

# Distinctly Different Sex Ratios in African and European Populations of *Drosophila melanogaster* Inferred From Chromosomewide Single Nucleotide Polymorphism Data

Stephan Hutter,<sup>\*,1</sup> Haipeng Li,<sup>\*,†</sup> Steffen Beisswanger,<sup>\*</sup> David De Lorenzo<sup>\*,2</sup>  
and Wolfgang Stephan<sup>\*</sup>

<sup>\*</sup>Section of Evolutionary Biology, Biocenter, University of Munich, 82152 Planegg-Martinsried, Germany and <sup>†</sup>Institute of Genetics, University of Cologne, 50674 Cologne, Germany

Manuscript received May 4, 2007

Accepted for publication July 15, 2007

## ABSTRACT

It has been hypothesized that the ratio of X-linked to autosomal sequence diversity is influenced by unequal sex ratios in *Drosophila melanogaster* populations. We conducted a genome scan of single nucleotide polymorphism (SNP) of 378 autosomal loci in a derived European population and of a subset of 53 loci in an ancestral African population. On the basis of these data and our already available X-linked data, we used a coalescent-based maximum-likelihood method to estimate sex ratios and demographic histories simultaneously for both populations. We confirm our previous findings that the African population experienced a population size expansion while the European population suffered a population size bottleneck. Our analysis also indicates that the female population size in Africa is larger than or equal to the male population size. In contrast, the European population shows a huge excess of males. This unequal sex ratio and the bottleneck alone, however, cannot account for the overly strong decrease of X-linked diversity in the European population (compared to the reduction on the autosome). The patterns of the frequency spectrum and the levels of linkage disequilibrium observed in Europe suggest that, in addition, positive selection must have acted in the derived population.

**I**N recent years genomic scans of DNA sequence variation using single nucleotide polymorphisms (SNPs) have been performed for multiple species. These studies became possible by the availability of full genome sequences, and data are now available from a variety of organisms such as *Drosophila melanogaster* (GLINKA *et al.* 2003; ORENGO and AGUADÉ 2004; OMETTO *et al.* 2005), humans (AKEY *et al.* 2004), and *Arabidopsis thaliana* (SCHMID *et al.* 2005). These data sets provide useful tools to address questions such as estimating population sizes and demographic histories of a species. One of the main conclusions of the work performed on *D. melanogaster* was that demographic events are major factors in shaping the patterns of DNA polymorphism (*e.g.*, ANDOLFATTO 2001; GLINKA *et al.* 2003; HADDRILL *et al.* 2005b). *D. melanogaster* is thought to have originated in sub-Saharan Africa and to have only relatively recently (10,000–15,000 years ago) colonized the rest of the world (DAVID and CAPY 1988;

LACHAISE *et al.* 1988). Populations that reside in the ancestral species range show signatures of population size expansion (GLINKA *et al.* 2003; POOL and AQUADRO 2006), while derived populations have polymorphism patterns compatible with population size bottlenecks (*e.g.*, ANDOLFATTO 2001; GLINKA *et al.* 2003). Theoretical studies have since utilized the data obtained from genome scans to estimate the parameters of these demographic events (HADDRILL *et al.* 2005b; LI and STEPHAN 2006).

The aforementioned SNP-based genome scans in *D. melanogaster* were performed solely for noncoding regions on the X chromosome. This leaves us virtually ignorant about noncoding autosomal variation on a chromosomal scale. In general, data quantifying the amount of autosomal variation based on SNPs is rather scarce for *D. melanogaster*. Only recently a study surveying SNP polymorphism in an autosomal set of coding regions was published (SHAPIRO *et al.* 2007). Yet this study does not address the relationship between autosomal and X-linked diversity. ANDOLFATTO (2001) used a data set compiled from different previous studies to compare levels of X-linked and autosomal diversity in African and non-African flies. The drawback of this study, however, is that it combines samples from many different publications. The data set is therefore not a representative population sample. In addition, the loci

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AM701830–AM706343.

<sup>1</sup>Corresponding author: Section of Evolutionary Biology, Biocenter, University of Munich, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany. E-mail: hutter@zi.biologie.uni-muenchen.de

<sup>2</sup>Present address: Department of Basic Medical Sciences, University of Lleida, 25199 Lleida, Spain.

are mostly genes that were chosen in the original studies because of their unusual patterns of polymorphism. Thus, they may not represent an unbiased set of loci such as those used in genome scans (see above). That being said, the comparison of X-linked and autosomal diversity showed a clear pattern: the relative levels of polymorphism on the X chromosome and the autosomes did not accord with standard neutral expectations. Under the standard neutral model the ratio of X-linked to autosomal polymorphism should be equal to the relative population sizes of the chromosomes, which is 3/4. In the data set of ANDOLFATTO (2001) the X chromosome exhibited more variation than expected under standard neutrality in Africa and too few polymorphisms outside Africa.

These deviations from standard neutral expectations can be caused by the action of natural selection (AQUADRO *et al.* 1994). In the case of positive Darwinian selection X-linked loci are believed to be more affected by genetic hitchhiking than autosomes (MAYNARD SMITH and HAIGH 1974; CHARLESWORTH *et al.* 1987), which will lead to an  $X/A$  ratio of diversity lower than the neutral expectation of 3/4. If background selection (CHARLESWORTH *et al.* 1993) is predominant, then this ratio is expected to be  $>3/4$ . Alternatively, nonselective forces can also alter the ratio of X-linked to autosomal diversity. There is evidence that effective population sizes for males and females are not equal in natural populations of *D. melanogaster* (reviewed in CHARLESWORTH 2001). If this is the case then the ratio of X-linked to autosomal variation will not correspond to the standard expectation. Since females carry two-thirds of the X chromosomes in a population but only one-half of the autosomes, changes in female population size will affect X chromosomes more strongly than autosomes. In other words, if the female population size is larger than the male one, the X-chromosomal to autosomal ratio of diversity is expected to be  $>3/4$ , while a smaller female population size leads to ratios  $<3/4$ .

KAUER *et al.* (2002) performed a genome scan of variability of 133 microsatellite loci in African and non-African populations on both the X chromosome and autosomes. Comparing heterozygosities of microsatellites can be problematic, because each locus has a specific mutation rate. Thus, there is a risk of having mutational biases. The data were again from multiple populations in and outside Africa, but the authors corrected for possible biases by forming across-population averages. The results of this study confirm the findings of ANDOLFATTO (2001). The authors conclude that background selection shaped polymorphism in the ancestral African populations while positive selection was prevalent outside Africa. However, unequal sex ratios could also have contributed to the patterns (KAUER *et al.* 2002). Both the SNP and the microsatellite studies did not account for the demographic history of *Drosophila*. Events such as population size expansions and bottle-

necks may have different effects on X chromosomes and autosomes, which may further complicate a direct comparison of both chromosomes.

We present here the first genome scan of an autosome in *D. melanogaster* using noncoding SNP markers. We surveyed a total of 378 loci located on chromosome 3 in a European population. In addition, we analyzed a random subset of 53 loci in an African population to get an estimate of autosomal diversity for an ancestral population. The individuals analyzed come from the same populations that have already been used in previous scans of X-linked diversity (GLINKA *et al.* 2003; OMETTO *et al.* 2005). By combining the data sets we now have the opportunity to study X chromosomes and autosomes within a single ancestral and a derived population. To make use of the additional information that SNP data can provide, we also analyzed statistics describing the frequency spectrum of mutations as well as linkage disequilibrium (LD). To address the possibility of unequal sex ratios and demography shaping polymorphism, we extended the likelihood method described in LI and STEPHAN (2006). This approach allows us to estimate the most likely sex ratios in both populations while taking their demographic history into account.

## MATERIALS AND METHODS

**Biological material:** The individuals come from a total of 24 highly inbred lines sampled from an African population from Zimbabwe and a European population from The Netherlands (GLINKA *et al.* 2003).

**Chromosomal inversions:** For inversion analysis we individually crossed male flies to virgin Canton-S females, homozygous for the standard chromosome arrangement. We prepared salivary glands from late F<sub>1</sub> third-instar larvae, maintained at 18°. Several larvae per *D. melanogaster* line were dissected to account for low-frequency inversions. Polytene chromosomes were prepared using the lacto-acetic orcein method and viewed under an inverted phase contrast microscope. Banding patterns and inversion breakpoints were identified according to standard chromosome maps (LEFEVRE 1976).

**Data collection:** Primers for 378 loci (125 on chromosome 3L, 253 on chromosome 3R) were designed on the basis of the *D. melanogaster* genome Release 3.2 (<http://www.flybase.org>). Loci were chosen evenly spaced along the chromosome and located in intronic or intergenic regions. All 378 genomic regions were sequenced in the European sample, while only a random subset of 53 fragments (17 on 3L, 36 on 3R) was sequenced in the African one.

Sequence data were obtained by means of capillary sequencing on both strands. As the *D. melanogaster* genome Release 4.2.1 became available during the course of our study, we rechecked all our loci for overlap with coding regions. One locus (namely 3-480) overlapped partially with a coding exon. The overlapping part was removed from the alignment and only the noncoding sequence was used for analysis. The new sequences were deposited into the European Molecular Biology Laboratory (EMBL) database under accession nos. AM701830–AM706343. To obtain an outgroup sequence for each of our fragments homologous sequences were searched in the *D. simulans* genome, Mosaic Assembly Release 1.0 (Genome

Sequencing Center, Washington University School of Medicine, St. Louis), via BLAST. Only for four fragments a homolog could not be found, resulting in a total of 374 alignments in the European and 51 alignments in the African population that contain outgroup information. On the basis of the annotation of the *D. melanogaster* genome Release 4.2.1, we updated our published data on the X-linked noncoding loci (OMETTO *et al.* 2005). Loci that showed complete or almost complete overlap with coding regions were removed from the data set. Where only a partial overlap existed, the coding regions were removed from the alignments. The final X-chromosomal data set consisted of 259 loci for the European and 249 loci for the African population.

**Statistical analysis:** We estimated basic population genetic statistics, such as levels of nucleotide diversity per site measured as average pairwise distance  $\pi$  (TAJIMA 1983) and Watterson's estimator  $\theta$  (WATTERSON 1975) as well as Tajima's  $D$  statistic (TAJIMA 1989). We calculated the  $P$ -value of Tajima's  $D$  on the basis of coalescent simulations assuming the standard neutral model (10,000 iterations). SCHAEFFER (2002) noted that it is difficult to compare values of Tajima's  $D$  between different loci, if they show differences in sample size and/or number of segregating sites. Schaeffer therefore suggests using the ratio of Tajima's  $D$  to  $D_{\min}$  to overcome this limitation, where  $D_{\min}$  is the absolute value of the theoretical minimum of  $D$ . Thus, the  $D/D_{\min}$  statistic and  $D$  have the same sign. Since the loci we are dealing with here do indeed vary greatly in sample size and number of segregating sites, we used the  $D/D_{\min}$  statistic (SCHAEFFER 2002) to summarize the frequency spectrum when building averages over multiple loci or comparing loci among each other. The LD statistic  $Z_{nS}$  (KELLY 1997) and divergence ( $K$ ) corrected for multiple hits (JUKES and CANTOR 1969) to *D. simulans* were estimated by the program VariScan (HUTTER *et al.* 2006). Since  $Z_{nS}$  values are biased by the sample size of the locus we cannot compare loci with different allele numbers directly (similar to Tajima's  $D$ ). We therefore used the following approach: only loci with eight alleles or more were considered. If a locus contained more than eight alleles all  $\binom{8}{n}$  combinations of alignments with exactly eight alleles were generated and the average  $Z_{nS}$  value was calculated. This statistic, which we call  $Z_{nSS}$ , was used to describe the locus-specific LD. To estimate recombination rates for each locus we used the program Recomb-Rate (COMERON *et al.* 1999), which follows an approach by KLIMAN and HEY (1993). Levels of recombination are given as recombination rate per base pair per generation  $\times 10^{-8}$ . A supplemental table describing the summary statistics for all loci used in this study is available at <http://www.genetics.org/supplemental/>.

**Demographic modeling and estimation of the sex ratio in the African population:** For a diploid sexual population, we denote the number of females as  $N_f$ , the number of males as  $N_m$ , the effective population size of the X chromosome as  $N_x$ , the effective population size of the autosome as  $N_a$ , and the sex ratio as  $\beta = N_f/N_m$ . We assume that the sex ratio is constant (over time) within a population, but may vary from one population to another one.

Then we have  $N_x = 9N_m N_f / (4N_m + 2N_f) = (9\beta / (4 + 2\beta)) N_m$ , and  $N_a = 4N_m N_f / (N_m + N_f) = (4\beta / (1 + \beta)) N_m$  (HEDRICK 2000), where  $N_x$  and  $N_a$  are large relative to the sample size. Their ratio is

$$N_x/N_a = 9(1 + \beta)/8(2 + \beta). \quad (1)$$

$N_x/N_a$  has the lower bound  $\lim_{\beta \rightarrow 0} (N_x/N_a) = 0.5625$  and the upper bound  $\lim_{\beta \rightarrow \infty} (N_x/N_a) = 1.125$ .

Following LI and STEPHAN (2006), we assume that the demographic history of the African population is characterized by an instantaneous expansion model. That is, the

effective population size of the X chromosome increased instantaneously from  $N_{x1}$  to  $N_{x0}$  at  $2N_{x0}t_x (= 2N_{a0}t_a)$  generations ago, where  $N_{x1}$  and  $N_{x0}$  are the ancestral and current effective population sizes of the X chromosome in the African population, respectively, and  $N_{a0}$  is the current effective population size of the autosome in the African population. Then we have  $t_a = (9(1 + \beta)/8(2 + \beta))t_x$ . Thus the time back to the expansion for the autosome (in units of  $2N_{a0}$ ) may be different from that for the X chromosome (in units of  $2N_{x0}$ ), while the strength of the expansion is the same (*i.e.*,  $N_{x0}/N_{x1} = N_{a0}/N_{a1}$ ), where  $N_{a1}$  is the ancestral effective population size of the autosome in the African population.

The average mutation rates of the X and the autosome are denoted by  $\bar{\mu}_x$  and  $\bar{\mu}_a$  (per base pair per generation), respectively. We assume a generation time of 10 generations per year. Mutation rates are estimated from divergence between *D. melanogaster* and *D. simulans*, assuming a species split 2.3 million years ago (LI *et al.* 1999). We then have  $\theta_x = 4N_{x0}\bar{\mu}_x$  and  $\theta_a = 4N_{a0}\bar{\mu}_a = (8(2 + \beta)/9(1 + \beta))\bar{\mu}_x\theta_x$ . Thus, the unknown parameters in the model are  $\theta_x$ ,  $t_x$ ,  $N_{x0}/N_{x1}$ , and  $\beta$ .

Following LI and STEPHAN (2006), we summarize the SNP data in terms of the mutation frequency spectrum (MFS). The likelihood for the  $k$ th locus on the X chromosome is then given as  $L_{xk} = P(\text{MFS} | \bar{C}_{xk}) = \prod_{i=1}^{n_k-1} P(\xi_{ik} | E(l_{ik}))$ , where  $\bar{C}_{xk}$  is a set of  $(n_k - 1)$  expected branch lengths under the demographic scenario for the X chromosome. The branch length is scaled so that 1 unit represents  $2N_{x0}$  generations,  $n_k$  is the sample size of the  $k$ th locus,  $\xi_{ik}$  is the number of derived mutations carried by  $i$  sampled chromosomes for the  $k$ th locus, and  $E(l_{ik})$  is the expected length of branches with  $i$  descendants for the  $k$ th locus under the demographic scenario.  $P(\xi_{ik} | E(l_{ik}))$  is given by the Poisson probability; *i.e.*,  $P(\xi_{ik} | E(l_{ik})) = \lambda_{ik}^{\xi_{ik}} e^{-\lambda_{ik}} / \xi_{ik}!$ , with  $\lambda_{ik} = E(l_{ik})\theta_{xk}/2$ , where  $\theta_{xk} = 4N_{x0}\mu_k$ , and  $\mu_k$  is the mutation rate of the  $k$ th locus (per locus).

The likelihood for the  $k$ th locus on the autosome is given as  $L_{ak} = P(\text{MFS} | \bar{C}_{ak}) = \prod_{i=1}^{n_k-1} P(\xi_{ik} | E(l_{ik}))$ , and  $P(\xi_{ik} | E(l_{ik})) = \lambda_{ik}^{\xi_{ik}} e^{-\lambda_{ik}} / \xi_{ik}!$ , with  $\lambda_{ik} = E(l_{ik})\theta_{ak}/2$ , where  $\theta_{ak} = 4N_{a0}\mu_k$ , and the branch length is scaled so that 1 unit represents  $2N_{a0}$  generations.  $\bar{C}_{ak}$  is defined in analogy to  $\bar{C}_{xk}$  for the X chromosome.

Then  $L = \prod_{k=1}^{m_a} L_{ak} \times \prod_{k=1}^{m_x} L_{xk}$ , where  $m_a$  and  $m_x$  are the numbers of loci on the autosome and the X chromosome, respectively. A grid search is performed to maximize the likelihood.

**Demographic modeling and estimation of the sex ratio in the European population:** Since there is convincing evidence that the European population is derived from an ancestral African population (GLINKA *et al.* 2003; BAUDRY *et al.* 2004), we use a two-population model (Figure 2 of LI and STEPHAN 2006) to infer the demographic history and the sex ratio of the European population. In the following, the indexes A and E distinguish the model parameters for the African and the European populations, respectively.

Similar to the definitions for the African population, we have  $\beta_A = N_{fA}/N_{mA}$ ,  $N_{xA0}/N_{aA0} = 9(1 + \beta_A)/8(2 + \beta_A)$ ,  $\theta_{xA} = 4N_{xA0}\bar{\mu}_x$ ,  $\theta_{aA} = 4N_{aA0}\bar{\mu}_a = (8(2 + \beta_A)/9(1 + \beta_A))\bar{\mu}_x\theta_{xA}$ , and  $t_{aA0} = (9(1 + \beta_A)/8(2 + \beta_A))t_{xA0}$ . For the European population, we have  $\beta_E = N_{fE}/N_{mE}$ ,  $N_{xE0}/N_{aE0} = 9(1 + \beta_E)/8(2 + \beta_E)$ ,  $\theta_{xE} = 4N_{xE0}\bar{\mu}_x$ , and  $\theta_{aE} = 4N_{aE0}\bar{\mu}_a = (8(2 + \beta_E)/9(1 + \beta_E))\bar{\mu}_x\theta_{xE}$ . We assume that the sex ratio of the European population is constant and may be different from that of the ancestral African population.

Following LI and STEPHAN (2006), we assume that the demographic history of the European population is characterized by an instantaneous bottleneck model. The demographic history of the European population for the X chromosome is parameterized by  $t_{xE0}$ ,  $t_{xE1}$ ,  $\text{strength}_{xE} (= N_{xE0}/N_{xE1})$ , and  $\text{ratio}_{xE} (= N_{xA0}/N_{xE0})$ , where time is given in units

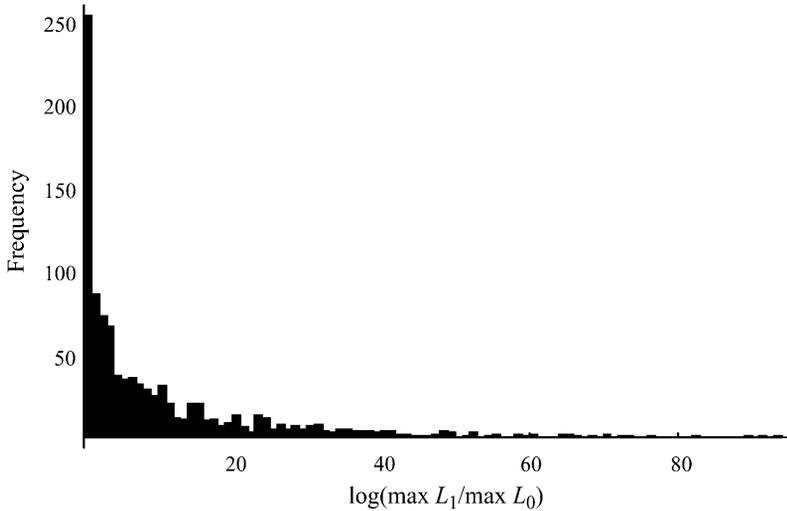


FIGURE 1.—Empirical distribution of  $\zeta = \log(\max L_1/\max L_0)$  that is obtained by analyzing 1000 simulated African data sets conditional on the null hypothesis (see text). The analysis is based on the profile likelihood for sex ratio. The critical value ( $\zeta_{95\%} = 44.46$ ) is obtained from the empirical distribution and is further used in the LRT.

of  $2N_{\text{xAO}}$  generations. Similarly, the demographic history of the European population for the autosome is parameterized by  $t_{\text{aE0}} (= (9(1 + \beta_{\text{A}})/8(2 + \beta_{\text{A}}))t_{\text{xEO}})$ ,  $t_{\text{aE1}} (= (9(1 + \beta_{\text{A}})/8(2 + \beta_{\text{A}}))t_{\text{xE1}})$ ,  $\text{strength}_{\text{aE}} (= N_{\text{aE0}}/N_{\text{aE1}} = \text{strength}_{\text{xE}})$ , and  $\text{ratio}_{\text{aE}} (= N_{\text{aAO}}/N_{\text{aEO}} = ((2 + \beta_{\text{A}})(2 + \beta_{\text{E}})/(1 + \beta_{\text{A}})(1 + \beta_{\text{E}}))\text{ratio}_{\text{xE}})$ , where time is measured in units of  $2N_{\text{aAO}}$  generations. Thus, the unknown parameters in the two-population model are  $t_{\text{xEO}}$ ,  $t_{\text{xE1}}$ ,  $\text{strength}_{\text{xE}}$ ,  $\text{ratio}_{\text{xE}}$ , and  $\beta_{\text{E}}$  because the parameters for the African (ancestral) population are estimated according to the procedure described above.

We summarize the SNP data in the two populations in terms of the joint mutation frequency spectrum. The maximum-likelihood method outlined previously (LI and STEPHAN 2006) is used to estimate the demographic scenario and the sex ratio in the derived European population. In this analysis, we used only the fragments that are sequenced in both populations.

**Likelihood-ratio test and likelihood-based confidence intervals:** The likelihood-ratio test (LRT) is a statistical test of the goodness-of-fit. If the null model and the alternative model are hierarchically nested, and the former model has one parameter less than the latter, then we define  $\zeta = \log(\max L_1/\max L_0)$ , where  $L_1$  and  $L_0$  are the likelihoods for the alternative and null models, respectively. Then we have  $\zeta \geq 0$  because of  $\max L_1 \geq \max L_0$ . Since  $\zeta$  may not be approximated by a  $\chi^2$ -distribution with 1 d.f., we obtain the empirical distribution of  $\zeta$  from 1000 simulated data sets under the null model. The LRT (a one-tail test) is conducted as follows: we reject the null model at the 5% significance level if  $\zeta_{95\%} < \zeta_{\text{obs}}$ , where  $\zeta_{95\%}$  is the critical value.

For the African population, polymorphism data sets of the X and the autosome are simulated conditional on the constant population size model and the local recombination rate (COMERON *et al.* 1999). The coalescent process was described previously (LI and STEPHAN 2006). We assume that there is no recombination within loci since the average fragment length is  $\sim 500$  bp. The sex ratio is 1. An empirical distribution of  $\zeta$  is shown in Figure 1.

For the European population, a two-population model is used to simulate the joint MFS (LI and STEPHAN 2006). The estimated African expansion scenario is used, and we assume that no bottleneck occurred when the European population is derived from the ancestral African population. The current population size is equal between the two populations. The joint MFS of the X chromosome is simulated conditional on the local recombination rate (COMERON *et al.* 1999). To sim-

ulate the joint MFS of the autosome, we assume that the autosomal loci are independent. The sex ratio is 1 for both populations.

In the case of a multiparameter model ( $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ ), we may be interested in the confidence interval (C.I.) of one parameter at a time, say in  $\vartheta_1$ . Let  $L_p(\vartheta_1) = \max_{\vartheta_2, \dots, \vartheta_k} L(\vartheta_1, \vartheta_2, \dots, \vartheta_k)$  be the profile likelihood. Then the likelihood-based approximate 95% C.I. is  $\{\vartheta_1, \log(L_p(\hat{\vartheta}_1)/L_p(\vartheta_1)) \leq \zeta_{95\%}\}$ , where  $\hat{\vartheta}_1$  is the maximum-likelihood estimate of  $\vartheta_1$  (PAWITAN 2001).

## RESULTS

**Inversion patterns:** In the African sample we detected inversions on chromosome 3L in line 145. Inversions on 3R were found in lines 131, 157, and 229. In the European sample we detected an inversion on 3R in line 13. We did not observe any 3L inversions in Europe. Genes included in an inversion do not recombine with genes on the standard chromosome, with the exception of rare double crossovers (WESLEY and EANES 1994; AULARD *et al.* 2002). Therefore, we excluded lines showing an inversion from subsequent analyses.

**Autosomal polymorphism patterns of the European population:** We analyzed a total of 378 fragments located on both arms of chromosome 3 (125 fragments on 3L and 253 fragments on 3R) in 11 inversion-free lines of the European population. The length of fragments (excluding insertions and deletions) ranges between 162 and 672 bp with a mean of 536 bp. On average data could be obtained from 10.8 lines. Fragments located on 3L have an average distance of 63 kb encompassing a total of 7.2 Mb and show recombination rates of  $3.7\text{--}5.0 \times 10^{-8}$ . Fragments located on 3R are on average 46 kb apart, spanning a total region of 11.7 Mb. Their recombination rates lie between  $1.2$  and  $3.9 \times 10^{-8}$ . Data from both chromosomal arms were pooled and analyzed jointly to cover a broad range of recombination rates.

**TABLE 1**  
**Summary statistics of the four data sets analyzed**

	Autosome		X chromosome	
	Europe	Africa	Europe <sup>b</sup>	Africa <sup>b</sup>
No. of loci	378	53	259	249
Average sample size	10.96	7.96	11.86	11.72
Average length of alignment <sup>a</sup>	536.1	510.5	527.7	501.3
Average (SE) recombination rates $\times 10^{-8}$	3.3 (0.06)	3.4 (0.16)	3.5 (0.07)	3.5 (0.07)
Average (SE) GC content in %	41.4 (0.003)	41.6 (0.007)	39.1 (0.004)	39.0 (0.004)
No. of invariant loci	1	0	21	0
Average (SE) $\theta$ in %	0.68 (0.02)	1.14 (0.11)	0.47 (0.03)	1.34 (0.05)
Average (SE) $\pi$ in %	0.66 (0.03)	1.04 (0.10)	0.48 (0.03)	1.17 (0.04)
Average (SE) divergence ( $K$ ) in %	5.0 (0.16)	5.2 (0.46)	6.8 (0.22)	6.4 (0.19)
Average (SE) $\theta/K$	0.16 (0.005)	0.25 (0.025)	0.08 (0.004)	0.22 (0.006)
Average (SE) $D/D_{\min}$	-0.08 (0.027)	-0.25 (0.052)	-0.09 (0.045)	-0.32 (0.018)
Significant Tajima's $D$ values (+/-)	15/32	0/3	25/50	0/21
Average (SE) $Z_{nSS}$	0.40 (0.010)	0.29 (0.019)	0.51 (0.018)	0.20 (0.005)

<sup>a</sup>Excluding indels.

<sup>b</sup>The updated data set (OMETTO *et al.* 2005) based on the annotation of the *D. melanogaster* genome Release 4.2.1.

Of the 378 loci surveyed, only a single fragment has no polymorphism. This lack of polymorphism does not result from an overly short alignment (639 bp) or reduced mutation rate (divergence to *D. simulans* is 0.084). The  $\theta$  value averaged over all fragments is 0.0068 with a standard error (SE) of 0.0002. This is only approximately half as high as the diversity levels reported for synonymous sites on autosomes in non-African populations (0.0155; ANDOLFATTO 2001). This difference can be explained by the fact that the numbers from ANDOLFATTO (2001) come from a data set that combined multiple populations, hence inflating estimates of diversity. Additionally there is evidence that there is a systematic difference between noncoding and silent sites, indicating that there seems to be more purifying selection on sites located in noncoding regions compared to nonsynonymous sites (ANDOLFATTO 2005). The mean (SE) divergence to *D. simulans* is 0.050 (0.0016). For the  $D/D_{\min}$  statistic describing the frequency distribution we observed a mean (SE) of -0.08 (0.027), which is close to the standard neutral expectation. When looking at Tajima's  $D$  values of the loci individually we found that a total of 47 of 377 fragments differ from standard neutral expectations; 15 are significantly positive and 32 significantly negative.  $Z_{nSS}$  values show an average (SE) of 0.40 (0.010). The summary statistics are shown in Table 1.

**Autosomal polymorphism patterns of the African population:** A random subset of 53 fragments was sequenced in the African population. Thirty-six fragments are located on 3R and 17 on 3L. Again only lines harboring no inversions were used. On average data could be obtained for 7.96 lines.

The mean (SE) level of diversity is 0.0114 (0.0011). No invariant loci were observed. Comparing this to

previous results from synonymous sites in African populations we can see that our diversity estimates are again lower (0.0161 on autosomes; ANDOLFATTO 2001).  $D/D_{\min}$  has a mean of -0.25 with a SE of 0.052. A locus-by-locus inspection of the frequency spectrum revealed that three loci deviate significantly from the standard neutral model; all of them have negative values. Values of  $Z_{nSS}$  show a mean (SE) of 0.29 (0.019). The numbers are summarized in Table 1.

**Contrasting autosomal patterns between populations:** At first we investigated if our African subset of loci deviates from the European set in terms of average recombination or mutation rate (estimated by divergence to *D. simulans*). Neither recombination rate (Mann-Whitney  $U$ -test,  $P = 0.77$ ) nor divergence (Mann-Whitney  $U$ -test,  $P = 0.91$ ) is statistically different. These factors should therefore not influence our comparisons between the populations. Average levels of diversity are significantly lower in Europe (Mann-Whitney  $U$ -test,  $P < 0.001$ ). The mean  $D/D_{\min}$  is significantly higher in Europe (Mann-Whitney  $U$ -test,  $P = 0.02$ ). Both populations are known to have undergone different demographic events (LI and STEPHAN 2006). Such events affect not only means of summary statistics describing patterns of polymorphism but also their variance. Therefore contrasting variances of statistics such as  $D/D_{\min}$  and  $Z_{nSS}$  can provide useful information on the demographic history (*e.g.*, HADDRILL *et al.* 2005b). Tables 2 and 3 summarize the variances of  $D/D_{\min}$  and  $Z_{nSS}$  for our data sets. To find out if empirical variances differ between chromosomes and populations we conducted Levene tests. Comparing variances of  $D/D_{\min}$  between European and African autosomes we find that the empirical variance is significantly higher in Europe (Levene test,  $P = 0.01$ ). To check if this larger variance

**TABLE 2**  
Pairwise tests for differences in variance of  $D/D_{\min}$

	$D/D_{\min}$		Levene test
	X chromosome	Autosome	
Europe	0.492	0.269	$P < 0.001$
Africa	0.085	0.144	$P = 0.012$
Levene test	$P < 0.001$	$P = 0.009$	

also results in an increase of significant Tajima's  $D$  values we conducted a Fisher's exact test. Neither the number of positive values ( $P = 0.14$ ) nor the number of negative values ( $P = 0.37$ ) is significantly increased in the European population. LD behaves similarly to the frequency spectrum: the mean value of  $Z_{\text{nsS}}$  (Mann-Whitney  $U$ -test,  $P < 0.001$ ) and its variance (Levene test,  $P < 0.001$ ) are significantly elevated in Europe.

**X-linked data:** We calculated the same statistics from the updated X-linked data set of OMETTO *et al.* (2005) for comparison although the updating produces little difference in the summary statistics (Table 1).

We compared the X-linked data between populations in the same way as the autosomal ones. Mutation and recombination rates do not differ between Africa and Europe (Mann-Whitney  $U$ -test,  $P = 0.30$  and  $P = 0.84$ , respectively) as found for the autosomal loci. Average levels of diversity and the variance are significantly reduced in Europe (Mann-Whitney  $U$ -test,  $P < 0.027$ , and Levene test,  $P < 0.001$ ). The average  $D/D_{\min}$  value is higher in Europe than in Africa, but this difference is not statistically significant (Mann-Whitney  $U$ -test,  $P = 0.27$ ). The failure to detect a difference here might be due to the high variance in the European population. It is significantly higher than that in Africa (Levene test,  $P < 0.001$ ) and in fact it is the highest in our data set (Table 2). This high variance in the frequency spectrum also leads to an increased number of significant Tajima's  $D$  values in Europe. Both the number of significantly positive and the number of significantly negative values are higher than expected when compared to those in the African population (Fisher's exact test,  $P < 0.001$  in both cases). Comparison of LD showed that the average value of  $Z_{\text{nsS}}$  is higher in Europe (Mann-Whitney  $U$ -test,

**TABLE 3**  
Pairwise tests for differences in variance of  $Z_{\text{nsS}}$

	$Z_{\text{nsS}}$		Levene test
	X chromosome	Autosome	
Europe	0.068	0.038	$P < 0.001$
Africa	0.007	0.019	$P = 0.060$
Levene test	$P < 0.001$	$P < 0.001$	

$P < 0.001$ ), and so is its variance (Levene test,  $P < 0.001$ ).

**Comparison of X chromosomes and autosomes within populations:** It is important to note that levels of divergence are different when comparing the chromosomes. Divergence is elevated on the X chromosome in the European (Mann-Whitney  $U$ -test,  $P < 0.001$ ) and the African data sets (Mann-Whitney  $U$ -test,  $P = 0.001$ ). This cannot be due to systematic differences in mutation rates since studies have shown that evolutionary rates do not differ between the chromosomes in *D. melanogaster* (BETANCOURT *et al.* 2002). HADDRILL *et al.* (2005a) found a negative correlation of divergence and GC content in introns and concluded that base composition influences the local mutation rate. Confirming this hypothesis, we observed a significantly elevated GC content in our autosomal loci for both populations (Mann-Whitney  $U$ -test,  $P < 0.001$  in Europe and  $P = 0.008$  in Africa). To account for the resulting differences in mutation rate we corrected the levels of diversity by dividing individual  $\theta$ -values by the local divergence. These  $\theta/K$  values were then used for estimating the ratios of X-chromosomal to autosomal diversity. Expanding the results of HADDRILL *et al.* (2005a) we find that the effect of base composition is not only confined to introns. We took the combined 232 purely intergenic loci from both chromosomes of the European data set and correlated the GC content with divergence. A significant negative correlation can be observed (Spearman's correlation coefficient  $R = -0.477$ ,  $P < 0.001$ ).

We also tested if levels of recombination are comparable between our X-linked and autosomal data sets. A Mann-Whitney  $U$ -test shows that levels of recombination are not different for either the African data sets ( $P = 0.47$ ) or the European ones ( $P = 0.057$ ). It should, however, be noted that the  $P$ -value of the European population is close to significance.

The ratios of X-linked to autosomal polymorphism are 0.49 for the European and 0.90 for the African population. Previous studies report much higher numbers. ANDOLFATTO (2001) finds a ratio of 0.66 in non-African and 1.60 in African flies using synonymous sites. KAUER *et al.* (2002) used microsatellite heterozygosity and observed ratios of 0.78 outside of Africa and 1.20 in Africa. It should be noted that these studies do not correct for possible differences in mutation rate so these ratios might be subject to mutational biases. When leaving  $\theta$ -values uncorrected for our data set we observe ratios of 0.66 and 1.17 in Europe and Africa, respectively. These numbers are in good agreement with the uncorrected numbers reported in previous studies. BEGUN and WHITLEY (2000) looked at ratios of diversity in non-African *D. simulans* populations. They could correct for different mutation rates since *D. melanogaster* was available as an outgroup. Using these corrected diversity levels they found a X-chromosomal to

autosomal ratio of polymorphism of 0.49, which is exactly the value we obtained for our European data set.

We wanted to know if the observed diversity ratios were significantly different from the expected 0.75 assuming a sex ratio of 1:1 in an equilibrium population. To do this we multiplied X-linked data by  $\frac{4}{3}$  and performed Mann–Whitney  $U$ -tests. We find that the X chromosome lacks diversity in Europe ( $P < 0.001$ ) but is too diverse in Africa ( $P < 0.001$ ). We also compared the patterns of the frequency spectrum and LD. In Europe the average values of  $D/D_{\min}$  do not differ between the X chromosome and the autosome (Mann–Whitney  $U$ -test,  $P = 0.48$ ) while the variance is increased for the X chromosome (Levene test,  $P < 0.001$ ). This larger variance leads to significantly more loci deviating from standard neutral expectations on the X chromosome. There are more loci with significantly positive (Fisher’s exact test,  $P = 0.003$ ) and significantly negative values ( $P < 0.001$ ) than expected in comparison with the autosome. The  $Z_{nSS}$  statistic shows an elevated level of LD on the X chromosome (Mann–Whitney  $U$ -test,  $P < 0.001$ ) along with an elevated variance (Levene test,  $P < 0.001$ ). A study examining the pattern of LD in non-African *D. simulans* found a very similar pattern (WALL *et al.* 2002). In the African population mean  $D/D_{\min}$  values do not differ statistically either (Mann–Whitney  $U$ -test,  $P = 0.13$ ) while the variance is higher on the autosome (Levene test,  $P = 0.012$ ). The mean LD is higher on the autosome (Mann–Whitney  $U$ -test,  $P = 0.001$ ) while the variances do not differ significantly (Levene test,  $P = 0.06$ ). The means and 95% confidence intervals of  $\theta/K$ ,  $D/D_{\min}$ , and  $Z_{nSS}$  for all four data sets are shown in Figure 2 for better comparison.

**Inferring the demographic history and the sex ratio in the African and European populations:** On the basis of X-linked data it has been proposed that the African population has expanded in recent time (OMETTO *et al.* 2005; LI and STEPHAN 2006). To further examine this hypothesis, we compare the instantaneous population expansion model with the standard neutral model, using both our X-linked and our autosomal data sets. That is, the null hypothesis is  $N_{xA0} = N_{xA1}$  and  $N_{aA0} = N_{aA1}$  (the simple model), and the alternative hypothesis is  $N_{xA0} \geq N_{xA1}$  and  $N_{aA0} \geq N_{aA1}$  (the complex model). The LRT ( $P < 0.01$ , the critical value  $\zeta_{99\%} = 41.21$ ) suggests that the expansion model [ $\max\log(L) = -4754.2$ ] explains the features of the polymorphism data in the African sample significantly better than the standard neutral model [ $\max\log(L) = -4896.3$ ].

Since the *D. melanogaster* lineage split from *D. simulans*  $\sim 2.3$  million years ago (LI *et al.* 1999) and the average divergence over loci on the X and third chromosome is 0.0667 and 0.0522, respectively,  $\bar{\mu} = 1.450 \times 10^{-9}$  and  $1.135 \times 10^{-9}$  per site per generation (assuming 10 generations per year). Thus, in the expansion model,  $\hat{\theta}_{xA0} (= 4N_{xA0}\bar{\mu}_x) = 0.050$  (*i.e.*,  $\hat{\theta}_{aA0} = 4N_{aA0}\bar{\mu}_a = 0.047$ ). Then  $\hat{N}_{xA0}$  and  $\hat{N}_{aA0}$  are  $8.621 \times 10^6$  and  $10.40 \times 10^6$ ,

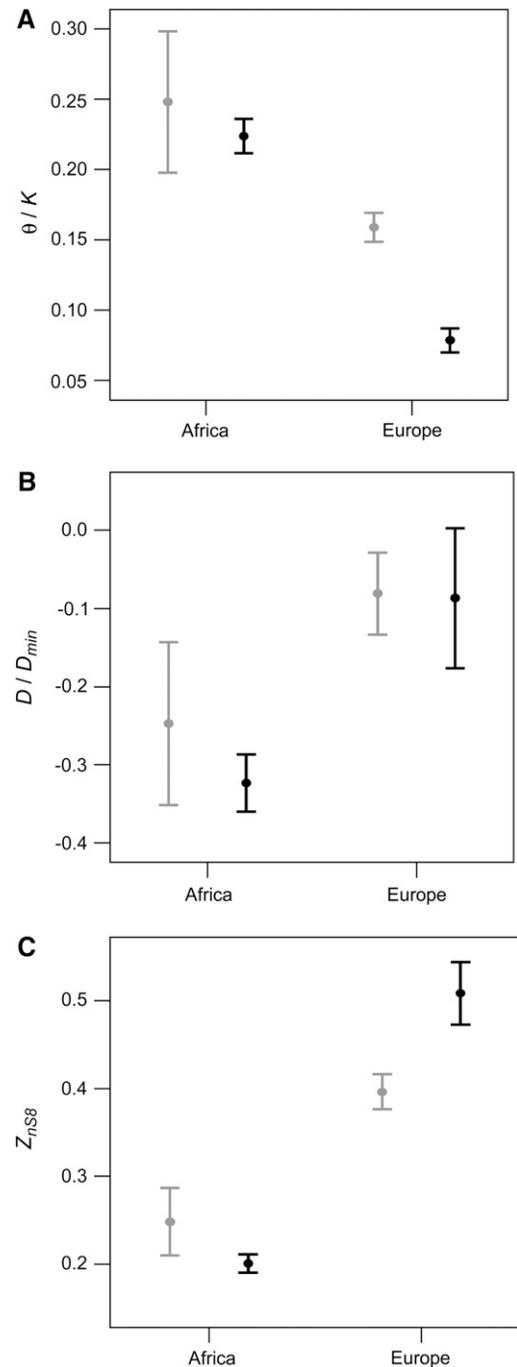


FIGURE 2.—Means of (A)  $\theta/K$ , (B)  $D/D_{\min}$ , and (C)  $Z_{nSS}$  for both populations. X chromosomes are solid and autosomes are shaded. The error bars indicate 95% confidence intervals.

respectively, and the ratio of population sizes ( $N_{xA0}/N_{aA0}$ ) is 0.829 with a 95% confidence interval of (0.636, 1.08). The ratio  $N_{xA0}/N_{aA0}$  is slightly less than the ratio of X-linked to autosomal polymorphism obtained from Watterson’s  $\theta$  (0.90). When the ratio of X-linked to autosomal polymorphism is inferred as the ratio  $N_{xA0}/N_{aA0}$ , it is assumed that the African population is under equilibrium. However, after the population size expands, genetic diversity on the X chromosome is expected to reach equilibrium before that of autosomes because of the smaller

effective population size of the X chromosome, resulting in an excess of diversity on the X (see also DISCUSSION).

The estimated time to the expansion in the past is 60,300 years with a 95% confidence interval of (13,800, 172,000) years, which is very similar to our previous estimate (LI and STEPHAN 2006). The strength of the expansion, measured by the ratio of the current size to the size before the expansion, is 5.0 (2.0, 12.0). The sex ratio ( $N_f/N_m$ ) is 1.8, but the LRT suggests that the number of females in the African population is not significantly larger than the number of males [ $P > 0.05$ ;  $\max\log(L) = -4756.6$  vs.  $-4754.0$ ; the critical value  $\zeta_{95\%} = 44.46$ ].

For the European population, the time of the out-of-Africa migration is 17,500 years, which is similar to our previous estimate from the X-linked data set (LI and STEPHAN 2006), and  $\hat{N}_{xEO} = 0.958 \times 10^6$ . We find  $\hat{\beta} = 0$  (0, 0.082), and the ratio of population sizes ( $N_{xEO}/N_{aEO}$ ) is 0.5625. The LRT suggests that there is a vanishingly small percentage of females in the European population [ $P < 0.01$ ;  $\max\log(L) = -12452.9$  vs.  $-12548.8$ ; the critical value  $\zeta_{99\%} = 8.45$ ]. The unrealistic estimate of the sex ratio suggests that the low diversity on the X chromosome in the European population cannot be explained by the bottleneck and the smaller effective population size for the X chromosome alone.

Compared to the autosomal diversity in the European population, the reduced diversity on the X chromosome could be due to purifying selection alone. POOL and AQUADRO (2006) found that purifying selection might at least be partly responsible for the occurrence of low-frequency variation in *D. melanogaster* populations. For this reason, following FU (1997) and SMITH and EYRE-WALKER (2002), we repeated the analyses disregarding the singletons (*i.e.*, the mutations carried by a single chromosome in the sample). In this case, we also have  $\hat{\beta} = 0$  (0, 0.31), which may suggest that purifying selection does not play a major role.

## DISCUSSION

**Different levels of nucleotide diversity between the X and the third chromosome and between ancestral and derived populations:** Our scan of sequence diversity on the third chromosome in an ancestral African and a derived European population of *D. melanogaster* shows a pattern similar to that already found by GLINKA *et al.* (2003) on the X chromosome in the same populations: the European population exhibits reduced levels of polymorphism. However, while the level of diversity drops to 35% on the X chromosome (relative to the ancestral one), the European autosome retains 62% of the ancestral  $\theta$ -value. Thus, the question arises what forces lead to these differences in reduction of polymorphism. In addition, the ancestral population shows a ratio of X-chromosomal to autosomal diversity of 0.90, which is significantly higher than the expecta-

tion of 0.75 under standard neutral conditions. Following the more severe drop-off of diversity on the X chromosome, this leads to a ratio of 0.49 in Europe that is significantly lower than the standard neutral expectation. Such disparities have already been reported in previous studies (ANDOLFATTO 2001; KAUER *et al.* 2002).

It is now well established that the reduction of variation in the derived population can be attributed mainly to a population size bottleneck during range expansion (HADDRILL *et al.* 2005b; OMETTO *et al.* 2005; LI and STEPHAN 2006). However, it appears that autosomes and sex chromosomes were affected differently during this colonization process. CHARLESWORTH (2001) suggested that an unequal sex ratio might explain these observations. If any of both sexes experiences a large variance in reproductive success (NUNNEY 1993) or if a substantial fraction of individuals fail to reproduce during their lifetime (CHARLESWORTH 2001), this will lead to a reduction of effective population size. Reproduction in natural populations of *Drosophila* is highly unlikely to occur by random mating, and in fact there is abundant evidence (CROW and MORTON 1955; BOULÉTREAU 1978; SOLLER *et al.* 1999) that sexual selection and environmental factors might lead to different population sizes for males and females. If these effects cause an unequal sex ratio, this will be reflected in the ratio of X-chromosomal to autosomal effective population sizes. Since males carry only one X chromosome as opposed to two in females, differences in male population size have a smaller effect on X-linked than on autosomal diversity. Therefore, unequal sex ratios will lead to a deviation from the standard expectation of 0.75 for ratios of diversity. A goal of this study was to obtain accurate estimates of the sex ratios in both populations, taking their demography into account.

### Demography and sex ratio in the African population:

Our estimation of the demographic history of the African population using the combined X-chromosomal and autosomal data set confirms the results of LI and STEPHAN (2006) where only X-linked data were used, although we find that the empirical distribution of  $\zeta$  does not follow a  $\chi^2$ -distribution (Figure 1), which was used in our previous analysis (LI and STEPHAN 2006). The most likely scenario for the African population is a population size expansion  $\sim 60,000$  years ago. The estimate of the sex ratio suggests that the female population size is 1.8 times larger than the male one in Africa. Although this difference is not significant, it may suggest that an unequal sex ratio played an important role in shaping X-chromosomal and autosomal diversity in Africa. CHARLESWORTH (2001) tried to explain the excess of female effective population size in the ancestral population: females are in good breeding condition in Africa, so there is little variance in reproductive success. Males, on the other hand, are subject to strong sexual selection that reduces the male effective population size (CROW and MORTON 1955; NUNNEY 1993).

In addition to our maximum-likelihood approach, an estimation of the sex ratio can be obtained directly from levels of diversity. If we take the ratio of  $\theta/K$  values as a proxy for the ratio of population sizes of X chromosomes and autosomes and plug them into Equation 1, we obtain a female/male ratio of 3.0 for the African population. This ratio is larger than the estimate (1.8) from our analysis. The difference between both values may be understood as follows. The populations are not in equilibrium. In such a case estimating population sizes from diversity (using Watterson's  $\theta$ ) will lead to an underestimation of effective population size. Since the X chromosome has a smaller population size than autosomes it will reach equilibrium faster after the expansion. Therefore the underestimation of population size will not be as extreme on the X as on the autosome. This leads to a bias toward higher female/male ratios. Even though we estimate an elevated population size for females, the population size of the X chromosome is still considerably smaller than that of autosomes. Following Equation 1 one would need a sevenfold excess of females to achieve equal population sizes for X chromosomes and autosomes, but our estimate is well below that value. Therefore the argument of a smaller X-chromosomal  $N_e$  still holds.

KAUER *et al.* (2002) used Equation 1 to test if unequal sex ratios could explain their data. But even when they assumed a 50-fold excess of female population size the X chromosome seemed too variable. This is because their ratio of microsatellite heterozygosity was 1.21, which is larger than the limiting value of 9/8 that can be achieved when male population size approaches zero. A reason for this high ratio could be a bias in mutation rate. If we leave  $\theta$  uncorrected for our SNP data set we find that the ratio equals 1.18 for the African population. This ratio corresponds well with the microsatellite data; so a biased mutation rate might indeed explain these findings.

Not only do we see differences in overall levels of diversity between X chromosomes and autosomes but there are also differences in the frequency spectrum.  $D/D_{\min}$  values are slightly (although not significantly) more negative on the X chromosome and the variance is reduced. Population size expansions are known to lead to an increase of low-frequency variants (*i.e.*, creating negative Tajima's  $D$  values) and reduce the variance of Tajima's  $D$  across the loci. The magnitude of this effect depends on the parameter values of the expansion. Although the population size expansion affects both the X chromosome and the autosome, the effect may not be the same, and the expansion scenario is different for the X chromosome and the autosome from a coalescent point of view. Since  $N_{xA0} < N_{aA0}$ , we have  $t_x = 0.035$  and  $t_a = 0.029$ . Coalescent simulations show that the expansion scenario for the X chromosome results in more negative values of Tajima's  $D$  and a lower variance than that for the autosome (results not shown). Therefore, our observation on  $D/D_{\min}$  supports the hypoth-

esis that the African population may have undergone a population size expansion.

$D/D_{\min}$  is not the only summary statistic that shows differences between chromosomes. The average  $Z_{nSS}$  is significantly lower on the X chromosome, implying that there is less LD. The variance is also lower on the X chromosome, and this difference is marginally significant ( $P = 0.06$ ). The effect of population size expansions on LD is very similar to that on the frequency spectrum. Average LD tends to get lower and the variance becomes smaller if the population underwent a size increase. Coalescent simulations show that expansions with X-chromosomal parameter values produce lower averages and smaller variances, in accordance with the trend we find in our data.

ANDOLFATTO (2001) suggested that the presence of common autosomal inversions in African *D. melanogaster* populations and the action of positive selection might play a role in reducing autosomal diversity compared to the X chromosome. If such chromosomal inversions are present at considerable frequencies, population recombination rates are lowered because inversions suppress crossing over when in a heterozygous state. The effect of genetic hitchhiking associated with positive selection will then be more pronounced on the autosomes compared to the X chromosome. If the action of positive selection was the main reason for the reduced variability on the autosome (relative to the X), then this should also be visible in the frequency spectrum. However, the observed patterns of  $D/D_{\min}$  argue against this hypothesis, as we see a larger proportion of low-frequency variation on the X chromosome compared to the autosome (Figure 2B). While we cannot exclude that inversions might affect patterns of nucleotide diversity in Africa, their relative contribution seems to be rather low compared to the effects of demography. Our analyses suggest that the population size expansion (~60,000 years ago) together with an unequal sex ratio can account for the genomewide patterns of polymorphism that we observe on the X chromosome and the autosome in the African population.

**Demography and sex ratio in the European population:** Our inference of the demography and sex ratio for the European population produced a surprising result. The most likely demographic scenario is a population bottleneck with subsequent expansion, but the estimated female/male ratio is zero, implying that an extremely small percentage of females are present in Europe. In such a case, we have  $N_f = 4/9N_x$ . We estimate that the current effective population for the X chromosome in the European population is  $0.958 \times 10^6$ , which suggests that the number of females is  $0.426 \times 10^6$ , whereas the number of males in the European population is much larger. Our estimate of the female/male ratio has an upper 95% confidence interval of 0.082. Therefore a 12-fold excess of male effective population size relative to females might be sufficient to explain the patterns we

observe. Such an extreme excess of males is highly unlikely in natural populations of *Drosophila*. On the one hand BOULÉTREAU (1978) showed that a large proportion of females fail to breed in European populations, which may effectively reduce their population size relative to males. On the other hand there is abundant sexual selection on males in natural populations that in turn reduces male effective population size (CROW and MORTON 1955; NUNNEY 1993). Depending on the relative magnitudes of these effects this might indeed lead to a male-biased sex ratio, but CHARLESWORTH (2001) showed that expectations of this bias are not as extreme as the ratios required to explain the data we observe. This suggests that the lower X-linked diversity cannot be explained by a biased sex ratio alone.

Previous work on non-African *D. melanogaster* (ANDOLFATTO 2001; KAUER *et al.* 2002) found higher ratios of X-chromosomal to autosomal diversity than we observed. However, since these analyses do not control for mutation rate, they may be biased (see above). Another study comparing X-chromosomal and autosomal polymorphism in non-African *D. simulans* (BEGUN and WHITLEY 2000) found exactly the same estimate we obtained in our European data set. This work provides a good comparison to our analysis, because levels of diversity were also corrected (by dividing by divergence). Mutational biases should therefore not affect the results.

While demography and sex ratio alone cannot account for the genomic polymorphism patterns observed they may still have played important roles in shaping polymorphism in Europe. Our analysis suggests that the sex ratio was inverted during the colonization process of Europe. As a consequence of this the X chromosome underwent a more severe bottleneck than the autosome. Population size reductions, much like expansions, have distinct effects on the frequency spectrum of polymorphisms and LD. In the case of a population size reduction the frequency spectrum tends to show an excess of intermediate-frequency polymorphisms. This leads to elevated Tajima's  $D$  values. In addition, the variance of Tajima's  $D$  tends to get larger. A similar pattern is created for the  $Z_{nSS}$  statistic. Average LD and its variance increase. It is important to note that these predictions are sufficiently understood only for simple models of population size reduction. The demographic history of the European population, however, seems to be rather complex (Figure 1b of LI and STEPHAN 2006). First, it was derived from an ancestral African population that was not in equilibrium (see above). Of course one could question if the Zimbabwe population actually reflects the true ancestral population. But recent studies have shown that there is a signal for expansion in the vast majority of African populations that have been surveyed (POOL and AQUADRO 2006). It is therefore very likely that the true ancestral population also showed this pattern. Second, the derived population experienced a population size expansion of its own,

subsequent to the colonization of Europe. It is difficult to assess how these different events will contribute to the overall patterns of polymorphism.

For average  $D/D_{min}$ , we find that it is nearly identical on the X chromosome and the autosome in Europe. Interestingly both values are very close to zero as in the case of a standard neutral population. However, since our estimation of the European demography rejects this model, this is most likely the result of a much more complex process. The variance of  $D/D_{min}$  shows a large difference between the chromosomes. The larger variance on the X chromosome may again be explained by the different times back to the bottleneck event (in the coalescent view). In the case of very recent bottlenecks slightly "older" events tend to cause more variance in  $D/D_{min}$ . A comparison of means and variances of  $Z_{nSS}$  between the chromosomes also shows a similar pattern. LD is higher on the X chromosome and has a larger variance. This is also expected under a simple bottleneck model. In a similar case, WALL *et al.* (2002) explained the patterns of LD they found in non-African *D. simulans* (which mimic our findings) with a simple bottleneck. Even though the European population has a complex demographic past the patterns of the frequency spectrum ( $D/D_{min}$ ) and LD agree well with the predictions of a simple population size reduction. In conclusion, the bottleneck seems to have been the dominant demographic force shaping the patterns of polymorphism we observe in Europe today.

However, the main question remains: What led to the extreme differences in average diversity between the X chromosome and the autosome? We have shown that demography and unequal sex ratio alone cannot account for these differences. BEGUN and WHITLEY (2000) propose that genetic hitchhiking caused by positive selection might have played a major role in reducing X-linked diversity in their *D. simulans* data, since it is thought that positive selection should affect X chromosomes more strongly than autosomes (AQUADRO *et al.* 1994). Theoretical work has shown that this claim holds if recombination occurs in both sexes (BETANCOURT *et al.* 2004). In the case of *Drosophila* where recombination occurs only in females, on the other hand, this effect is visible only if mutations are partially recessive (BETANCOURT *et al.* 2004). Thus, if a large fraction of advantageous mutations are indeed recessive (for instance, as suggested by ZEYL *et al.* 2003), the effect should also be visible in *Drosophila*.

Positive selection might also help explain the pattern we see for the frequency spectrum in the European population: a population size bottleneck creates an excess of intermediate-frequency variants and this effect is larger for the X chromosome. Genetic hitchhiking, on the other hand, tends to create low-frequency polymorphisms (BRAVERMAN *et al.* 1995) and also is supposed to influence the X chromosome to a larger extent (see above). If both forces act simultaneously their effects on the

average levels of Tajima's  $D$  might cancel out. At the same time this will result in a large variance of Tajima's  $D$ . Since we expect the effect of both forces to be more pronounced on the X chromosome, this might explain why the X chromosome harbors more loci that deviate from standard neutral expectations (32% of all loci containing polymorphism) than the autosome (only 12%) and, at the same time, average levels of  $D/D_{\min}$  are approximately equal. The effects on average levels of  $D/D_{\min}$  cancel out while the increases in variance add up.

The effect of hitchhiking on the genomewide averages of LD is complex and not well understood. PRZEWORSKI (2002) shows that recent selective sweeps can substantially increase levels of LD. However, the created signal disappears rapidly after fixation of the selected mutation. Furthermore, if there is recurrent positive selection in the population the overall effect of multiple sweeps might even lead to a slight decrease in LD. A more recent study (STEPHAN *et al.* 2006) shows that hitchhiking can destroy preexisting LD. Depending on the haplotype on which a beneficial mutation appears it will lead to either an increase or a decrease of already present LD. The average over all possibilities leads to a level of LD that will then be slightly lower after the sweep than it was before the emergence of the positively selected mutation. In summary, both studies suggest that the overall effect of hitchhiking is not one of an increase of LD as has been previously thought. On a genomewide level, hitchhiking rather tends to slightly decrease LD. It appears that the pattern of LD we observe in our data was therefore mainly shaped by demographic events as these leave a more pronounced signature. This confirms the conclusions made by HADDRILL *et al.* (2005b) concerning average levels of LD in the European population. The effects of selection on the variance of LD are clearer. STEPHAN *et al.* (2006) showed that recurrent hitchhiking increases the variance of LD. This increase in variance will be more pronounced on the X chromosome since we expect positive selection to be more prevalent there. Therefore, positive selection could have reinforced the differences in variance of LD between the two chromosomes created by the bottleneck.

In summary, the observed patterns of polymorphism provide evidence that population size bottlenecks and positive selection have acted simultaneously in the European population.

We thank K. Bhuiyan, A. Fabry, and A. Wilken for excellent technical assistance. We also thank J. Baines, J. Hermisson, P. Pennings, and P. Pfaffelhuber for constructive and helpful advice and discussions. We are grateful to the Volkswagen Foundation for financial support (grant no. I/78 815).

#### LITERATURE CITED

- AKEY, J. M., M. A. EBERLE, M. J. RIEDER, C. S. CARLSON, M. D. SHRIVER *et al.*, 2004 Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* **2**: e286.
- ANDOLFATTO, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**: 279–290.
- ANDOLFATTO, P., 2005 Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**: 1149–1152.
- AQUADRO, C. F., D. J. BEGUN and E. C. KINDAHL, 1994 Selection, recombination, and DNA polymorphism in *Drosophila*, pp. 46–56 in *Non-Neutral Evolution: Theories and Molecular Data*, edited by B. GOLDING, Chapman & Hall, New York.
- AULARD, S., J. R. DAVID and F. LEMEUNIER, 2002 Chromosomal inversion polymorphism in Afrotropical populations of *Drosophila melanogaster*. *Genet. Res.* **79**: 49–63.
- BAUDRY, E., B. VIGINIER and M. VEUILLE, 2004 Non-African populations of *Drosophila melanogaster* have a unique origin. *Mol. Biol. Evol.* **21**: 1482–1491.
- BEGUN, D. J., and P. WHITLEY, 2000 Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* **97**: 5960–5965.
- BETANCOURT, A. J., D. C. PRESGRAVES and W. J. SWANSON, 2002 A test for faster X evolution in *Drosophila*. *Mol. Biol. Evol.* **19**: 1816–1819.
- BETANCOURT, A. J., Y. KIM and H. A. ORR, 2004 A pseudohitchhiking model of X vs. autosomal diversity. *Genetics* **168**: 2261–2269.
- BOULÉTREAU, J., 1978 Ovarian activity and reproductive potential in a natural population of *Drosophila melanogaster*. *Oecologia* **35**: 319–342.
- BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY and W. STEPHAN, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**: 783–796.
- CHARLESWORTH, B., 2001 The effect of life-history and mode of inheritance on neutral genetic variability. *Genet. Res.* **77**: 153–166.
- CHARLESWORTH, B., J. A. COYNE and N. H. BARTON, 1987 The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* **130**: 113–146.
- CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- COMERON, J. M., M. KREITMAN and M. AGUADÉ, 1999 Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. *Genetics* **151**: 239–249.
- CROW, J. F., and N. E. MORTON, 1955 Measurement of gene-frequency drift in small populations. *Evolution* **5**: 202–214.
- DAVID, J. R., and P. CAPY, 1988 Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet.* **4**: 106–111.
- FU, Y. X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.
- GLINKA, S., L. OMETTO, S. MOUSSET, W. STEPHAN and D. DE LORENZO, 2003 Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics* **165**: 1269–1278.
- HADDRILL, P. R., B. CHARLESWORTH, D. L. HALLIGAN and P. ANDOLFATTO, 2005a Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content. *Genome Biol.* **6**: R67.
- HADDRILL, P. R., K. R. THORNTON, B. CHARLESWORTH and P. ANDOLFATTO, 2005b Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res.* **15**: 790–799.
- HEDRICK, P. W., 2000 *Genetics of Populations*. Jones & Bartlett, Sudbury, MA.
- HUTTER, S., A. J. VILELLA and J. ROZAS, 2006 Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinform.* **7**: 409.
- JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism*, edited by H. N. MUNRO. Academic Press, New York.
- 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.
- KELLY, J. K., 1997 A test of neutrality based on interlocus associations. *Genetics* **146**: 1197–1206.
- KLIMAN, R. M., and J. HEY, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol. Biol. Evol.* **10**: 1239–1258.

- LACHAISE, D., M. L. CARIOU, J. R. DAVID, F. LEMEUNIER, L. TSACAS *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol. Biol.* **22**: 159–225.
- LEFEVRE, G., 1976 A photographic representation and interpretation of the polytene chromosomes of *Drosophila melanogaster* salivary glands, pp. 32–66 in *The Genetics and Biology of Drosophila*, Vol. 1a, edited by M. ASHBURNER and E. NOVITSKI. Academic Press, New York.
- LI, H., and W. STEPHAN, 2006 Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet.* **2**: e166.
- LI, Y. J., Y. SATTA and N. TAKAHATA, 1999 Paleo-demography of the *Drosophila melanogaster* subgroup: application of the maximum likelihood method. *Genes Genet. Syst.* **74**: 117–127.
- MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**: 23–35.
- NUNNEY, L., 1993 The influence of mating system and overlapping generations on effective population size. *Evolution* **47**: 1329–1341.
- 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 positive selection in a European population of *Drosophila melanogaster*: multilocus pattern of variation and distance to coding regions. *Genetics* **167**: 1759–1766.
- PAWITAN, Y., 2001 *In All Likelihood: Statistical Modeling and Inference Using Likelihood*. Oxford University Press, New York.
- POOL, J. E., and C. F. AQUADRO, 2006 History and structure of sub-Saharan populations of *Drosophila melanogaster*. *Genetics* **174**: 915–929.
- PRZEWORSKI, M., 2002 The signature of positive selection at randomly chosen loci. *Genetics* **160**: 1179–1189.
- SCHAEFFER, S. W., 2002 Molecular population genetics of sequence length diversity in the *Adh* region of *Drosophila pseudoobscura*. *Genet. Res.* **80**: 163–175.
- SCHMID, K. J., S. RAMOS-ONSINS, H. RINGYS-BECKSTEIN, B. WEISSHAAR and T. MITCHELL-OLDS, 2005 A multilocus sequence survey in *Arabidopsis thaliana* reveals a genomewide departure from a neutral model of DNA sequence polymorphism. *Genetics* **169**: 1601–1615.
- SHAPIRO, J. A., W. HUANG, C. ZHANG, M. J. HUBISZ, J. LU *et al.*, 2007 Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl. Acad. Sci. USA* **104**: 2271–2276.
- SMITH, N. G., and A. EYRE-WALKER, 2002 Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022–1024.
- SOLLER, M., M. BOWNES and E. KUBLI, 1999 Control of oocyte maturation in sexually mature *Drosophila* females. *Dev. Biol.* **208**: 337–351.
- STEPHAN, W., Y. S. SONG and C. H. LANGLEY, 2006 The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics* **172**: 2647–2663.
- TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- WALL, J. D., P. ANDOLFATTO and M. PRZEWORSKI, 2002 Testing models of selection and demography in *Drosophila simulans*. *Genetics* **162**: 203–216.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- WESLEY, C. S., and W. F. EANES, 1994 Isolation and analysis of the breakpoint sequences of chromosome inversion In(3L)Payne in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **91**: 3132–3136.
- ZEYL, C., T. VANDERFORD and M. CARTER, 2003 An evolutionary advantage of haploidy in large yeast populations. *Science* **299**: 555–558.

Communicating editor: M. VEUILLE