The Influence of Demography and Weak Selection on the McDonald–Kreitman Test: An Empirical Study in Drosophila

John Parsch, Zhi Zhang, and John F. Baines

Department of Biology, University of Munich, Germany

The McDonald–Kreitman (MK) test, which compares the ratio of polymorphism to divergence at nonsynonymous and synonymous sites, is frequently used to detect adaptive evolution in protein-coding sequences. Because the two classes of sites share a common evolutionary history, the MK test is thought to be robust to most demographic factors. However, weak selection on nonsynonymous sites can bias the MK test, especially when a species’ effective population size has not been constant. Here, we present an empirical analysis of the influence of demography on the MK test by comparing test results for a common set of 136 genes, including a set of sex-biased genes that shows a strong signal of adaptive evolution, in two Drosophila melanogaster populations: an ancestral population from Africa and a derived population from Europe. The latter has undergone a relatively recent bottleneck, which has reduced its effective population size. We find that the MK test has less power to detect positive selection in the European population for two reasons. First, the overall reduced level of standing variation decreases the statistical power of the test. Second, the segregation of slightly deleterious nonsynonymous mutations biases the MK test away from detecting positive selection. The latter effect is stronger for X-linked genes, which have experienced the greatest reduction in effective population size outside of Africa, and also leads to the underestimation of rates of adaptive protein evolution by multilocus implementations of the MK test. Interestingly, a subset of autosomal female-biased genes shows an increased signal of adaptive evolution in the European population. This is inconsistent with currently accepted demographic scenarios and may reflect female-specific changes in selective constraint following the colonization of non-African habitats.

Introduction

Since its introduction in 1991, the McDonald–Kreitman (MK) test has become a standard method for the detection of positive (or balancing) selection on protein-coding sequences (McDonald and Kreitman 1991; MacCallum and Hill 2006). The test requires both polymorphism and divergence data and is conceptually very simple: A two-by-two contingency table is constructed with columns for divergent and polymorphic sites, and rows for nonsynonymous and synonymous sites. The cell entries in the table represent the observed numbers of divergent nonsynonymous sites ($D_n$), polymorphic nonsynonymous sites ($P_n$), divergent synonymous sites ($D_s$), and polymorphic synonymous sites ($P_s$). Synonymous sites are assumed to evolve neutrally and departures from neutral evolution are detected by testing for significant differences between the ratios of the two rows. A significant excess of nonsynonymous divergence ($D_n/D_s > P_n/P_s$) is interpreted as a signal of positive selection, whereas a significant excess of nonsynonymous polymorphism ($D_n/D_s < P_n/P_s$) is interpreted as a signal of balancing selection. In addition to the neutrality of synonymous sites, the above interpretations assume that mutations fall into one of three categories: neutral, strongly beneficial, or strongly deleterious.

The action of weak selection, in which the absolute value of the selection coefficient $s$ is approximately equal to the inverse of the effective population size $N_e$, can affect the MK test in several ways. This especially applies to cases where $N_e$ has not been constant over the evolutionary history of the species investigated. For example, McDonald and Kreitman (1991) pointed out that an expansion in population size could lead to a false inference of positive selection. This could occur if slightly deleterious nonsynonymous mutations were fixed in a small ancestral population, but equally deleterious mutations no longer segregate in the current population. This is possible in cases of very recent population expansion or in which there has been a prolonged, severe bottleneck since the time of species divergence (Eyre-Walker 2002). A more pervasive problem, namely, the segregation of slightly deleterious mutations, biases the MK test in the opposite direction (Charlesworth and Eyre-Walker 2008). These mutations contribute much more to $P_n$ than to $D_n$ and thus will bias the MK test toward the detection of balancing selection (i.e., away from the detection of positive selection). This problem will be exacerbated in populations that have undergone a recent reduction in $N_e$.

Although theoretical aspects of the influence of weak selection and demography on the MK test are well known, empirical data regarding their effect on tests applied to natural population samples are lacking. To address this, we analyzed polymorphism and divergence in protein-coding genes from two population samples of Drosophila melanogaster, one from the species’ ancestral range in sub-Saharan Africa (Zimbabwe) and one from a derived population in Europe (the Netherlands). The demographic history of these two populations has been studied extensively and noncoding DNA polymorphism data suggest that the African population has undergone a mild expansion over the past 60,000 years, whereas the European population experienced a severe bottleneck about 16,000 years ago accompanying the out-of-Africa migration (Li and Stephan 2006; Stephan and Li 2006). Populations outside of Africa have not yet recovered from this bottleneck and, consequently, show an overall reduction in polymorphism relative to the African population (Begun and Aquadro 1993; Schlötterer et al. 1997; Glinka et al. 2003; Ometto et al. 2005). This reduction is
...strongest on the X chromosome (Andolfatto 2001; Kauer et al. 2002; Hutter et al. 2007). Its complex population history, together with the abundance of both positive and weak selection in the species (Sawyer et al. 2007), makes Drosophila melanogaster an excellent model system for investigating the effects of demography on the MK test.

Our sample of 136 genes contains roughly equal numbers that show enriched expression in males (male-biased genes), enriched expression in females (female-biased genes), or equal expression in the two sexes (unbiased genes). Within each expression class, approximately two-thirds of the genes are autosomal and one-third are X-linked. Previous studies of the African population have indicated that male-biased genes show elevated rates of adaptive protein evolution (Pröschel et al. 2006) and that this is especially pronounced for X-linked, male-biased genes (Baines et al. 2008). In contrast, unbiased genes (both autosomal and X-linked) show little signal of adaptive protein evolution. Here we present DNA sequence polymorphism data from the same set of genes in a European sample of equal size. This allows us to determine the effect of population size changes on the MK test for several categories of genes. We find that the MK test has less power to detect positive selection in the European population. For autosomal genes, this is mainly a result of the reduced number of segregating sites in the European population. For the X-linked genes, which have experienced a greater reduction in Ne, the segregation of slightly deleterious nonsynonymous mutations in Europe also reduces the power of the MK test. Interestingly, autosomal female-biased genes appear to have fewer deleterious mutations segregating in the European population. Such a pattern is not consistent with the inferred changes in population size that accompanied the colonization of non-African habitats and suggests that there may have been strong selective constraints on female reproduction during the out-of-Africa migration.

Materials and Methods
Gene Selection and African Polymorphism Data

The 136 protein-coding genes are comprised of 91 autosomal genes reported by Pröschel et al. (2006) and 45 X-linked genes reported by Baines et al. (2008). A complete list is provided in supplementary table S1 (Supplemental Material online). Within each set, the genes were further classified according to their relative expression in males and females, as determined by a consensus of three independent microarray experiments (Parisi et al. 2003; Ranz et al. 2003; Gibson et al. 2004; Gnäd and Parsch 2006). The autosomal set contains 33 male-biased, 28 female-biased, and 30 unbiased genes. The X-linked set contains 17 male-biased, 13 female-biased, and 15 unbiased genes. For both the autosomal and X-linked gene sets, male-biased and female-biased genes show a strong enrichment of expression in testis and ovary, respectively (Chintapalli et al. 2007; Baines et al. 2008). In all cases, the genes were chosen on the basis of their relative expression level in the two sexes, without consideration of function or interspecific divergence. The vast majority had no annotated function. Genes were also chosen such that other factors that may influence molecular evolution, such as local recombination rate, coding sequence length, and intron/exon structure, were comparable among groups (Pröschel et al. 2006; Baines et al. 2008).

Polymorphism and divergence statistics for the 136 genes in the African population were previously reported by Pröschel et al. (2006) and Baines et al. (2008), who surveyed 12 highly inbred isofemale lines from Lake Kariba, Zimbabwe (Begun and Aquadro 1993; Glinka et al. 2003). The average number of alleles reported per gene was 11. When available, a single sequence from an inbred strain of Drosophila simulans from Chapel Hill, NC (Meiklejohn et al. 2004) was used to determine divergence. Otherwise, the D. simulans genome sequence was used (Drosophila 12 Genomes Consortium 2007).

European Polymorphism Data

To survey DNA sequence polymorphism in a derived population of D. melanogaster, a sample of 12 highly inbred isofemale lines from Leiden, the Netherlands (Glinka et al. 2003) was used. The coding regions of the 136 genes were polymerase chain reaction (PCR) amplified using primers and cycling conditions described previously (Pröschel et al. 2006; Baines et al. 2008). The PCR products were purified with ExoSAP-IT (USB, Cleveland, OH) and sequenced from both strands using BigDye chemistry and a 3730 automated sequencer (Applied Biosystems, Foster City, CA). The average number of alleles successfully amplified and sequenced per gene was 11.4. All new sequences have been submitted to the GenBank/EMBL databases under the accession numbers FM244915–FM246454. Interspecific divergence was determined using the same D. simulans sequences described above for the African population.

DNA polymorphism and divergence statistics, including single-locus MK test statistics, were calculated with DnaSP 4 (Rozas et al. 2003). Multilocus MK test statistics were calculated using the program DoFE (Distribution of Fitness Effects), which was kindly provided by A. Eyre-Walker.

Simulation of MK-Table Data

MK-table data were simulated under the assumption of neutrality with the program ms (Hudson 2002) following the approach of Andolfatto (2008), but modified to apply to protein-coding sequences in which synonymous and nonsynonymous sites are interdigitated. For this, we assumed that every third site in the sequence was nonsynonymous and all other sites were synonymous. Although ms assumes an infinite sites model for mutations, the relative position of each mutation along a sequence is provided as a four-digit decimal between zero and one. In practice, if a mutation’s position was evenly divisible by 0.0003 it was assigned as nonsynonymous. This arrangement was not chosen to reflect the nature of the genetic code in a protein-coding sequence (in which the number of
nonsynonymous sites outnumbers the number of synonymous sites), but rather to account for the strong purifying selection that acts at some nonsynonymous sites. For example, the observed mean $D_h$ for our autosomal genes is $\sim 2$ times higher than the observed mean $D_s$. Thus, our arrangement leads to simulated counts that reflect the observed data. All simulations generated population samples of 12 sequences that diverged from a single outgroup sequence 12.8 $N_e$ generations ago, with these values also being chosen to best reflect the observed data. Andolfatto (2008) reported that factors such as recombination rate variation and demography could increase the type-I error rate of the MK test in cases where the neutral and selected sites were not interdigitated. To test if these factors affect the MK test when applied to interdigitated neutral and selected sites, we determined the proportion of 10,000 simulation runs with $P < 0.05$ ($G$-test) for a wide range of population-scaled mutation ($\theta$) and recombination ($\rho$) rate parameters, as well as for a bottleneck scenario that has been proposed for D. melanogaster. Only simulated MK tables with marginal counts $> 5$ and at least one count in each cell were used for calculation of $P$ values. We observed no increase in the type-I error rate for any combination of parameters (supplementary fig. S1, Supplementary Material online).

To examine the effect of $\theta$ on the type-II error rate of the MK test, we performed simulations as described above, but included positive selection by increasing the counts of $D_h$ according to the parameter $\alpha$, which indicates the proportion of nonsynonymous differences between species that were fixed by positive selection. Similarly, to investigate the effects of segregating deleterious nonsynonymous mutations on the type-II error of the MK test, we increased counts of $D_s$ according to the parameter $\delta$, which indicates the proportion of nonsynonymous polymorphisms that are deleterious. For analyses of type-II error rates, we performed two-tailed $G$-tests (as would be done in practice), but only counted significant tests ($P < 0.05$) consistent with positive selection (i.e., $D_d/D_s > P_d/P_s$).

Results and Discussion

Effective Population Sizes of the Autosomes and X Chromosome in Europe and Africa

Previous studies have used DNA sequence polymorphism in intronic and intergenic regions to infer the demographic history of the African (Zimbabwe) and European (the Netherlands) populations of D. melanogaster considered here (Glinka et al. 2003; Ometto et al. 2005; Li and Stephan 2006; Hutter et al. 2007). These studies assumed that noncoding regions evolve neutrally and, thus, provide reliable information about the $N_e$ of a population. However, there is evidence that noncoding regions are subject to more purifying and positive selection than synonymous sites (Andolfatto 2005). To determine if this selection has an effect on demographic inference, we examined levels of synonymous polymorphism in the complete set of 136 protein-coding genes and compared them with the data from noncoding regions. In total, we examined 26,500 synonymous sites, 1,773 (6.7%) of which were polymorphic in at least one of the populations. Overall, our results agreed well with those of Hutter et al. (2007). For the autosomes, synonymous nucleotide polymorphism ($\theta_{syn}$) is $\sim$1.4-fold higher in Africa than in Europe; for the X chromosome, it is $\sim$2.4-fold higher (table 1). Several factors may contribute to the stronger out-of-Africa bottleneck observed for the X chromosome. First, if an equal number of males and females passed through the bottleneck, it is expected to have a larger impact on the X chromosome than the autosomes, because there are only three X chromosomes for every four autosomes in the population (Wall et al. 2002; Pool and Nielsen 2007). Second, an increase in the effective number of males relative to females in the derived population and/or the colonization of the non-African habitat through a series of founder events involving multiply mated females could lead to a greater than expected reduction in X-linked diversity (Charlesworth 2001; Hutter et al. 2007; Pool and Nielsen 2008). Finally, an increased rate of selective sweeps on the X chromosome could disproportionately reduce its level of standing variation following a bottleneck (Kauer et al. 2002; Hutter et al. 2007; Singh et al. 2007). A detailed treatment of the above scenarios can be found in Singh et al. (2007).

In the African population, $\theta_{syn}$ is slightly higher for the X chromosome than for the autosomes (table 1). This can partially be explained by a higher mutation rate on the X chromosome (Begun et al. 2007; Hutter et al. 2007), as synonymous divergence ($K_{syn}$) is also higher for the X chromosome (table 1). However, even after correcting for this mutational difference, the ratio of X-linked to autosomal polymorphism is 1.06 and is significantly higher than the value of 0.75 expected under an equal sex ratio (Mann–Whitney $U$-test, $P = 0.008$). This suggests that males have a smaller $N_e$ than females in the African population, which is consistent with greater sexual selection acting on males (Nunney 1993; Charlesworth 2001). In the European population, the situation is reversed and the mutation-corrected ratio of X to autosomal synonymous polymorphism is 0.64. This does not differ significantly from the expected ratio of 0.75 (Mann–Whitney $U$-test, $P = 0.39$), but is in qualitative agreement with previous findings that males have a larger $N_e$ than females in the European population (Hutter et al. 2007).

In both of the above cases, our estimate of the mutation-corrected ratio of X to autosomal polymorphism was higher than that of Hutter et al. (2007), who estimated ratios of 0.90 and 0.49 for the African and European populations, respectively. The main reason for this discrepancy appears to be the elevated ratio of X-linked to autosomal divergence at noncoding sites relative to synonymous sites. For noncoding sites, Hutter et al. (2007) estimated that divergence on the X chromosome is $\sim$30% greater than that on the autosomes, whereas for synonymous sites we estimate that X-linked divergence is $\sim$7% greater than autosomal divergence (table 1). Because the ratio of polymorphism to divergence is used to correct for a possible mutational bias when estimating $N_e$, this leads to synonymous sites giving higher estimates than noncoding sites for the $N_e$ of the X chromosome. The increased divergence of X-linked noncoding sequences could be explained if positive selection acts more frequently on X-linked than autosomal loci.
There is some evidence of a “fast-X” effect for *Drosophila* protein-coding genes (Counterman et al. 2004; Musters et al. 2006; Baines et al. 2008; but see Thornton et al. 2006), however, comparable data for noncoding regions are currently lacking.

**Effect of Demography on the MK Test**

To determine the effect of demography on the MK test, we compared test results for all 136 genes in the African and European populations. For each population, a common sample size of 12 inbred lines was used for the polymorphism survey. Previous studies of the African population identified 20 genes that gave a significant MK test for positive selection (Pro¨schel et al. 2006; Baines et al. 2008). For the same set of genes, only nine showed a significant signal of positive selection in the European population (table 2), which is significantly fewer than in the African population (Fisher’s exact test, \( P < 0.048 \)). Seven genes were significant in both populations. Thus, two genes were significant only in Europe and 13 genes were significant only in Africa.

The above results indicate that the power to detect positive selection by the MK test is reduced in the European population. An obvious contributor to this difference is that the overall lower levels of polymorphism in the European population reduce the power of the MK test by lowering the counts of \( P_s \) and \( P_a \) in the MK tables. To estimate the magnitude of this effect, we performed coalescent simulations with parameters relevant to the two *D. melanogaster* populations (see Materials and Methods). Figure 1A shows the effect of \( \theta \) (in terms of \( 4N_e L \), where \( \mu \) is the mutation rate per site per generation and \( L \) is the sequence length) on the type-II error rate of the MK test for several values of the selection parameter \( \alpha \), which indicates the proportion of nonsynonymous differences between species that were fixed by positive selection (Smith and Eyre-Walker 2002). For *Drosophila*, previous studies have estimated that \( \alpha \) falls within the range of 50–95% (Eyre-Walker 2006; Andolfatto 2007; Sawyer et al. 2007). Our own data suggest an \( \alpha \) value of \( \approx 60\% \) (see below). Assuming \( \alpha = 0.6 \) and \( \theta \) values that correspond to our observed data for the autosomes (5.4 and 3.1 for Africa and Europe, respectively), we expect the type-II error rate to be 11% higher for the European population. For the X chromosome (\( \theta = 5.4 \) and \( \theta = 1.5 \) for Africa and Europe, respectively) we expect the type-II error rate to be 16% higher for the European population.

**Segregation of Slightly Deleterious Mutations**

In addition to decreasing the overall power of the MK test by lowering counts in the polymorphism cells of
MK table, a reduction in \( N_e \) may also bias the MK test away from detecting positive selection by increasing the ratio \( P_d/P_s \). Overall, our data suggest that many slightly deleterious nonsynonymous mutations segregate in both the African and the European population. For both populations, the proportion of nonsynonymous polymorphisms present at low frequency (\( \leq 15\% \)) is significantly greater than the proportion of synonymous polymorphisms (\( \chi^2 \)-test, \( P < 0.001 \)). The segregation of deleterious mutations has the greatest effect on the European X chromosome, which has the smallest \( N_e \). This is evident from the ratio of nonsynonymous to synonymous polymorphism, \( \theta_{\text{non}}/\theta_{\text{syn}} \). Overall, the European X chromosome has the lowest \( \theta_{\text{syn}} \) and the highest \( \theta_{\text{non}}/\theta_{\text{syn}} \) (fig. 3). Similarly, the European X chromosome has the highest overall \( P_d/P_s \) (table 3). For the 45 X-linked genes in our survey, the summed value of \( P_d/P_s \) in the European population is significantly greater than that in the African population (\( P = 0.03 \), as determined from 10,000 bootstrap replicates). For the 91 autosomal genes there is no difference in the summed value of \( P_d/P_s \) between Europe and Africa (\( P = 0.5 \)).

The above differences between the X chromosome and autosomes are expected, given our inferences about their relative population sizes in Europe and Africa. Assuming a gamma distribution of fitness effects for deleterious mutations, the effect of population size differences on \( P_d/P_s \) can be approximated as

\[
\frac{P_d}{P_s}\big|_{\text{Europe}} / \frac{P_d}{P_s}\big|_{\text{Africa}} \sim \left( \frac{N_e|_{\text{Europe}}}{N_e|_{\text{Africa}}} \right)^{\beta},
\]

where \( \beta \) is the shape parameter of the gamma distribution (Welch et al. 2008). For *Drosophila*, \( \beta \) has been estimated to be in the range 0.3–0.4 (Loewe et al. 2006; Kightley and Eyre-Walker 2007). Assuming \( \beta = 0.3 \) and the population size differences inferred from synonymous polymorphism (table 1), we expect \( \frac{P_d}{P_s}\big|_{\text{Europe}} / \frac{P_d}{P_s}\big|_{\text{Africa}} \) to be 1.30 for X-linked genes, which agrees well with our observed ratio of 1.26. For autosomal genes, we expect \( \frac{P_d}{P_s}\big|_{\text{Europe}} / \frac{P_d}{P_s}\big|_{\text{Africa}} \) to be 1.12. However, the observed ratio over all autosomal genes (0.99) suggests that the difference in \( N_e \) has no effect on \( P_d/P_s \). This discrepancy may result from differences in selection among sex-biased genes (discussed below).

To further examine the effect of segregating deleterious mutations on the MK test, we performed coalescent simulations of MK-table data in which the counts of \( P_d \) were increased according to the parameter \( \delta \), where \( \delta \) represents the fraction of nonsynonymous polymorphisms that are deleterious. Figure 1B shows the effect of \( \delta \) on the type-II error rate of the MK test for a range of \( z \) values and with other parameters fixed at values appropriate for our *Drosophila* data. As expected, the power to detect positive selection decreases as \( \delta \) increases. When differences in both \( \theta \) (fig. 1A) and \( \delta \) (fig. 1B) are taken into account, the type-II error rate is expected to be 27% higher in the European population than in the African population for parameters relevant to our X-chromosomal data. For parameters relevant to our autosomal data, the type-II error rate is expected to be 16% higher in the European population than in the African population.

**The Effect of Sex-Biased Gene Expression**

Previous studies of the African *D. melanogaster* population have shown that the rate of adaptive evolution differs among genes with respect to their expression level in the two sexes (Pröschel et al. 2006; Baines et al. 2008). Sex-biased genes, especially those with male-biased expression, show a strong signal of adaptive protein evolution. In contrast, unbiased genes show little signal of adaptive evolution and appear to evolve mainly under purifying selection. This same pattern is seen for the European population. Of the nine genes that give a significant MK test for positive selection in the European population, six are male biased, three are female biased, and none are unbiased (table 1). To determine the type and strength of selection affecting each group of genes, we estimated the selection parameter \( z \) using the maximum likelihood method of Bierne and Eyre-Walker (2004), which is applied to the individual MK tables within each expression class. With this approach, we also see a signal of positive selection for the sex-biased genes, especially those with male-biased expression, and a signal of weak purifying selection for unbiased genes (fig. 4). As mentioned above, the low frequency of nonsynonymous polymorphisms in the African and European populations (fig. 2) suggests that the segregation of slightly deleterious mutations may lead to the underestimation of \( z \).
To correct for this, we also calculated \( \alpha \) after excluding all polymorphisms (both synonymous and nonsynonymous) segregating at a frequency of less than 15% (Charlesworth and Eyre-Walker 2008). In all cases, \( \alpha \) increased after removal of the low-frequency polymorphisms (fig. 4), although the relative values among male-biased, female-biased, and unbiased genes remained unchanged.

The segregation of slightly deleterious mutations appears to have the largest effect on unbiased genes, which previously have been shown to evolve primarily under purifying selection (Pröschel et al. 2006; Baines et al. 2008). For both autosomal and X-linked unbiased genes, \( \alpha \) is higher in the African population than in the European population (fig. 4). Furthermore, for both the autosomes and the X chromosome, \( P_n/P_s \) is higher in the European population than in the African population (table 3). For the X chromosome, the difference in \( P_n/P_s \) between Europe and Africa is significant (\( P = 0.05 \); here and in the following \( P \) values were determined from 10,000 bootstrap replicates). For the autosomes, the difference in \( P_n/P_s \) between Europe and Africa is smaller and is not significant (\( P = 0.33 \)). These results are consistent with a greater reduction in \( N_e \) for the X chromosome in the European population.

Table 3
Summary of Polymorphism and Divergence

<table>
<thead>
<tr>
<th>Bias</th>
<th>Chr</th>
<th>( D_s )</th>
<th>( P_s )</th>
<th>( D_n )</th>
<th>( P_n )</th>
<th>( D_n/D_s )</th>
<th>( P_n/P_s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Auto</td>
<td>741</td>
<td>447</td>
<td>368</td>
<td>112</td>
<td>0.50</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>464</td>
<td>235</td>
<td>645</td>
<td>96</td>
<td>1.39</td>
<td>0.41</td>
</tr>
<tr>
<td>Female</td>
<td>Auto</td>
<td>624</td>
<td>233</td>
<td>299</td>
<td>90</td>
<td>0.48</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>624</td>
<td>233</td>
<td>299</td>
<td>90</td>
<td>0.48</td>
<td>0.39</td>
</tr>
<tr>
<td>Unbiased</td>
<td>Auto</td>
<td>436</td>
<td>266</td>
<td>118</td>
<td>80</td>
<td>0.27</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>227</td>
<td>182</td>
<td>129</td>
<td>96</td>
<td>0.57</td>
<td>0.53</td>
</tr>
<tr>
<td>All</td>
<td>X</td>
<td>1,801</td>
<td>946</td>
<td>785</td>
<td>282</td>
<td>0.44</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>947</td>
<td>546</td>
<td>959</td>
<td>236</td>
<td>1.01</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Overall, the male-biased genes show the least effect of the out-of-Africa population bottleneck, with only slight differences in \( \alpha \) between Africa and Europe (fig. 4). Male-biased genes also have very similar \( P_n/P_s \) in Africa and Europe for both autosomal and X-linked genes (table 3; \( P > 0.15 \); in both cases). These results suggest that the out-of-Africa bottleneck had little effect on the frequency of segregating deleterious nonsynonymous mutations in male-biased genes. This may be a result of differences in \( N_e \) between males and females in the two populations. Because all of the male-biased genes analyzed here show strong expression enrichment in male reproductive tissues, it is likely that they are subject to purifying selection only when they are in the male genetic background. On the basis of X-linked to autosomal polymorphism ratios in noncoding DNA, Hutter et al. (2007) estimated that the ratio of male to female \( N_e \) is \( \approx 0.6 \) in Africa, whereas it is at least 12 in Europe. Although the latter may be an overestimate if there has been an increase in positive selection acting on X-linked loci in Europe, it is likely that there has been at least some increase in the male–female sex ratio in the European population relative to the African population. For males, this change in sex ratio could counteract the overall reduction...
in Ne resulting from the out-of-Africa bottleneck, causing purifying selection on male-biased genes to be as effective in Europe as it is in Africa.

Female-biased genes show a more complex pattern that differs between the autosomes and the X chromosome. X-linked female-biased genes have lower z in Europe than in Africa (fig. 4). They also have higher $P_d/P_s$ in Europe than in Africa (table 3; $P = 0.02$). Thus, purifying selection appears to be less efficient in the European population, which is consistent with a large reduction in $Ne$ of both the X chromosome and females in Europe. Autosomal female-biased genes, however, do not follow the expected pattern. Instead, they show higher z in the European population (fig. 4) and have significantly lower $P_d/P_s$ in the European population than in the African population ($P = 0.01$; table 3). These observations are inconsistent with the inferred demographic changes that accompanied the colonization of Europe and suggest that purifying selection is more effective on autosomal female-biased genes in the European population. This could be explained if some of the nonsynonymous mutations segregating in female-biased genes have sexually antagonistic effects that are beneficial to females but harmful to males. Such mutations could reach higher frequency in Africa, where $Ne$ is larger for females than for males. This scenario could also account for the observed difference between autosomal and X-linked female-biased genes: if the deleterious effects of these mutations are recessive, they would be less likely to accumulate on the X chromosome, where they could be selected against in hemizygous males. The expression data, however, provide little support for sexual antagonism. Of the 28 autosomal female-biased genes, only three show a signal of expression in tissues other than ovary: CG6554 and CG1209 show weak expression in accessory gland and CG3975 shows weak expression in brain and midgut (Chintapalli et al. 2007). Although all of these genes have higher z and lower $P_d/P_s$ in Europe than in Africa and could be candidates for sexual antagonism, excluding them from the female-biased genes does not substantially alter the overall patterns reported above. Without these three genes, $P_d/P_s$ is still lower in Europe (0.26) than in Africa (0.35; $P = 0.06$) and the maximum likelihood estimate of z is 0.27 in Europe and 0.07 in Africa (0.54 and 0.39 after exclusion of low-frequency polymorphisms, respectively). The possibility remains, however, that low levels of male expression are present but not detected among these remaining 25 genes.

The autosomal female-biased gene results also could be explained if the selection coefficients of deleterious nonsynonymous mutations differ between the two populations, such that mutations that are neutral or very slightly deleterious in Africa are much more deleterious in Europe. Because the female-biased genes show enriched expression in ovary, the presence of such population-specific deleterious mutations could reduce fertility in European females relative to African females and provide a biological basis for the apparent reversal of sex ratio between the two populations. Experimental data relating genetic variation in female-biased genes to female fertility under different environmental conditions are needed to test this possibility.

Conclusion

Our results indicate that, in general, it is best to use polymorphism data from the ancestral African population for MK tests. For single-locus tests, one advantage is the greater polymorphism in the ancestral population, which increases the power to detect statistically significant differences in the two-by-two contingency table. A second advantage is that slightly deleterious nonsynonymous polymorphisms, which bias the MK test away from the detection of positive selection, segregate at lower frequency in the ancestral population. The latter is also important for multilocus MK tests that estimate the fraction of adaptive amino acid replacements for a group of genes or for the whole genome, as these tests are affected more by the segregation of deleterious mutations than by the numbers of polymorphic sites when the number of genes analyzed is large. Finally, it should be noted that in both the ancestral and the derived population, most segregating nonsynonymous polymorphisms appear to be deleterious and kept in low frequency by purifying selection. Thus, in all cases, the exclusion of low-frequency polymorphisms is recommended when estimating the rate of adaptive protein evolution.

Supplementary Material

Supplementary table S1 and supplementary figure S1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank K. Azadov, Y. Cämmerer, H. Gebhart, and M. Pröschel for help with sequencing. We also thank S. Hutter, W. Stephan, A. Eyre-Walker, and an anonymous reviewer for constructive comments on the manuscript. This work was supported by Deutsche Forschungsgemeinschaft grant PA 903/4.

Literature Cited


Hideki Inman, Associate Editor

Accepted December 21, 2008