. SCHNEIDER, 2005 Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinformatics Online* **1**: 47–50.
- FALUSH, D., M. STEPHENS and J. K. PRITCHARD, 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.
- FAY, J. C., and C. I. WU, 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.
- FU, Y.-X., 1995 Statistical properties of segregating sites. *Theor. Popul. Biol.* **48**: 172–197.
- FU, Y.-X., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GIRDLER, R. W., 1991 The Afro-Arabian rift system—an overview. *Tectonophysics* **197**: 139–153.
- HADRILL, P. R., K. R. THORNTON, B. CHARLESWORTH and P. ANDOLFATTO, 2005 Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res.* **15**: 790–799.
- HARR, B., M. KAUER and C. SCHLÖTTERER, 2002 Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **99**: 12949–12954.
- HEY, J., and R. M. KLIMAN, 2002 Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* **160**: 595–608.
- HUDSON, R. R., 2000 A new statistic for detecting genetic differentiation. *Genetics* **155**: 2011–2014.
- HUDSON, R. R., 2001 Two-locus sampling distributions and their application. *Genetics* **159**: 1805–1817.
- HUDSON, R. R., 2002 Generating samples under a Wright-Fisher neutral model. *Bioinformatics* **18**: 337–338.
- HUDSON, R. R., M. KREITMAN and M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.
- HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992 Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.
- JENSEN, J. D., Y. KIM, V. BAUER DUMONT, C. F. AQUADRO and C. D. BUSTAMANTE, 2005 Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* **170**: 1401–1410.
- KASTANIS, P., E. ELIPOPOULOS, G. GOULIELMOS, S. TSAKAS and M. LOUKAS, 2003 Macroevolutionary relationships of species of *Drosophila melanogaster* group based on mtDNA sequences. *Mol. Phylogenet. Evol.* **28**: 518–528.
- 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.
- KAUER, M., D. DIERINGER and C. SCHLÖTTERER, 2003 Non-neutral admixture of immigrant genotypes in African *Drosophila melanogaster* populations from Zimbabwe. *Mol. Biol. Evol.* **20**: 1329–1337.
- KERN, A. D., and D. J. BEGUN, 2005 Patterns of polymorphism and divergence from noncoding sequences of *Drosophila melanogaster* and *D. simulans*: evidence for nonequilibrium processes. *Mol. Biol. Evol.* **22**: 51–62.
- KIM, Y., and W. STEPHAN, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**: 765–777.
- KINDAHL, E. C., 1994 Recombination and DNA polymorphism on the third chromosome of *Drosophila melanogaster*. Ph.D. Thesis, Cornell University, Ithaca, NY.
- KO, W.-Y., R. M. DAVID and H. AKASHI, 2003 Molecular phylogeny of the *Drosophila melanogaster* species subgroup. *J. Mol. Evol.* **57**: 562–573.
- KOPP, A., and J. R. TRUE, 2002 Phylogeny of oriental *Drosophila melanogaster* species group: a multilocus reconstruction. *Syst. Biol.* **51**: 786–805.
- KUMAR, S., K. TAMURA and M. NEI, 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinformatics* **5**: 150–163.
- LACHAISE, D., and J.-F. SILVAIN, 2004 How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster*-*D. simulans* palaeogeographic riddle. *Genetica* **120**: 17–39.
- LACHAISE, D., M. CARIOU, J. R. DAVID, F. LEMEUNIER, L. TSACAS *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup, pp. 159–225 in *Evolutionary Biology*, Vol. 22, edited by M. K. HECHT, B. WALLACE and G. T. PRANCE. Plenum Press, New York.
- LACHAISE, D., M. HARRY, M. SOLIGNAC, F. LEMEUNIER, V. BÉNASSI *et al.*, 2000 Evolutionary novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from São Tomé. *Proc. R. Soc. Lond. Ser. B* **267**: 1487–1495.
- LEMEUNIER, F., 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.
- MANTEL, N., 1967 The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**: 209–220.
- 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* **145**: 1713–1725.
- OMETTO, L., S. GLINKA, D. DE LORENZO and W. STEPHAN, 2005 Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. *Mol. Biol. Evol.* **22**: 2119–2130.
- ORENGO, D. J., and M. AGUADÉ, 2004 Detecting the footprint of selection in a European population of *Drosophila melanogaster*: multilocus pattern of variation and distance to coding regions. *Genetics* **167**: 1759–1766.
- POOL, J. E., V. BAUER DUMONT, J. L. MUELLER and C. F. AQUADRO, 2006 A scan of molecular variation leads to the narrow localization of a selective sweep affecting both Afrotropical and cosmopolitan populations of *Drosophila melanogaster*. *Genetics* **172**: 1093–1105.
- PRITCHARD, J. K., M. STEPHENS and P. DONNELLY, 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- REICH, D. E., M. CARGILL, S. BOLK, J. IRELAND, P. C. SABETI *et al.*, 2001 Linkage disequilibrium in the human genome. *Nature* **411**: 199–204.
- ROSENBERG, N. A., 2004 DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* **4**: 137.
- ROZAS, J., J. C. SÁNCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003 DnaSP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- RUSSO, C. A. M., N. TAKEZAKI and M. NEI, 1995 Molecular phylogeny and divergence times of *Drosophilid* species. *Mol. Biol. Evol.* **12**: 391–404.
- SCHLÖTTERER, C., H. NEUMEIER, C. SOUSA and V. NOLTE, 2006 Highly structured Asian *Drosophila melanogaster* populations: A new tool for hitchhiking mapping? *Genetics* **172**: 287–292.

- STANLEY, S. M., P. A. PARSONS, G. E. SPENCE and L. WEBER, 1980 Resistance of species of the *Drosophila melanogaster* subgroup to environmental extremes. *Aust. J. Zool.* **28**: 413–421.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TAMURA, K., S. SUBRAMANIAN and S. KUMAR, 2004 Temporal pattern of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol. Biol. Evol.* **21**: 36–44.
- THORNTON, K., 2003 libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* **19**: 2325–2327.
- THORNTON, K., and P. ANDOLFATTO, 2006 Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* **172**: 1607–1619.
- VAN HERREWEGE, J., and J. R. DAVID, 1997 Starvation and desiccation tolerances in *Drosophila*: comparison of species from different climatic origins. *Ecoscience* **4**: 151–157.
- VEUILLE, M., E. BAUDRY, M. COBB, N. DEROME and E. GRAVOT, 2004 Historicity and population genetics of *Drosophila melanogaster* and *D. simulans*. *Genetica* **120**: 61–70.
- VOUIDIBIO, J., P. CAPY, D. DEFAYE, E. PLA, J. SANDRIN *et al.*, 1989 Short-range genetic structure of *Drosophila melanogaster* populations in an Afrotropical urban area and its significance. *Proc. Natl. Acad. Sci. USA* **86**: 8442–8446.
- WALL, J. D., P. ANDOLFATTO and M. PRZEWORSKI, 2002 Testing models of selection and demography in *Drosophila simulans*. *Genetics* **162**: 203–216.

Communicating editor: M. K. UYENOYAMA