The Impact of Founder Events on Chromosomal Variability in Multiply Mating Species

John E. Pool and Rasmus Nielsen

Department of Integrative Biology, University of California, Berkeley

In species with heterogametic males, the relative levels of X chromosome versus autosome diversity hold key information about the evolutionary forces at work in a population. It has been shown that population size changes alter the ratio of X linked to autosomal (X/A) variation, with population size reductions and recent bottlenecks leading to decreased X/A diversity ratios. Here we use theory and simulation to investigate a separate demographic effect—that of founder events involving multiply mated females—and find that it leads to much stronger reductions in X/A diversity ratios than are produced by simple population size changes. Investigating the potential of this process to account for sharply reduced X-linked diversity in European *Drosophila melanogaster*, we find that this model yields predictions that are compatible with the empirical data.

Under the simplest population genetic model, the ratio of X linked to autosomal variation (hereafter "X/A diversity ratio") would follow from the numbers of each chromosome present in a mating pair, with an expected value of 0.75. However, a variety of evolutionary forces ensure that this prediction is rarely observed in nature. First, the rate of mutation can differ between X-linked and autosomal nucleotides. In mammals, autosomal mutation rates are higher than X-linked rates, owing to a higher mutation rate in the male germ line (reviewed in Li et al. 2002). In *Drosophila*, mutation rates are more similar (Bauer and Aquadro 1997), although recent evidence has pointed to a slightly higher X-linked rate (Begun et al. 2007).

Demographic factors also have a strong potential to influence X/A diversity ratios. Simulation studies first suggested that mitochondrial (e.g., Fay and Wu 1999) and X-linked (Wall et al. 2002) loci are affected differently by population bottlenecks than are autosomal loci. Population size reductions and recent bottlenecks are predicted to reduce X/A diversity ratios, whereas population expansion should yield the opposite effect (Pool and Nielsen 2007). Chromosomal diversity ratios are also impacted by the effective numbers of females and males in a population (which may be altered by sex-specific variances in reproductive success), with an excess of females elevating X-linked diversity relative to autosomal diversity and an excess of males giving a lower X/A diversity ratio (Caballero 1995; Charlesworth 2001).

The impact of natural selection on X/A diversity ratios depends strongly upon the selection and dominance coefficients of the relevant mutations—this is particularly because the X chromosome's male hemizygosity makes recessive mutations immediately "visible" to selection. Background selection (the effect of linkage to deleterious mutations) has a stronger effect on diversity when deleterious mutations can drift to appreciable frequency. Recessive deleterious mutations are more likely to persist on the autosomes, where they are masked from selection in the heterozygous state. Therefore, background selection is predicted to reduce autosomal diversity more than X-linked diversity and to elevate X/A diversity ratios (Charlesworth et al. 1993; Charlesworth 1994, 1996).

Hitchhiking of recessive beneficial alleles is also predicted to be more efficient on the X chromosome (Avery 1984; Charlesworth et al. 1987; Aquadro et al. 1994), and positive selection may therefore reduce X/A diversity ratios. However, in the case of *Drosophila*-like recombination (none in males), the opposite is true if beneficial mutations are mostly dominant on average: positive selection will lead to a greater reduction in autosomal diversity, inflating X/A diversity ratios (Betancourt et al. 2004). Hitchhiking may also be less efficient on the X chromosome if beneficial alleles are derived from standing variation as opposed to new mutations (Orr and Betancourt 2001), which may be particularly relevant for populations entering new environments.

Among the best studied species in terms of X-linked and autosomal diversity is *Drosophila melanogaster*. This species includes both ancestral range populations from sub-Saharan Africa and "cosmopolitan" populations, which expanded into Europe and Asia roughly 10,000 years ago (Lachaise et al. 1988) and suffered a reduction in genetic diversity (Begun and Aquadro 1993). Large multilocus studies using both microsatellites (Kauer et al. 2003) and DNA sequence polymorphism (Hutter et al. 2007) provided further evidence for a genetic bottleneck in the history of cosmopolitan samples (from Europe), but found these populations to have much stronger reductions in X-linked diversity than autosomal diversity (relative to a sub-Saharan population). Hutter et al. (2007) estimated demographic parameters for a population bottleneck model and allowed differing sex ratios for the 2 populations but concluded that demographic factors alone could not account for the observed X-linked and autosomal diversity reductions and argued that positive selection in the European population must be invoked.

One limitation of most demographic inference involving recently founded populations is the assumption of a simple bottleneck model with instantaneous population size reduction and subsequent recovery (fig. 1*A*). Although this model may be useful in capturing many features of empirical data, it is also worth considering ways in which the actual colonization process might have differed from this scenario. For example, Edmonds et al. (2004) used spatially explicit simulations to show that mutations at the leading edge of a range expansion can reach higher frequencies than would otherwise be expected.

Key words: founder events, multiple mating, X-linked variation, autosomal variation, *Drosophila melanogaster*.

E-mail: jpool@berkeley.edu.

Mol. Biol. Evol. 25(8):1728–1736. 2008
doi:10.1093/molbev/msn124
Advance Access publication May 28, 2008

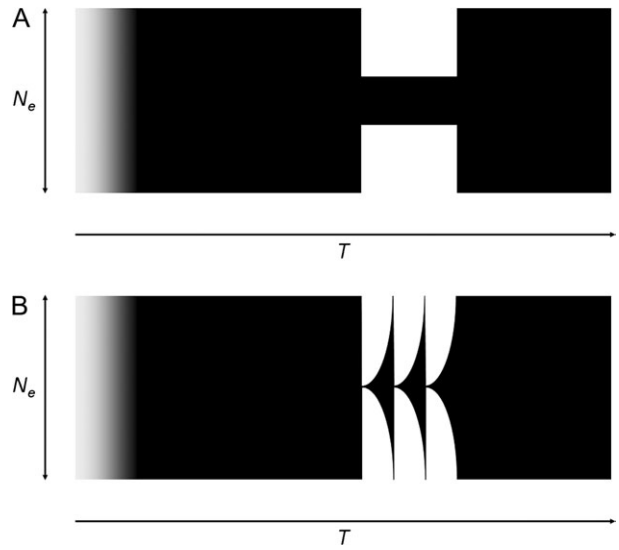

FIG. 1.—Illustration of the population bottleneck (*A*) and serial founder event (*B*) models, showing effective population size (N_e) over time (*T*).

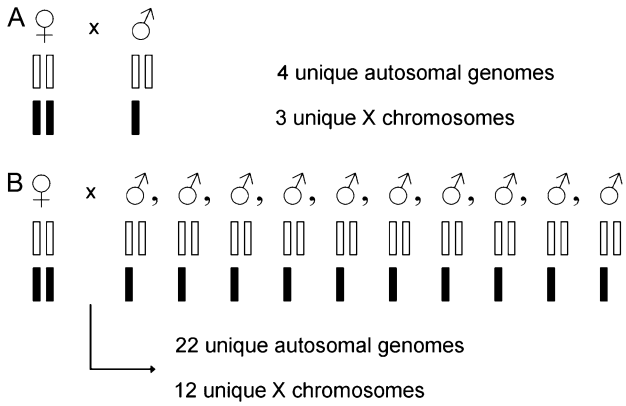

FIG. 2.—The unique X chromosomes and autosomal genomes introduced into a new population by a singly mated founder female (*A*) and by a founder female that had mated with 10 males in the ancestral population (*B*).

Here we focus on another departure from the traditional bottleneck model. A typical "3-epoch" size change model generally requires a large number of founders to leave the ancestral population simultaneously. For example, a model inferred from X-linked polymorphism data for *D. melanogaster* (Thornton and Andolfatto 2006) postulates 36,000 individuals splitting off from an ancestral population of size 2,400,000. In contrast, we examine a serial founder event model in which a small number of multiply mated females leave the ancestral range to found a new population, and this process may be repeated several times in the formation of the new population (fig. 1*B*).

Drosophila melanogaster females are known to mate multiply. Early genetic studies using allozyme markers found evidence for multiple paternity among the offspring from about half of wild-caught females (Milkman and Zeitler 1974; Ochando et al. 1996). Studies using more variable markers (microsatellites) have found higher levels of multiple mating. Harshman and Clark (1998) found that 16 of 19 wild-caught females produced progeny from 2 or more males, whereas Imhof et al. (1998) genotyped a larger number of offspring from 4 wild caught females and found that they produced progeny from 4 to 6 males.

Within an equilibrium population, multiple mating will not necessarily alter X/A diversity ratios (unless the mating system leads to sex-specific differences in reproductive success). However, we suggest that a founder event involving a small number of multiply mated females can give rise to a new population with a significantly lower X/A diversity ratio than the ancestral population. Due to multiple mating, there are effectively more male than female founders of the new population, and each of these males will contribute 2 unique autosomal genomes but only 1 X chromosome to the new population's gene pool (fig. 2). This excess of male founders may lead to reduced X-linked diversity relative to autosomal levels. A similar explanation was offered to account for low X-linked diversity in a recently founded population of *Drosophila pseudoobscura* (Reiland et al. 2002). Curiously, this model has not been suggested in the case of *D. melanogaster*, and its impact on X/A diversity ratios remains unclear.

Here we investigate the effect of founder events with multiple mating on X-linked and autosomal diversity. We find that demographic histories involving such events can produce considerably lower X/A diversity ratios than simple population size reduction or bottleneck models and that some parameter combinations can yield X-linked and autosomal diversity reductions similar to those observed in European *D. melanogaster*.

Methods

The Model

We propose a model in which *f* females, each having mated with *m* unique males, leave the ancestral population and found a separate (recently derived) population. A total of *b* sequential founder events occur at *T* time units before the present, where time is measured in number of generations divided by $2N_e$ of the ancestral range population, $2N_{anc}$. We allow the recently derived population to have a different N_e than the ancestral range population, such that $N_{der} = aN_{anc}$. Here we use *a* to indicate the ratio of autosomal effective population sizes.

We assume that population size recovery from each founder event is sufficiently rapid that drift and mutation during this period can be ignored, and we assume that founder events happen in close succession, such that drift and mutation between them can be ignored as well. For this model, the expected coalescence time between a pair of chromosomes is

$$E[C] = \int_0^T \frac{T}{ah}\, e^{-T/ah} dt + e^{-T/ah}(T + hP) = ah\ (1 - e^{-T/ah}(1 - P/a)), \quad (1)$$

where *P* is the probability of escape from the founder events (i.e., the probability that a pair of alleles do not coalesce during the founder events) and *h* is an inheritance scalar

equal to 1 for autosomal loci and three-fourth for X-linked loci under the assumption of equal sex ratios. Each founder female has the same probability of being chosen as the mother of a particular allele/lineage, and each of this female's mates has an equal probability of being chosen as the father. For X-linked loci, the escape probability is given by

$$P_X = \prod_1^b \left(1 - \left(\frac{2}{9f} + \frac{1}{9fm}\right)\right). \quad (2)$$

And for autosomal loci, it is

$$P_A = \prod_1^b \left(1 - \left(\frac{1}{8f} + \frac{1}{8fm}\right)\right). \quad (3)$$

In the presence of unequal sex ratios, the relationship between chromosomal effective population sizes is defined by

$$r = \frac{N_X}{N_A} = \frac{9(1+s)}{8(2+s)}, \quad (4)$$

where s is the ratio of females to males. Allowing the sex ratio to differ before and after the founder events, we introduce the notation r_{anc} and r_{der}, to denote the ratio of effective population sizes in the ancestral range and recently derived populations, respectively. The expected coalescence times for X-linked and autosomal loci in the ancestral range and recently derived populations (assuming demographic equilibrium for the former) then become

$$\begin{aligned} E[C_{A,anc}] &= 1, \\ E[C_{X,anc}] &= r_{anc}, \\ E[C_{A,der}] &= a - e^{-T/a}(a - P_A), \\ E[C_{X,der}] &= ar_{der} - e^{-T/ar_{der}}(ar_{der} - P_X r_{anc}). \end{aligned} \quad (5)$$

We are primarily interested in comparing the expected nucleotide diversity between the recently derived and ancestral range populations. Of course, the expected nucleotide diversity is proportional to the expected coalescence time for a pair of alleles (e.g., Hudson 1983) so that ratios of expected nucleotide diversity can be calculated directly as the ratios of expected coalescence times.

Comparing Theoretical Predictions to Empirical Data

Predictions from the serial founder event model were compared with the empirical data from *D. melanogaster* analyzed by Hutter et al. (2007). That study involved the collection of DNA sequence polymorphism data at 53 autosomal loci in a sub-Saharan (ancestral range) population from Zimbabwe compared with 378 loci in a cosmopolitan population from The Netherlands (a cosmopolitan, recently derived sample). Nucleotide diversity (π) at these loci was compared against the X-linked loci studied by Glinka et al. (2003) and Ometto et al. (2005), of which 246 were sequenced in both of the above population samples. We confine our analysis to these data sets because they involve the same samples and similar locus characteristics. For our purposes, the European sample is also preferable over studies involving *D. melanogaster* from North America, where recent sub-Saharan admixture has altered levels of diversity (Caracristi and Schlötterer 2003; Haddrill et al. 2005).

We examined predictions for a maximum of 5 sequential founder events, each involving 1–8 females (f). The number of females was allowed to differ between sequential founder events, but the number of males per female (m) was fixed at 1, 3, or 5. We investigated founder events occurring $t = 0.005$, 0.01, and 0.02 ancestral, autosomal coalescent units before the present, and we studied both the case of equal population sizes before and after the founder events ($a = 1$), and the case where the recently derived population's size only recovers to one-tenth of the ancestral size ($a = 0.1$).

We studied 2 contrasting models of sex ratios in the recently derived and ancestral range populations. First, we examined the case of equal numbers of females and males in both populations ($N_f = N_m$). In this model, we assumed that the difference between X-linked and autosomal nucleotide diversity in the sub-Saharan population (0.0116 vs. 0.0104) resulted entirely from a higher X-linked mutation rate. Note that a slightly higher X-linked mutation rate is consistent with recent analyses (Begun et al. 2007; Hutter et al. 2007). Second, we examined a model of sex ratio reversal in which the ancestral range (sub-Saharan) population has an excess of females, whereas the post-founder event, cosmopolitan population has an excess of males, as suggested by Charlesworth (2001). Here, the sub-Saharan sex ratio was estimated by assuming that the ratio of X linked to autosomal interspecific divergence rates in the Hutter et al. (2007) data set is proportional to the relative X-linked and autosomal mutation rates. To obtain the relative effective population sizes of the X chromosome and autosomes in the sub-Saharan population, we estimate it as:

$$\hat{r}_{anc} = \frac{N_X}{N_A} = \frac{\pi_X/K_X}{\pi_A/K_A} = \frac{0.0116/0.684}{0.0104/0.515} = 0.838, \quad (6)$$

where K_x and K_A are the number of substitutions per site between *D. melanogaster* and *Drosophila simulans* for X-linked and autosomal loci, respectively (Hutter et al. 2007). Solving from equation (4), this corresponds to an ancestral sex ratio of $s = 1.92$ or about 2 females per male. For the recently derived population, we chose an arbitrary value of $N_x/N_A = 0.625$, which corresponds to a sex ratio of $s = 0.25$ (4 males per female).

We used theoretical predictions to identify parameter combinations potentially compatible with the empirical data, thus reducing the parameter space to be explored by subsequent simulation. First, we obtained 90% bootstrap confidence intervals (CIs) for the cosmopolitan sample's X-linked and autosomal diversity reductions by sampling loci (with replacement) from the empirical data to produce 1,000,000 resampled data sets. Then, for all possible combinations of the parameters and models listed above, we generated theoretical predictions for the X-linked and autosomal diversity reductions (as defined above). If a parameter set's predicted X-linked and autosomal diversity reductions

both fell within the 90% bootstrap CIs of the empirical data, that model was retained for further examination (via simulation); the remaining parameter sets were discarded.

In a similar fashion, theoretical predictions were also examined for population size reduction and population bottleneck models, using equations for X-linked and autosomal diversity reductions provided by Pool and Nielsen (2007). For the reduction (2-epoch) model, the principal parameters are T, the time (in ancestral, autosomal coalescent units) since the size change, and a, the population size change factor. Predictions were examined for all combinations of $T = \{0.00001, 0.00002, 0.00003, \ldots, 0.16\}$ and $a = \left\{\frac{1}{2^1}, \frac{1}{2^2}, \frac{1}{2^3}, \ldots, \frac{1}{2^{12}}\right\}$. The bottleneck model includes parameters T (time since the bottleneck ended), d (duration of the bottleneck), and c (the ratio of the bottlenecked size to the present population size). Predictions were examined for all combinations of $T = \{0.005, 0.01, 0.02, \ldots, 0.16\}$, $d = \{0.00001, 0.00002, 0.00003, \ldots, T - 0.00001\}$, and $c = \left\{\frac{1}{2^1}, \frac{1}{2^2}, \frac{1}{2^3}, \ldots, \frac{1}{2^{15}}\right\}$. As with the founder event analysis, the bottleneck model was investigated for the cases of equal and unequal pre- and postbottleneck N_e ($a = 1$ and $a = 0.1$). The reduction and bottleneck scenarios were also investigated under both the equal sex ratio and the sex ratio reversal models described above.

Coalescent Simulations

A coalescent simulation program was written to generate 2 population data sets under the multiple mating, serial founder event model. The simulations follow a standard coalescent process (Kingman 1982a, 1982b; Hudson 1983) from the present time back to the founder events and also from the population split time back to the common ancestor of the sampled chromosomes. Only the founder event generations are modeled differently. For each founder event, the numbers of X chromosomes and autosomes present in female and male founders are simply obtained based on f and m. Each of the remaining lineages in the recently derived population first randomly choose whether they came from a female or male founder (autosomes come from a female with probability 1/2, X chromosomes with probability 2/3), then randomly choose a specific founder chromosome. When 2 or more lineages choose the same founder chromosome, coalescence occurs (multiple coalescent events are permitted). Once all founder events have been simulated, the populations are merged and the standard coalescence process resumes until a common ancestor is reached.

Data sets were simulated to resemble the one analyzed by Hutter et al. (2007). A total of 246 X-linked loci were simulated for both populations (each with sample size $n = 12$). For autosomal loci, 53 were simulated for both populations (with $n = 8$ for the ancestral range sample and $n = 11$ for the recently derived sample) along with 325 loci that were simulated for the recently derived sample only. All simulated loci were 500 bp in length (close to the average length for both the X-linked and autosomal data sets), and within the X-linked and autosomal data sets, each was given the same mutation rate (θ was calibrated according to empirical Zimbabwe X-linked and autosomal diversity). Following Hutter et al. (2007), no intralocus recombination within these short fragments was simulated. For each parameter combination, 10,000 replicate data sets were produced.

Results

Predictions of the Model

The primary goal of the present study was to examine the effect of multiple mating during founder events on X-linked and autosomal diversity. Above, theoretical models were described that give the predicted population diversity ratio (recently founded population/ancestral range population) under a serial founder event model with multiple mating, for X-linked and autosomal loci. Below, we refer to these quantities as X-linked retention (the proportion of ancestral X-linked nucleotide diversity retained in the recently founded population, X_{ret}) and autosomal retention (A_{ret}). It is also informative to consider a "ratio of ratios" (R_R), which can be defined either as the ratio of the X-linked and autosomal population diversity ratios, $R_R = \frac{X_{ret}}{A_{ret}} = \frac{\theta_{X\cdot Der}/\theta_{X\cdot Anc}}{\theta_{A\cdot Der}/\theta_{A\cdot Anc}}$, or as the ratio of the X/A diversity ratios from the recently derived and ancestral range populations, $R_R = \frac{\theta_{X\cdot Der}/\theta_{A\cdot Der}}{\theta_{X\cdot Anc}/\theta_{A\cdot Anc}}$. Of particular interest are demographic scenarios that give a low value of X_{ret} (a severe reduction in X-linked diversity) but a relatively higher A_{ret} (a milder reduction in autosomal diversity) and therefore yield low values of R_R.

We hypothesized that founder events involving multiply mated females would lead to greater contrast between X-linked and autosomal diversity reductions, compared with models involving simple population size changes. Figure 3 confirms this prediction. Here, founder event models with varying degrees of multiple mating (m) were calibrated to have the same autosomal diversity reduction (A_{ret}) by adjusting the numbers of founder females. Models with higher m achieve greater X-linked diversity reductions (lower X_{ret}) with the same A_{ret} and will thus have lower R_R. The largest drop in X_{ret} is between $m = 1$ and $m = 2$, and a moderate level of multiple mating is sufficient to have a strong effect on X-linked versus autosomal diversity reductions.

Figure 3 also includes population size reduction and population bottleneck models that produce similar autosomal diversity reductions (A_{ret}). Other reduction and bottleneck parameter combinations with comparable A_{ret} predictions yield very similar X_{ret} as the models shown (data not shown). As indicated by figure 3, founder event models with multiple mating produce lower R_R than the reduction and bottleneck models and therefore have more potential to account for low X/A diversity ratios in recently founded populations.

Comparison to Empirical Data

One species with recently founded populations that show reduced X/A diversity ratios is *D. melanogaster*. Large, multilocus studies have shown that cosmopolitan populations of *D. melanogaster* have a stronger reduction

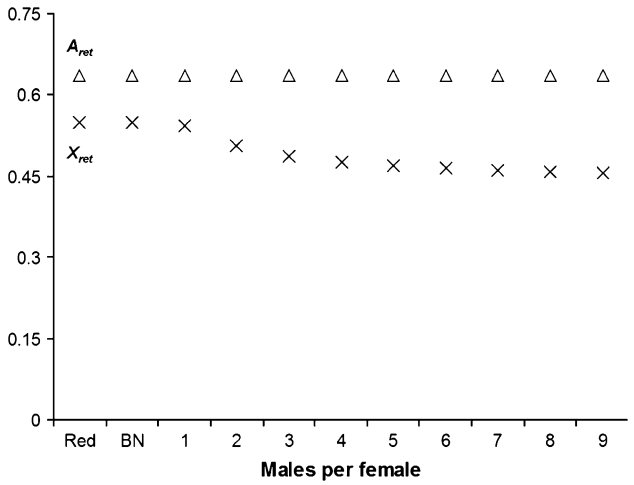

FIG. 3.—Proportion of X-linked and autosomal diversity retained (X_{ret} and A_{ret}) under sample population size reduction (Red), population bottleneck (BN), and multiple mating founder event models. The reduction model involved a decrease to 0.0215 of the initial N_e at $T = 0.005$ (ancestral, autosomal) coalescent units before the present. The bottleneck model began with a reduction to 0.00452 of the initial N_e, lasting 0.001 coalescent units and followed by a recovery to the initial size 0.005 coalescent units before the present (adjusting bottleneck severity and duration did not alter results). The founder event models included 5 founder events occurring 0.005 coalescent units ago, equal population sizes and sex ratios, and numbers of founder females calibrated to give equal A_{ret} for each level of multiple mating ($m = 1$ through $m = 9$).

in X-linked than autosomal diversity, relative to ancestral range populations from sub-Saharan Africa (Kauer et al. 2003; Hutter et al. 2007). As a first step toward identifying demographic parameter sets that are compatible with empirical data, we generated 90% bootstrap CIs for X_{ret} and A_{ret}, based on resampling from the Hutter et al. (2007) multilocus sequence data set. The 90% CIs obtained were $X_{ret} = 0.413$ [0.377, 0.449] and $A_{ret} = 0.636$ [0.554, 0.741]. Theoretical predictions for X_{ret} and A_{ret} under a given demographic model were then generated and compared against the above CIs. In practice, these criteria screen for parameter sets that fall within the lower 95% confidence limit for X_{ret} and the upper 95% confidence limit for A_{ret} because no otherwise acceptable parameter sets were rejected due to low X_{ret} or high A_{ret}.

Analysis of theoretical predictions was also performed for population size reduction and population bottleneck models. For the reduction model, none of the of 360,000 different parameter settings analyzed yielded X_{ret} and A_{ret} within the bootstrap CIs. The same was true for the bottleneck model: none of the 1,632,636 parameter combinations examined gave expected values within the CIs. Thus, in agreement with the conclusions of Hutter et al. (2007), we find that demographic models involving simple population size changes (with or without population differences in sex ratio) are unable to account for the X-linked and autosomal diversity reductions observed in cosmopolitan *D. melanogaster*.

In contrast, for the founder event model with multiple mating, a number of parameter combinations (1,952 of 46,296) provided expected values within the bootstrap CIs for X_{ret} and A_{ret}. These parameter sets represent founder event models that might be compatible with the empirical data and included models with (1,157) and without (795) a sex ratio reversal (excess females in Africa, excess males in Europe). The high proportion of rejected parameter combinations is not surprising as most models will leave either too little or too much diversity on both the X chromosome and the autosomes. For each of the 1,952 parameter sets not rejected in the above step, replicate data sets were then simulated (as described in the Methods) to resemble the X-linked and autosomal sequence polymorphism data analyzed by Hutter et al. (2007). In each case, mean simulated values of X_{ret} and A_{ret} were found to agree very closely with theoretical predictions (data not shown).

As shown in table 1, the lowest mean R_R values (equal to X_{ret}/A_{ret}) from the simulated parameter sets were 0.669 (for the sex ratio reversal model) and 0.671 (for the equal sex ratio model), both slightly less extreme than the empirical value (0.649). However, these simulated values represent averages from 10,000 replicate data sets, and even though each data set consists of several hundred loci, there is considerable variance among replicates. To account for this uncertainty, we calculated the R_R P value, defined as the proportion of simulated data sets giving a lower R_R than the empirical value. The above-mentioned parameter sets had R_R P values of 0.414 (sex ratio reversal model) and 0.413 (equal sex ratio model), indicating that they generate more extreme R_R values than the empirical data over 40% of the time. Parameter sets with the highest R_R P values included models with and without a sex ratio reversal and models with equal and unequal population sizes (table 1). A total of 1,802 parameter combinations yielded R_R P values greater than 0.05 (supplementary table S1, Supplementary Material online) and thus cannot be rejected as demographic hypotheses for European *D. melanogaster* based on this aspect of the data.

All 1,714 parameter combinations with multiple mating ($m = 3$ or $m = 5$) that matched theoretical predictions for X_{ret} and A_{ret} also had simulated R_R P values greater than 0.05 (table 2). Of the much smaller number of parameter sets without multiple mating ($m = 1$) that matched the theoretical criteria, the only ones with R_R $P > 0.05$ involved sex ratio reversal, a smaller N_e in the founded population, and t greater than 0.05 (table 2). The highest R_R P value for models without multiple mating was 0.0802, and the lowest average R_R from these simulations was 0.768. Thus, founder event models without multiple mating may have some weak potential to account for the observed X_{ret} and A_{ret} from cosmopolitan *D. melanogaster*, but the parameter space is relatively limited, and these models are still fairly unlikely to account for the empirical data.

Although all the simulated parameter sets had theoretically predicted X_{ret} and A_{ret} within the 90% CIs of the empirical data, those with the lowest R_R P values tended to have slightly lower X_{ret} and A_{ret} values than the empirical data. To verify that these models can account for both the low X_{ret} and the relatively higher A_{ret} observed in *D. melanogaster*, we also monitored the proportion of simulated replicates that gave both a lower X_{ret} and a greater A_{ret} than the empirical data. This combined proportion (CombPr) is not a standard P value; in the best case, if the empirical X_{ret} and A_{ret} represented the median simulated values from

Table 1
Founder Events Models with Highest R_R *P* Values

m	*T*	*a*	r_{der}	*f*	X_{ret}	A_{ret}	R_R	R_R *P*	CombPr
5	0.005	1	0.625	{5, 3, 1, 1, 1}	0.380	0.568	0.669	0.420	0.0952
5	0.01	1	0.625	{4, 3, 1, 1, 1}	0.378	0.565	0.670	0.414	0.0852
5	0.005	0.1	0.625	{8, 8, 1, 1, 1}	0.380	0.567	0.671	0.414	0.0945
5	0.005	1	0.75	{5, 3, 1, 1, 1}	0.381	0.569	0.671	0.413	0.0965
5	0.005	0.1	0.625	{8, 7, 1, 1, 1}	0.379	0.566	0.670	0.412	0.0957
5	0.005	0.1	0.625	{4, 1, 1, 1}	0.380	0.567	0.671	0.412	0.0918
5	0.01	0.1	0.625	{1, 1, 1}	0.379	0.566	0.672	0.409	0.0962
5	0.02	1	0.625	{8, 2, 1, 1, 1}	0.380	0.567	0.672	0.408	0.0923
5	0.005	1	0.625	{2, 1, 1, 1}	0.381	0.570	0.671	0.408	0.0996
5	0.005	1	0.625	{6, 3, 1, 1, 1}	0.383	0.571	0.672	0.405	0.1032
5	0.01	1	0.625	{5, 3, 1, 1, 1}	0.383	0.570	0.673	0.404	0.0983
5	0.005	1	0.75	{2, 1, 1, 1}	0.383	0.570	0.673	0.402	0.0943
5	0.005	1	0.625	{4, 4, 1, 1, 1}	0.383	0.571	0.673	0.402	0.1061
5	0.005	0.1	0.625	{3, 2, 2, 1, 1}	0.379	0.564	0.673	0.402	0.0850
5	0.005	1	0.625	{7, 3, 1, 1, 1}	0.385	0.574	0.672	0.401	0.1083
5	0.01	1	0.75	{4, 3, 1, 1, 1}	0.380	0.566	0.673	0.399	0.0882
5	0.02	1	0.625	{4, 3, 1, 1, 1}	0.384	0.571	0.674	0.396	0.1041
5	0.005	1	0.75	{6, 3, 1, 1, 1}	0.383	0.570	0.673	0.395	0.0978
5	0.01	1	0.625	{2, 1, 1, 1}	0.385	0.571	0.675	0.394	0.1007
5	0.005	0.1	0.625	{5, 1, 1, 1}	0.384	0.572	0.673	0.394	0.1035
5	0.01	0.1	0.625	{4, 3, 2, 1, 1}	0.379	0.563	0.675	0.393	0.0796
5	0.005	1	0.75	{4, 4, 1, 1, 1}	0.384	0.571	0.674	0.393	0.1011
5	0.01	1	0.625	{6, 3, 1, 1, 1}	0.386	0.574	0.674	0.392	0.1087
5	0.005	0.1	0.75	{6, 6, 1, 1, 1}	0.378	0.560	0.675	0.390	0.0790

NOTE.—*m*, males per female; *t*, time since founder events (ancestral, autosomal coalescent units); *a*, ratio of population sizes (N_{der}/N_{anc}); r_{der}, ratio of X linked to autosomal population size in the recently derived population (where 0.75, r_{anc} is also 0.75; where 0.625, r_{anc} is 0.838; see Methods); *f*, number of females in each successive founder event; X_{ret} and A_{ret}, average proportion of ancestral diversity retained for X-linked and autosomal loci; R_R, ratio of ratios (X_{ret}/A_{ret}); R_R *P*, proportion of simulated replicates with simulated R_R less than empirical R_R; and CombPr, proportion of simulated replicates with X_{ret} less than empirical value and A_{ret} greater than empirical value. Supplementary table S1 (Supplementary Material online) provides a continuation of this list.

Table 2
Breakdown of the Numbers of Parameter Sets with Simulated R_R *P* Value > 0.05

	$a = 1$; equal sex ratios			$a = 0.1$; equal sex ratios		
	$T = 0.005$	$T = 0.01$	$T = 0.02$	$T = 0.005$	$T = 0.01$	$T = 0.02$
$m = 1$	0 (12)	0 (7)	0 (3)	0 (12)	0 (10)	0 (3)
$m = 3$	45 (45)	43 (43)	39 (39)	71 (71)	95 (95)	85 (85)
$m = 5$	42 (42)	41 (41)	41 (41)	55 (55)	80 (80)	111 (111)
	$a = 1$; sex ratio reversal			$a = 0.1$; sex ratio reversal		
	$T = 0.005$	$T = 0.01$	$T = 0.02$	$T = 0.005$	$T = 0.01$	$T = 0.02$
$m = 1$	0 (13)	0 (13)	0 (9)	0 (29)	20 (52)	68 (77)
$m = 3$	47 (47)	46 (46)	45 (45)	83 (83)	116 (116)	171 (171)
$m = 5$	43 (43)	43 (43)	39 (39)	57 (57)	103 (103)	171 (171)

NOTE.—Numbers in parentheses indicate parameter sets satisfying theoretical predictions for X_{ret} and A_{ret}; numbers outside parenthesis indicate how many of those parameter sets had simulated R_R *P* values > 0.05; *m* indicates number of males per founder female; *T* is the time since the founder events in coalescent units scaled by the ancestral autosomal N_e; and *a* is the ratio of population sizes (N_{der}/N_{anc}).

a given demographic model, the expected value of CombPr would be 0.25. The highest CombPr values observed for simulated parameter sets were 0.130 (sex ratio reversal model) and 0.135 (equal sex ratio model), and many founder event models with multiple mating were found to generate the contrasting X-linked and autosomal diversity reductions observed in cosmopolitan *D. melanogaster* with appreciable frequency (supplementary table S1, Supplementary Material online).

Discussion

Above, we have demonstrated that founder events involving multiply mated females can have a considerably stronger effect on X/A diversity ratios than simple population size changes, giving rise to recently derived populations with disproportionately reduced X-linked diversity. This demographic effect may be an important determinant of chromosomal diversity in new and geographically expanding populations of insects that mate multiply, and perhaps also for plant species that have sex chromosomes, because pollen may travel to a new population more frequently than seeds.

Although we focus on the X chromosome and autosomes, this process should also influence diversity comparisons involving uniparentally inherited markers. Due to the greater male contribution to genetic diversity in a population founded by multiply mated females, relatively more

Y-linked than mitochondrial diversity should be retained (although considerable variance may surround these expectations due to the lack of recombination on these chromosomes). This prediction cannot be tested in *D. melanogaster* because mitochondrial DNA is not expected to behave as a selectively neutral marker in this species: a very recent global replacement of maternally inherited *Wolbachia* endosymbionts (Riegler et al. 2005) implies a similar history for mitochondrial DNA in *D. melanogaster*.

In the genus *Drosophila*, X-linked and autosomal diversity data have been collected for ancestral range and recently derived populations of 4 species: *D. melanogaster* (e.g., Kauer et al. 2003; Hutter et al. 2007), *D. simulans* (Andolfatto 2001; Schöfl et al. 2006), *D. pseudoobscura* (Reiland et al. 2002), and *Drosophila subobscura* (Pascual et al. 2007). In each case, the recently derived population shows a lower X/A diversity ratio than the ancestral range population (Pool and Nielsen 2007). Population size changes might account for at least part of each population difference (Wall et al. 2002; Pool and Nielsen 2007). However, at least in the case of *D. melanogaster*, this factor does not seem to fully account for the recently derived population's reduced X-linked diversity (Hutter et al. 2007; this analysis).

Here we have shown that demographic models are capable of accounting for the differing diversity reductions of X-linked and autosomal loci in cosmopolitan *D. melanogaster*. Although previous studies (e.g., Hutter et al. 2007) were correct in concluding that a simple bottleneck model (with or without a change in sex ratio) is insufficient to produce this pattern, we find that sequential founder events involving multiply mated females can replicate the observed diversity reductions rather closely.

The serial founder event model with multiple mating is distinct from the sex ratio reversal hypothesis proposed by Charlesworth (2001) to account for earlier data in *D. melanogaster*. Charlesworth (2001) suggested that the relatively high X/A diversity ratio of sub-Saharan populations could result from a smaller effective number of males (due to sexual selection), whereas the lower X/A diversity ratio of cosmopolitan samples could result from a relatively small effective number of females (with most females in poor breeding condition at any given time, due to less favorable environments). We investigated population size reduction, population bottleneck, and multiple mating founder event models, both with and without sex ratio reversals. Even with sex ratio reversal, we could not identify any models involving simple population size changes (reductions or bottlenecks) that were compatible with the empirical data. In contrast to our multiple mating founder event model, which will generate the lowest R_R values for recent founder events, sex ratio reversals may yield a greater effect for more ancient population divergences (Charlesworth 2001). We analyzed reductions and bottlenecks up to 0.16 coalescent units in the past, but even population split times of this age are probably inconsistent with the paucity of unique variation (putative "neomutations") found in cosmopolitan *D. melanogaster* (Baudry et al. 2004). Thus, the combined effect of simple population size changes and sex ratio reversal does not seem capable of explaining the empirical data. Sex ratio reversal did allow founder event models with multiple mating to generate slightly lower R_R values, so if there has been such a reversal in *D. melanogaster*, it may still contribute to the observed population differences in X/A diversity ratios.

The serial founder event model considered here is consistent with many aspects of the biology and history of *D. melanogaster*. Wild-caught females often produce progeny from several different males (e.g., Imhof et al. 1998). The species is capable of a very high reproductive rate, so rapid population growth in a newly colonized environment is quite plausible, and genetic drift during recovery from founder events should then be minimal (as assumed under our model). The hypothesis of a series of founder events is consistent with the range expansion inferred for *D. melanogaster*, which, from a likely ancestral range in eastern Africa (Veuille et al. 2004; Haddrill et al. 2005; Pool and Aquadro 2006), expanded north from sub-Saharan Africa. This expansion appears to have originated in the equatorial rift zone and likely followed the Nile Valley into northern Africa (Pool and Aquadro 2006). The serial founder event model might accurately reflect the spread of *D. melanogaster* from one human settlement to the next along the Nile, along with the crossing of arid, sparsely inhabited regions in North Africa and the Middle East. Although we assume a maximum of 5 founder events in this study, theoretical predictions for X-linked and autosomal diversity reductions are very similar for a larger number of founder events (e.g., $b = 100$) involving larger numbers of founder females (data not shown).

The multiple mating model examined here assumes each male that mated with a given female has an equal probability of producing offspring in the newly founded population. In reality, it appears that females often produce a large proportion of offspring from one or a few males and smaller proportions from other males (e.g., Imhof et al. 1998), a pattern that may result from the order of mating and from sperm competition (e.g., Harshman and Clark 1998). In this respect, our multiple mating model might overestimate the genetic diversity contributed by m males mating with each founder female. However, our model may also underestimate male genetic contributions by not accounting for males that actually accompany founder females to the new environment, potentially mating with founder females and their female offspring. Additional data on paternity from wild-caught *D. melanogaster* may allow more realistic models of multiple mating. In the current model, it may be preferable to think of m as the effective number of males per founder female. Importantly, the bulk of the "multiple mating effect" during founder events can be generated by fairly modest values of m (fig. 3) that are likely to be reasonable for *D. melanogaster*.

In this study, we have assumed that sub-Saharan populations of *D. melanogaster* are at demographic equilibrium. These populations generally show an excess of low-frequency mutations, but the relative contributions of demographic processes (such as population growth) and purifying selection (keeping deleterious alleles at low frequency) are unclear (Pool and Aquadro 2006). In any case, it seems unlikely that a departure from demographic equilibrium in sub-Saharan *D. melanogaster* would have a strong effect on our analysis of population diversity ratios.

The low X/A diversity ratio in cosmopolitan *D. melanogaster* has often been attributed, at least in part, to an increased rate of positive selection in these populations (e.g., Kauer et al. 2003; Hutter et al. 2007). This explanation relies on the assumption that selection will act more efficiently on the X chromosome, which depends on the properties of advantageous mutations (e.g., Orr and Betancourt 2001; Betancourt et al. 2004), and although it is likely that cosmopolitan populations of *D. melanogaster* have adapted to new environments, it is not yet clear whether any acceleration in the rate of genetic hitchhiking has been sufficiently dramatic to grossly alter chromosomal patterns of diversity. Clearly, the findings of this study do not exclude the likely action of positive selection in both cosmopolitan and sub-Saharan *D. melanogaster*, but they do suggest that it is not strictly necessary to invoke "out-of-Africa" adaptation as an explanation for the differing X/A diversity ratios of cosmopolitan and sub-Saharan *D. melanogaster*.

Another factor suggested to impact X linked versus autosomal diversity in *D. melanogaster* is local selection on polymorphic autosomal inversions (e.g., Andolfatto 2001, Aulard et al. 2002). The lines sequenced for third chromosome loci by Hutter et al. (2007) were selected to be free of inversion polymorphism. Still, it remains possible that diversity on these chromosomes has been influenced by restricted recombination and the effect of linked selection. Indeed, we found that third chromosome loci within 1 cM of a common inversion's breakpoint had substantially lower diversity in the Zimbabwe ($\pi = 0.00607$ vs. 0.0115; Mann–Whitney $P < 0.01$) and The Netherlands ($\pi = 0.00404$ vs. 0.00788; Mann–Whitney $P < 0.01$) samples than loci more than 1 cM away from such breakpoints (supplementary table S2, Supplementary Material online). However, in spite of this strong effect of inversions on chromosomal variability within each population, the same test showed no effect of proximity to breakpoints on population diversity ratios ($\pi_{Neth}/\pi_{Zim} = 0.699$ and 0.783; Mann–Whitney $P = 0.50$). Therefore, inversions would not be expected to interfere with our analysis of the cosmopolitan sample's diversity reductions on the X and third chromosomes.

This study has focused specifically on predicted and observed levels of X-linked and autosomal nucleotide diversity. However, other aspects of the empirical data from *D. melanogaster* are also consistent with a more severe genetic bottleneck for the X chromosome than the autosomes, as would be generated by founder events involving multiply mated females. Provided that nearly all variation is not lost, more severe bottlenecks should yield higher values of 1) the coefficient of variation of π (the dispersion of locus-specific values around the mean), 2) the variance of Tajima's (1989) *D* statistic among loci, and 3) linkage disequilibrium (Z_{nS8}; Kelly 1997; Hutter et al. 2007). For each of these 3 statistics, the Zimbabwe population has fairly similar X-linked and autosomal values. The Netherlands population has somewhat higher autosomal values than Zimbabwe but considerably greater X-linked values of all 3 (supplementary table S3, Supplementary Material online).

The model considered here should inform future research aimed specifically at inferring demographic parameters in *D. melanogaster* and other species that exhibit multiple mating. Ideally, such studies will consider multiple aspects of X-linked and autosomal polymorphism. In *Drosophila*, however, the choice of appropriate data summaries is not trivial. For example, the 3 statistics listed above are all quite sensitive to assumptions regarding within-locus recombination (which will often be present in *Drosophila*, even for short loci). The frequency spectrum of variable sites offers another potential source of information concerning population history. But it is clear that in *Drosophila*, a large fraction of the genome is under selective constraint (e.g., Andolfatto 2005; Begun et al. 2007), and the extent to which sub-Saharan populations show an excess of rare alleles depends partly upon the inferred constraint of local genomic regions (Pool and Aquadro 2006). Ultimately, the greatest insight concerning the evolutionary forces at work in populations may come from analyses that seek to infer both demographic events and positive selection (e.g., Wright et al. 2005; Li and Stephan 2006), while also implementing reasonable models of mutation, recombination, and selective constraint.

Supplementary Material

Supplementary tables S1–S3 are available at *Molecular Biology and Evolution*online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This research was supported by an National Institutes of Health (NIH) Kirschstein-National Research Service Award postdoctoral fellowship (F32 HG004182) to J.E.P. and an NIH research grant (UO1HL084706) to R.N.

Literature Cited

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 CF, Begun DJ, Kindahl EC. 1994. Selection, recombination and DNA polymorphism in Drosophila. In: Golding B, editor. Non-neutral evolution. New York: Chapman & Hall. p. 46–56.

Aulard S, David JR, Lemeunier F. 2002. Chromosomal inversion polymorphism in Afrotropical populations of *Drosophila melanogaster*. Genet Res. 79:49–63.

Avery PJ. 1984. The population genetics of haplo-diploids and X-linked genes. Genet Res. 44:321–341.

Baudry E, Viginier B, Veuille M. 2004. Non-African populations of *Drosophila melanogaster* have a unique origin. Mol Biol Evol. 21:1482–1491.

Bauer VL, Aquadro CF. 1997. Rates of DNA sequence evolution are not sex-biased in *Drosophila melanogaster* and *D. simulans*. Mol Biol Evol. 14:1252–1257.

Begun DJ, Aquadro CF. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature. 365:548–550.

Begun DJ, Holloway AK, Stevens K, et al. (13 co-authors). 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 5:e310.

Betancourt AJ, Kim Y, Orr HA. 2004. A pseudohitchhiking model of X vs. autosomal diversity. Genetics. 168:2261–2269.
Caballero A. 1995. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics. 139:1007–1011.
Caracristi G, Schlötterer C. 2003. Genetic differentiation between American and European Drosophila melanogaster populations could be attributed to admixture of African alleles. Mol Biol Evol. 20:792–799.
Charlesworth B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet Res. 63:213–227.
Charlesworth B. 1996. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet Res. 134:1289–1303.
Charlesworth B. 2001. The effect of life-history and mode of inheritance on neutral genetic variability. Genet Res. 77:153–166.
Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 130:113–146.
Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics. 134:1289–1303.
Edmonds CA, Lillie AS, Cavalli-Sforza LL. 2004. Mutations arising in the wave front of an expanding population. Proc Natl Acad Sci USA. 101:975–979.
Fay JC, Wu C-I. 1999. A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol Biol Evol. 16:1003–1005.
Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D. 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. Genetics. 165:1269–1278.
Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. 2005. Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Genome Res. 15:790–799.
Harshman LG, Clark AG. 1998. Inference of sperm competition from broods of field-caught *Drosophila*. Evolution. 52:1334–1341.
Hudson RR. 1983. Properties of a neutral allele model with intragenic recombination. Theor Popul Biol. 23:183–201.
Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W. 2007. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. Genetics. 177:469–480.
Imhof M, Harr B, Brem G, Schlötterer C. 1998. Multiple mating in wild *Drosophila melanogaster* revisited by microsatellite analysis. Mol Ecol. 7:915–917.
Kauer M, Dieringer D, Schlötterer C. 2003. A microsatellite variability screen for positive selection associated with the "out of Africa" habitat expansion of *Drosophila melanogaster*. Genetics. 165:1137–1148.
Kelly JK. 1997. A test of neutrality based on interlocus associations. Genetics. 146:1197–1206.
Kingman JFC. 1982a. On the genealogy of large populations. J Appl Probab. 19A:27–43.
Kingman JFC. 1982b. The coalescent. Stoch Process Appl. 13:235–248.
Lachaise D, Cariou M, David JR, Lemeunier F, Tsacas L, Ashburner M. 1988. Historical biogeography of the *Drosophila melanogaster* species subgroup. In: Hecht MK, Wallace B, Prance GT, editors. Evolutionary biology. New York: Plenum Press. Vol. 22p. 159–225.
Li H, Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. PLoS Genet. 2:e166.
Li W-H, Yi S, Makova K. 2002. Male-driven evolution. Curr Opin Genet Dev. 12:650–656.
Milkman R, Zeitler RR. 1974. Concurrent multiple paternity in natural and laboratory populations of *Drosophila melanogaster*. Genetics. 78:1191–1193.
Ochando MD, Reyes A, Ayala FJ. 1996. Multiple paternity in two natural populations (orchard and vineyard) of *Drosophila*. Proc Natl Acad Sci USA. 93:11769–11773.
Ometto L, Glinka S, De Lorenzo D, Stephan W. 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.
Orr HA, Betancourt AJ. 2001. Haldane's sieve and adaptation from the standing genetic variation. Genetics. 157:875–884.
Pascual M, Chapuis MP, Mestres F, Balanya J, Huey RB, Gilchrist GW, Serra L, Estoup A. 2007. Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. Mol Ecol. 16:3069–3083.
Pool JE, Aquadro CF. 2006. History and structure of sub-Saharan populations of *Drosophila melanogaster*. Genetics. 174:915–929.
Pool JE, Nielsen R. 2007. Population size changes reshape genomic patterns of diversity. Evolution. 61:3001–3006.
Reiland J, Hodge S, Noor MAF. 2002. Strong founder effect in *Drosophila pseudoobscura* colonizing New Zealand from North America. J Hered. 93:415–420.
Riegler M, Sidhu M, Miller WJ, O'Neill SL. 2005. Evidence for a global *Wolbachia* replacement in *Drosophila melanogaster*. Curr Biol. 15:1428–1433.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.
Thornton K, Andolfatto P. 2006. Approximate Bayesian inference reveals evidence for a recent, severe, bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics. 172:1607–1619.
Veuille M, Baudry E, Cobb M, Derome N, Gravot E. 2004. Historicity and population genetics of *Drosophila melanogaster* and *D. simulans*. Genetica. 120:61–70.
Wall JD, Andolfatto P, Przeworski M. 2002. Testing models of selection and demography in *Drosophila simulans*. Genetics. 162:203–216.
Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. Science. 308:1310–1314.

Jeffrey Thorne, Associate Editor

Accepted May 21, 2008