Surprising Differences in the Variability of Y Chromosomes in African and Cosmopolitan Populations of *Drosophila melanogaster*

Amanda M. Larracuente*,†,1 and Andrew G. Clark*
*Molecular Biology and Genetics, Cornell University, New York 14853, †Department of Biology, University of Rochester, Rochester, New York 14627

ABSTRACT The nonrecombining *Drosophila melanogaster* Y chromosome is heterochromatic and has few genes. Despite these limitations, there remains ample opportunity for natural selection to act on the genes that are vital for male fertility and on Y factors that modulate gene expression elsewhere in the genome. Y chromosomes of many organisms have low levels of nucleotide variability, but a formal survey of *D. melanogaster* Y chromosome variation had yet to be performed. Here we surveyed Y-linked variation in six populations of *D. melanogaster* spread across the globe. We find surprisingly low levels of variability in African relative to Cosmopolitan (*i.e.*, non-African) populations. While the low levels of Cosmopolitan Y chromosome polymorphism can be explained by the demographic histories of these populations, the staggeringly low polymorphism of African Y chromosomes cannot be explained by demographic history. An explanation that is entirely consistent with the data is that the Y chromosomes of Zimbabwe and Uganda populations have experienced recent selective sweeps. Interestingly, the Zimbabwe and Uganda Y chromosomes differ: in Zimbabwe, a European Y chromosome appears to have swept through the population.

THE *Drosophila melanogaster* Y chromosome is highly functionally specialized in male-related activities (Hardy *et al.* 1981; Kennison 1981; Goldstein *et al.* 1982). Although it comprises ~13% of the male genome (40 Mb; Hoskins *et al.* 2002), the *D. melanogaster* Y chromosome has only 13 known protein-coding genes (Goldstein *et al.* 1982; Carvalho *et al.* 2000, 2001; Vibranovski *et al.* 2008; Krsticevic *et al.* 2010), all thought to be active only in primary spermatocytes in the testis (Hardy *et al.* 1981). The remainder of the Y chromosome is heterochromatic and dense in repetitive elements (Hoskins *et al.* 2002).

Although the *D. melanogaster* Y chromosome is highly repetitive and gene poor, several lines of evidence suggest that it harbors functionally important variation (Chippindale and Rice 2001; Rohmer *et al.* 2004; Lemos *et al.* 2008; Lemos *et al.* 2010). The Y chromosome has contributed to variation in thermotolerance across *D. melanogaster* populations (Rohmer *et al.* 2004), has epistatic effects on male fitness (Chippindale and Rice 2001), and has epigenetic effects on the expression of genes across the genome (Lemos *et al.* 2008, 2010; Jiang *et al.* 2010), including specifically in the male germline (Zhang 2000). The functional variation on the Y chromosome has been somewhat of a paradox because, while there is evidence for structural polymorphism (Lyckegaard and Clark 1989), the haploid transmission of the Y chromosome makes it far more difficult to maintain substantial levels of genetic variation (Clark 1987). In one small survey of Y chromosomal variation in a protein-coding gene, *kl-5* (*Dhc–Yh3*), only a single segregating site was discovered in 11 lines of *D. melanogaster* and no variants were found in 10 lines of *Drosophila simulans* (Zurovcova and Eanes 1999). A more recent study in a closely related species, *D. simulans*, showed low levels of polymorphism in fragments sequenced from the Y-linked genes *kl-2* and *ORY* (Kopp *et al.* 2006). While no large survey of *D. melanogaster* Y chromosome variation has yet been reported, these small surveys suggest that the Drosophila Y chromosome is nearly devoid of nucleotide variation.

The Y chromosome has an effectively clonal inheritance: it is passed from father to son each generation without recombining. As a result, the Y chromosome is particularly

doi: 10.1534/genetics.112.146167
Manuscript received May 29, 2012; accepted for publication October 12, 2012
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.146167/-/DC1.
[1]Corresponding author: University of Rochester, Dep of Biology, River Campus, 480 Hutchison Hall, Rochester, NY 14627. E-mail: alarracu@bio.rochester.edu

sensitive to the demographic history of populations. Population size fluctuations are a potential cause of reduced Y chromosome variation. *D. melanogaster* originated in sub-Saharan Africa and colonized Europe beginning roughly 16,000 years ago (Thornton and Andolfatto 2006) and reached the Americas only in the past few hundred years (David and Capy 1988). Thus, the demographic history of Cosmopolitan (non-African) populations includes population bottlenecks corresponding to colonization events. Population size changes can have opposite effects on different regions of the genome, depending on the model parameters (Pool and Nielsen 2007). For example, because of its smaller effective population size, the X chromosome is expected to recover from population size changes faster than the autosomes. Outside of Africa, the ratio of X-to-autosome diversity is lower than the expected 3/4 (Andolfatto 2001). A recent, severe bottleneck could potentially explain the dearth of variation on the X chromosome (Pool and Nielsen 2007). These effects should be more pronounced for haploid regions of the genome such as the Y chromosome and mitochondria (Pool and Nielsen 2007)—a theoretical prediction that has yet to be tested with empirical data.

We surveyed Y-linked polymorphism in *D. melanogaster* across the globe to investigate patterns of variation and to explore the effects of population bottlenecks on patterns of Y chromosome variation and differentiation between populations. We surveyed introns of four Y-linked genes in six populations and show that there indeed are low levels of polymorphism on the Y chromosome. Surprisingly, we observed a striking reduction of variation on the Y chromosome in African relative to Cosmopolitan populations—a pattern that contrasts with polymorphism data for the X chromosome and autosomes. We conclude that while recent natural selection has clearly shaped Y chromosome evolution in African populations, the demographic history of the Cosmopolitan populations is capable of explaining the patterns of diversity on the Y chromosome. Although we cannot formally reject neutrality of the Cosmopolitan Y chromosomes under bottleneck models, other attributes of the data suggest that our ability to detect positive selection on these chromosomes may be limited.

Materials and Methods

Fly strains

We surveyed *D. melanogaster* isofemale lines from six populations: Zimbabwe (24 lines), Uganda (20 lines), Beijing (17 lines), Tasmania (19 lines), Netherlands (19 lines), and Pennsylvania (24 lines). The lines from the Netherlands were a gift courtesy of Zoltan Bochdanovits (Bochdanovits and de Jong 2003). The lines from Tasmania are a gift from Ary Hoffman. Lines from Beijing were provided by Chip Aquadro (Begun and Aquadro 1995). The lines from Zimbabwe (ZW) are from Victoria Falls, provided by Bill Ballard. The Uganda lines are a gift from John Pool.

Resequencing of PCR products

Genomic DNA was isolated from single male flies using Puregene DNA extraction kit (Puregene). We amplified fragments from the introns of four Y-linked genes: *kl-5*, *kl-3*, *kl-2*, and *ORY* totaling 7.8 kb in each of 123 fly lines. The fragments correspond to introns in each of the four genes: *kl-5* (*kl-5*_int8 forward primer 5′ ACTCTCGACCCA CACCTTTG and reverse primer 5′ GCTGCCAACTGATCCAA AAT; and *kl-5*_int10 forward primer 5′ TGTCAAATTTAA GAAAAAGAGAGAGG and reverse primer 5′ AGATTTGT CCTGCAGCTCATC), *kl-3* (*kl-3*_int2 forward primer 5′ CGTTTTGGCCATCCTAAAAA and reverse primer 5′ CTCC TTTGAATGTGGGCAAT; and *kl-3*_int6 forward primer 5′ GCCGAAATGGTCGATATGAT and reverse primer 5′ TGGATG CCAGTTTCCTTTTT), *kl-2* (*kl-2*_int1 forward primer 5′ GC AGCAAATAAAAGCGAAGC and reverse primer 5′ TGTAA ACCCAATACGCACGA; and *kl-2*_int2 forward primer 5′ TTTTTAAAACTACCAACCTCTGCT and reverse primer 5′ AATAAAAGCTCGCGAAACGA), and *ORY* (*ORY*_intu1 forward primer 5′ TTATAGCATTCCCTCTATTTTGCTG and reverse primer 5′ CAGTAAATCCAAAAATTGTCATTCC; and *ORY*_intu2 forward primer 5′ ATTCGGAGTTTACTTTGATA-CATGG and reverse primer 5′ ATCAAGCTGTTATCAAAA GTTCAGC). Primer pairs *kl-5*_int10 and *kl-2*_int2 were anchored in 20 and 29 bp of an exon, respectively. The coding sequences (totaling 49 bp) were trimmed for the analysis.

Primers were tested in both males and virgin females to confirm Y linkage of sequences. Conditions for PCR varied with template and primers and are available on request. Unincorporated nucleotides were removed using Exonuclease I/Shrimp alkaline phosphatase. PCR resequencing was performed using the ABI Prism Big Dye cycle sequencing kit according to the manufacturer's protocol and sequencing reactions were purified using a Sephadex column. Both the forward and reverse strand of each PCR product was sequenced using an ABI 3730 automated sequencer.

Raw sequence chromatograms were edited by eye using Sequencher v. 4.10.1 (Gene Codes, Ann Arbor, MI) and assembled into contigs where each putative polymorphic site was scrutinized. Fragments were resequenced when traces did not provide at least twice the coverage over each SNP. Sequences were exported, concatenated (for each analysis), and formatted using custom PERL scripts. Alignments were performed using Sequencher v. 4.10.1. Some sequences in some lines repeatedly resulted in either weak PCR bands or ambiguous chromatograms—those sequences were dropped from the analysis.

Because data quality issues encountered in amplifying repetitive regions of the genome are expected to falsely increase diversity and we find little diversity in our data set, we do not expect that our data suffer from many errors at the PCR or sequencing level. Nonetheless, because of the

repetitive nature of the Y chromosomes, every measure was taken to ensure that orthologous regions of the Y chromosome were amplified in each line: only a single band was amplified in males prior to sequencing, and there was no evidence of "heterozygous" sites in our traces, indicating that our primers annealled uniquely in the *D. melanogaster* genome. Any traces with ambiguous chromatograms were removed from the data set.

Polymorphism analysis

We estimated θ_w (population mutation rate per silent nucleotide site), θ_π (measure of nucleotide diversity per silent site), and Tajima's D (a summary of the frequency spectrum) using the program compute in the analysis package v. 0.8.0 associated with the libsequence version 1.7.0 library (Thornton 2003). Because of the considerable population structure in *D. melanogaster*, summary statistics for each population were calculated separately (Supporting Information, Figure S1; Table 1).

Recombination

We estimated the minimum number of recombination events using the RecMin software (http://www.stats.ox.ac.uk/~myers/RecMin/) as R_m (Hudson and Kaplan 1985) and R_h (Myers and Griffiths 2003).

X-linked polymorphism

All X-linked data were obtained from the literature. Briefly, Pool and Aquadro (2006) surveyed 10 lines of flies from a China and Uganda population at 4 noncoding X-linked loci at least 10 kb from coding loci in regions of high recombination (Table S1). Haddrill *et al.* (2005) surveyed the introns of 10 X-linked genes across regions of the X chromosome with high recombination in Zimbabwe, Netherlands, and Pennsylvania populations (Table S1). Kolaczkowski *et al.* (2011) surveyed variation across the genome of 19 individuals of a Tasmanian population of *D. melanogaster* using Illumina technology. For the Tasmania population, we used the estimate of pairwise diversity from Kolaczkowski *et al.* (2011) (Table S1). Our Y-linked data come from these same populations and, in most cases, the same lines.

Coalescent methods for tests of neutrality

All coalescent simulations were performed with custom C++ programs using the libsequence v. 1.7.0 library (Thornton 2003). To determine the significance of the reduction of π_Y, or $\Delta\pi$ (the reduction in variation on the Y, estimated as $[\pi_X - \pi_Y]/\pi_X$), we simulated 1×10^5 X-linked and Y-linked genealogies under the infinite sites model making the assumption that $\theta_Y = \theta_X/3$ (*i.e.*, we assume that there is one Y chromosome for every three X chromosomes in a population). We used the average θ_X estimated from Haddrill *et al.* (2005) for the following populations: Zimbabwe (0.01327); Netherlands (0.00414); and Pennsylvania (0.00557). For Uganda and Beijing, we used estimates of θ_X from Pool and Aquadro (2006), 0.012 and 0.0036, respectively (Table S1). Because comparable X-linked data (PCR resequence data from noncoding X-linked loci) are not currently available for Tasmania, we assumed a uniform range of θ_X values that encompass the Cosmopolitan samples that we have considered [$\theta_X \sim$ U(0.0020–0.0045)]. We generated P-values using the empirical cumulative distribution function (ecdf) in R. The false discovery rate (FDR) was estimated using the p.adjust package in R (Benjamini and Hochberg 1995).

Demographic models

Population bottlenecks: We used models of simple population bottlenecks to determine whether the Y chromosome polymorphism and frequency spectra from the Netherlands and Beijing populations could be explained by past demographic changes. The model describes a population that drops in size at time t_b ($t_b = t_r + d$, where t_r is the time to recovery from the bottleneck and d is the duration of the bottleneck) in the past according to severity parameter f ($f = N_b/N_0$, where N_0 is the N_e after the bottleneck moving forward in time, or the current population size) and recovers from the bottleneck at time t_r to N_0/N_A, where N_A is the ancestral population size before the bottleneck (Figure 3 and Figure S2). We also considered models where the bottlenecked population came from an ancestral population that was expanding in either a stepwise, or exponential manner (Figure S2). The parameter t_{grow} describes when the ancestral population began expanding in the past and was estimated in Li and Stephan (2006) using Zimbabwe data. Neutral mutations were placed on the tree according to a Poisson distribution with mean $\theta t/2$, where θ is the neutral mutation rate. For models incorporating expansion in the ancestral population, we used an average X-linked θ estimated by Li and Stephan (2006) of 0.0499. For models that just consider bottlenecks from a stable ancestral African population, we drew θ for each X-linked locus from a uniform distribution bounded by the 95% confidence intervals of θ_w obtained in Haddrill *et al.* [2005; $\theta \sim$U(0.01, 0.015)] estimated from the X-linked Zimbabwe population. In all models, we make the assumption that $\theta_Y = \theta_X/3$, $\rho_X/\theta = 10$ (Haddrill *et al.* 2005; Thornton and Andolfatto 2006) and that $\rho_Y = 0$. We performed simulations under six population bottleneck models consistent with parameter values estimated in the literature for Beijing and Netherlands populations (Table S3). The six models can be grouped into two types for each population: the B1 models that consider an older, shorter, and more severe bottlenecked European population (t_r=0.048, d=0.001, f =0.002, N_0/N_A = 0.125; Table S3; *e.g.*, Li and Stephan 2006) and B2 models that consider a longer bottleneck (t_r=0.004, d =0.018, f = 0.029, N_0/N_A = 1.0; Table S3; *e.g.*, Thornton and Andolfatto 2006). Each of the two model types then had different scenarios: the founding population came from an ancestral African population that was expanding exponentially, according to a stepwise function, or was of stable size. The beginning of the African expansion was assumed to be about

60,000 years ago (Li and Stephan 2006), and growth rates and times were scaled to match N_0 for each population (Table S3). All Beijing population models include a second, recent bottleneck from an ancestral European population using parameters comparable to those estimated by Laurent *et al.* (2011; B1: t_{r2} =0.019, d_2 =0.001, f_2 =0.02, N_0/N_{A2} = 0.254; B2: t_{r2} =0.002, d_2 =0.0001, f_2 =0.02, N_0/N_{A2} = 1.0). All time is in scaled units of $4N_{e0}$ generations and was scaled appropriately for the Y chromosome in the simulations ($t_Y = 3t_X$). N_0 corresponds to estimates from the literature under the different demographic models (Netherlands B1, 1.075×10^6, Li and Stephan 2006; B2, 2.4×10^6, Thornton and Andolfatto 2006; Beijing B1, 4.14×10^5, Li and Stephan 2006; B2, 2.4×10^6). Each demographic scenario was simulated 10,000 times and two-sided *P*-values for the X- and Y-linked empirical summary statistics were generated using the ecdf function in R. *P*-values were adjusted for false-positive rates using a Benjamini–Hochberg FDR rate and the *Q* and *P* values are reported in Table 3 and Table S4, respectively.

A demographic model for the Pennsylvania population was assigned under the assumption that North American populations were founded from an ancestral European population a few hundred years ago (B1, t_{r2} =0.0006, d_2 =0.0003, f_2 =0.001, N_0/N_{A2} = 1.0; B2, t_{r2} =0.0003, d_2 =0.0001, f_2 =0.001, N_0/N_{A2} = 1.0; Table S3). N_0 has not been specifically estimated for the Pennsylvania population; therefore, we assumed a simple bottleneck (where the population recovers to the pre-bottleneck size at t_{r2}) for both types of models (Pennsylvania B1, 1.075×10^6 ; B2, 2.4×10^6). It is important to note that these parameter values have not been inferred using X-linked or autosomal data; however, the results do not change significantly when the timing of this bottleneck is moved (data not shown). Because North American populations have higher nucleotide diversity than European populations (Haddrill *et al.* 2005; Caracristi and Schlötterer 2003), North American populations may have a more complicated demographic history than a simple bottleneck from a European ancestor. Caracristi and Schlötterer (2003) hypothesized that East coast North American populations are admixed with African alleles. To account for this possibility, we also simulated the Pennsylvania population under bottleneck models with varying degrees of African admixture (0, 5, 10, 15, and 20%). Two scenarios are possible: the European ancestor of North American populations was admixed with an African population or the African alleles entered the Pennsylvania population after the bottleneck, perhaps because of admixture with Caribbean populations (Caracristi and Schlötterer 2003). Because West coast populations appear more similar to European populations than East coast populations and because the North America–African genetic distance is smaller than European–African genetic distance (Caracristi and Schlötterer 2003), it seems more likely that admixture occurred after the North American bottleneck. Nonetheless, we simulated under both admixture scenarios: African–European admixture (AF–EU) and African–Pennsylvanian admixture (AF–PA), where admixture occurred at two fixed times corresponding to 300 years ago in the AF–EU model and 125 years ago for the AF–PA model (see Figure S3 and Table S6). These simulations were performed in ms (Hudson 2002) 10,000 times each for 10 X-linked and a single Y-linked locus, with the following assumptions:$\theta_Y = \theta_X/3$, $\rho_Y = 0$ and $\rho_X/\theta = 10$. Our conclusions for the Pennsylvania population do not change with the inclusion of admixture in the Pennsylvania bottleneck models (Table 4 and Table S6). The similarity between Tasmanian and Pennsylvania Y chromosomes indicates that they may originate from a similar European population; therefore, we also simulated Tasmanian populations with admixture in the event that the European ancestor had been admixed with an African population (Table S6).

Although the demographic history may differ between Uganda and Zimbabwe populations, we used the Zimbabwe demographic model as an approximation for the Uganda population. Likewise, we used demographic models for Pennsylvania, Netherlands, and Beijing populations as approximations for the Tasmania population and chose the most appropriate demographic model (*i.e.*, Pennsylvania; see Table 4, Figure 3, and Figure S2) for future simulations. Again, each demographic scenario was simulated 10,000 times and two-sided *P*-values for the X- and Y-linked empirical summary statistics were generated using the ecdf function in R. *P*-values were adjusted for false-positive rates using a Benjamini–Hochberg FDR rate and the *Q*-values are reported in Tables 3 and 4.

We repeated the neutral coalescent rejection sampling procedure described above except this time incorporating the demographic history described by the "best-fitting" model (the one that could describe the most aspects of the data) shown in boldface type in Tables 3 and 4, Table S3, and Table S4. The FDR was estimated using the p.adjust package in R (Benjamini and Hochberg 1995). *Q*-values <0.05 (corresponding to an FDR of 5%) are considered significant.

Population expansion: We used a simple model of population expansion to consider the possibility that a history of population growth in Africa could explain the reduction in diversity and skewed frequency spectrum on the Y chromosome in the African lines. The assumptions of the models were the same as those described for the population bottleneck models described above (*i.e.*, $\rho_X/\theta = 10$, $\rho_Y = 0$, $\theta_Y = \theta_X/3$). We ran one model with exponential growth and one with a simple stepwise growth (Figure 3 and Figure S2). The model with exponential growth describes a population that began expanding at time t_{grow} in the past at rate λ until the present time t_{rgrow}. We chose parameter values consistent with those estimated in Li and Stephan (2006) for the Zimbabwe population: $\lambda \sim U(15,25)$, $t_{rgrow} = 0$ (assuming $N_{e0} = 2.4 \times 10^6$). The model with stepwise population growth describes a population that expanded to size N_0/N_A at time t_{grow} in the past. Again, we chose parameter values consistent

Table 1 Summary statistics of Y chromosome variation in each population

Population	n	m	S	S_1	π	θ_w	$\Delta\pi$	π/π_0	D_{Taj}
Zimbabwe	24	7729	3	3	0.00323	0.01039	**0.9976*****	0.0074	−1.733*
Uganda	20	7523	2	2	0.00266	0.0075	**0.9978******	0.0067	−1.513*
Beijing	17	7618	7	4	0.01696	0.02772	0.9529**	0.1428	−1.368
Tasmania	19	7544	7	4	0.01370	0.02667	0.9355**	0.1951	−1.631*
Netherlands	19	7519	6	5	0.01505	0.02425	0.9637***	0.1103	−1.305
Pennsylvania	24	7172	6	6	0.00849	0.02365	0.9848***	0.0460	−2.048*

N, number of lines surveyed within each population. m, number of aligned bases considered in the summary statistic calculations. S, number of segregating sites, including indels. S_1, number of singletons or polymorphisms found only once in the sample. π, average percentage pairwise nucleotide diversity per 100 sites. θ_w, average percentage nucleotide diversity per site. $\Delta\pi$ is calculated as $(\pi_X - \pi_Y)/\pi_X$. Assuming k is 1/3 (equal numbers of breeding males and females), the neutral expectation of $\Delta\pi$ is ~0.667. π/π_0 is the reduction in π compared to the neutral expectation of π of the Y chromosome based on coalescent simulations of 10^5 neutral genealogies. D_{Taj}, Tajima's D statistic to summarize the frequency spectrum. *, $P < 0.05$; **, $P \sim 1 \times 10^{-3}$;***, $P \sim 1 \times 10^{-4}$.

with those estimated in Li and Stephan (2006) for the Zimbabwe population: $N_0/N_A \sim U(2.0,11.5)$, $t_{grow} = 0.0833$, $t_{rgrow} = 0$. All time is in scaled units of $4N_{e0}$ generations and was adjusted for the Y chromosome in the simulations ($t_{grow} = 0.2499$). Each demographic scenario was simulated 10,000 times and two-sided P-values for the X- and Y-linked empirical summary statistics were generated using the ecdf function in R. We used these same models to obtain P-values for the Uganda population, which also shows evidence for population expansion although the true parameter values may differ between Zimbabwe and Uganda. The FDR was estimated using the p.adjust package in R (Benjamini and Hochberg 1995).

Estimating the reduction in N_e on the Y chromosome: To identify the minimum reduction in N_{eY} required to produce the $\Delta\pi$ observed, we used an approximate Bayesian computation (ABC) approach with rejection sampling conditional on $\Delta\pi$ and π_Y (Przeworski 2003; Thornton and Andolfatto 2006) . We estimate the reduction in effective population size on the Y chromosome compared to the X chromosome (N_{eY}/N_{eX}) as the posterior probability of the scaling factor of θ, k. We assumed that population recombination rates follow $\rho_X/\theta = 10$ and $\rho_Y = 0$. The observed data in this case are the Y chromosome polymorphism data collected in this study, the X-linked polymorphism data compiled from the literature for the Zimbabwe (10 loci), Netherlands (10 loci), Pennsylvania (10 loci), Uganda (4 loci), and Beijing populations (4 loci). We assumed that the Tasmanian and Pennsylvanian populations have similar demographic histories because the Americas and Australia were founded around the same time (David and Capy 1988).

The rejection sampling algorithm is as follows:

1. Draw θ_X,ρ, and k from prior distributions, where θ is the X-linked population mutation rate, ρ is the X-linked population recombination rate, and k is the scaling factor for θ_Y (k represents N_{eY}/N_{eX}).
2. Estimate θ_Y as $k\theta_X$.
3. Simulate genealogies for $nsam_X$ independent X-linked loci and one Y-linked locus under the neutral coalescent model based on the empirical sample size for each locus, where $nsam_X$ is the number of loci in the empirical data set.
4. Continue if $|\theta_{X\text{-observed}} - \theta_{X\text{-simulated}}| \leq \varepsilon_\theta$
5. Calculate summary statistics for the simulated genealogies according to θ_X or θ_Y, assuming an infinite-sites mutation model.
6. Accept or reject the chosen parameter values conditional on $|\Delta\pi_{observed} - \Delta\pi_{simulated}| \leq \varepsilon_\pi$, and $|\pi_{Y\text{-observed}} - \pi_{Y\text{-simulated}}| \leq \varepsilon_\pi$ and record accepted parameter values.
7. Return to step 1 and continue simulations until 1000 samples from the joint posterior probability distribution are collected.

The prior distributions used were $\theta_X \sim U(0.01, 0.15)$ and $k \sim U(1 \times 10^{-7}, 1)$. The tolerance parameters, ε_θ and ε_π, were set to 80 and 2% of the observed θ, and $\Delta\pi$ or π, respectively. The same simulations were performed under a standard neutral model. For the Pennsylvania population admixture models, we used the same rejection sampling scheme, except it was performed using custom Perl scripts and ms (Hudson 2002).

Positive selection

We estimated the time to the most recent hard selective sweep in a coalescent framework using ABC with rejection sampling conditional on multiple summaries of each Y-linked data set. Because there is no crossing over on the *D. melanogaster* Y chromosome, selective sweeps were modeled as absolute bottlenecks, where at time t_{sweep}, all remaining lineages coalesced. We modeled Zimbabwe and Uganda populations separately. We incorporated population demographic histories into the selective sweep simulations so that the X-linked and Y-linked loci coalesce in the simulations according to the appropriate demographic model until the sweep, at which time all remaining Y-linked lineages coalesced.

The rejection sampling algorithm is as follows:

1. Draw θ_X and t_{sweep} from prior distributions, where θ_X is the X-linked population mutation rate, ρ is the X-linked population recombination rate, and k is the scaling factor for θ_Y (θ_Y is the population mutation rate for the Y chromosome and k represents N_{eY}/N_{eX}).
2. Estimate θ_Y as $k\theta_X$.

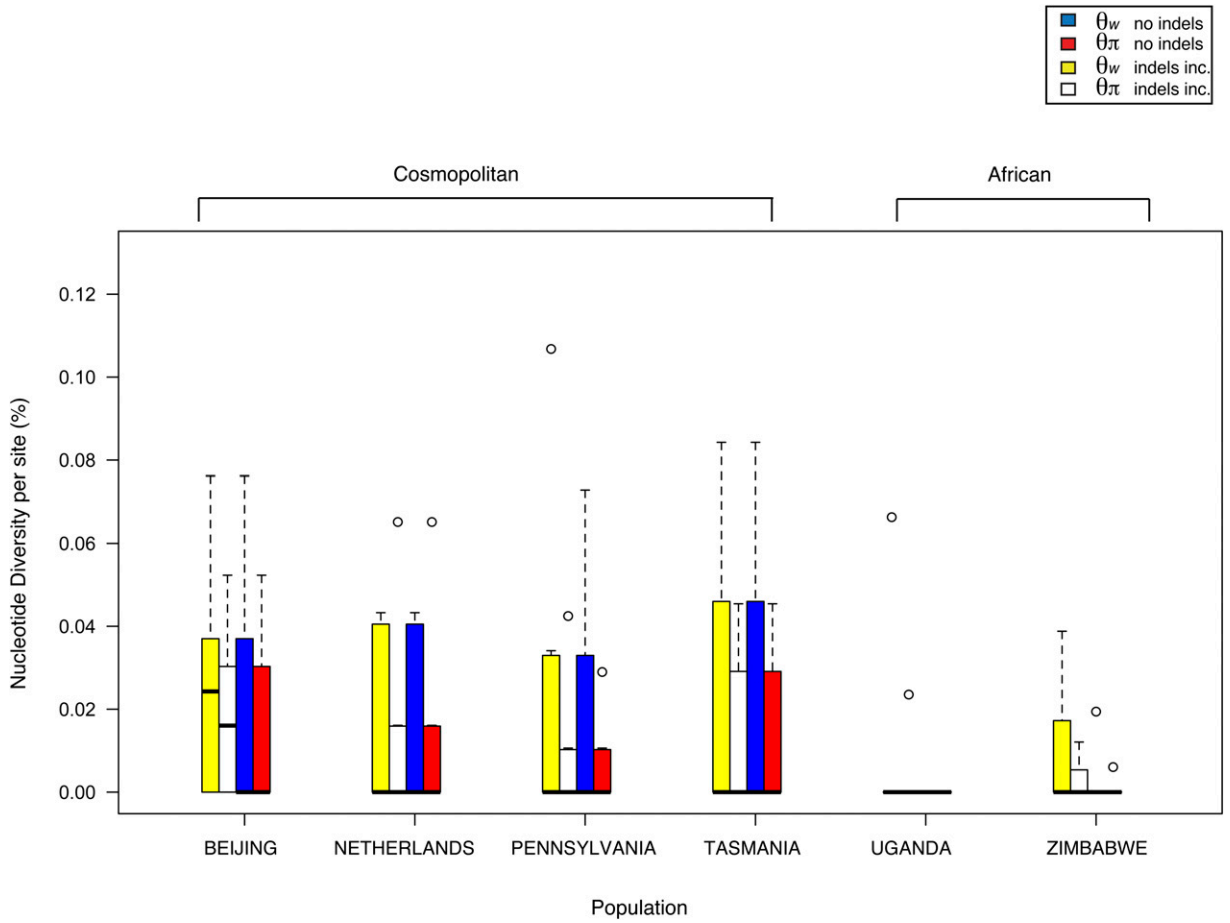

Figure 1 Levels of nucleotide variation by population. Standard box plots showing Watterson's estimator of nucleotide diversity per site from the number of segregating sites (θ_w) and the pairwise nucleotide diversity per site (θ_π) for each Y-linked fragment sequenced in the four non-African and two African populations. The horizontal lines interrupting the boxes indicate the median, the whiskers extend to the most extreme data point no more than 1.5 times the interquartile range from the box, and open circles indicate outliers. Estimates were calculated both including and excluding insertion–deletion mutations (indels) and only the analysis including indels is considered for the rest of the analysis.

3. Simulate genealogies for nsam_X independent X-linked loci and one Y-linked locus under the neutral coalescent model based on the empirical sample size for each locus, where nsam_X is the number of loci in the empirical data set.
4. Continue if $|\theta_{\text{observed}} - \theta_{\text{simulated}}| \leq \varepsilon_\theta$ for both the X- and Y-linked loci.
5. Calculate summary statistics for the simulated genealogies according to θ_X or θ_Y, assuming an infinite-sites mutation model.
6. Accept or reject the chosen parameter values conditional on $|\Delta\pi_{\text{observed}} - \Delta\pi_{\text{simulated}}| \leq \varepsilon$, $|D_{\text{Taj-Yobserved}} - D_{\text{Taj-Ysimulated}}| \leq \varepsilon$, $|S_{1\text{-Yobserved}}\text{-}S_{1\text{-Ysimulated}}| \leq \varepsilon$ and $|\pi_{\text{Yobserved}} - \pi_{\text{Ysimulated}}| \leq \varepsilon_\pi$ and record accepted parameter values.
7. Return to step 1 and continue simulations until 10,000 samples from the joint posterior probability distribution are collected.

The prior distributions used were: $\theta_X \sim \text{U}(0.0005, 0.03)$, and $t_{\text{sweep}} \sim \text{U}(0, 10^{-6})$. The tolerance parameter, ε, was set to 35% of the observed $\Delta\pi$, D_{Taj}, π_Y and S_1 (number of singletons).

Results

Patterns of polymorphism

When compared to the X chromosome, the Y chromosome harbors less variation than expected under the standard neutral model (Table 1). Contrary to patterns of variation on the X chromosome and autosomes, African populations have significantly less Y-linked variation than Cosmopolitan populations (Table 1 and Figure 1; $P = 0.002$ considering indels and $P = 0.008$ without considering indels, Mann-Whitney U test). Specifically, there is a ~400-fold reduction in pairwise nucleotide diversity on the Y chromosome compared to the X chromosome in Africa (414-fold for Zimbabwe and 383-fold for Uganda), whereas there is a 6-fold (Tasmania) to 66-fold (Pennsylvania) reduction in pairwise nucleotide diversity outside of Africa. The estimated reduction in variation on the Y chromosome compared to the neutral

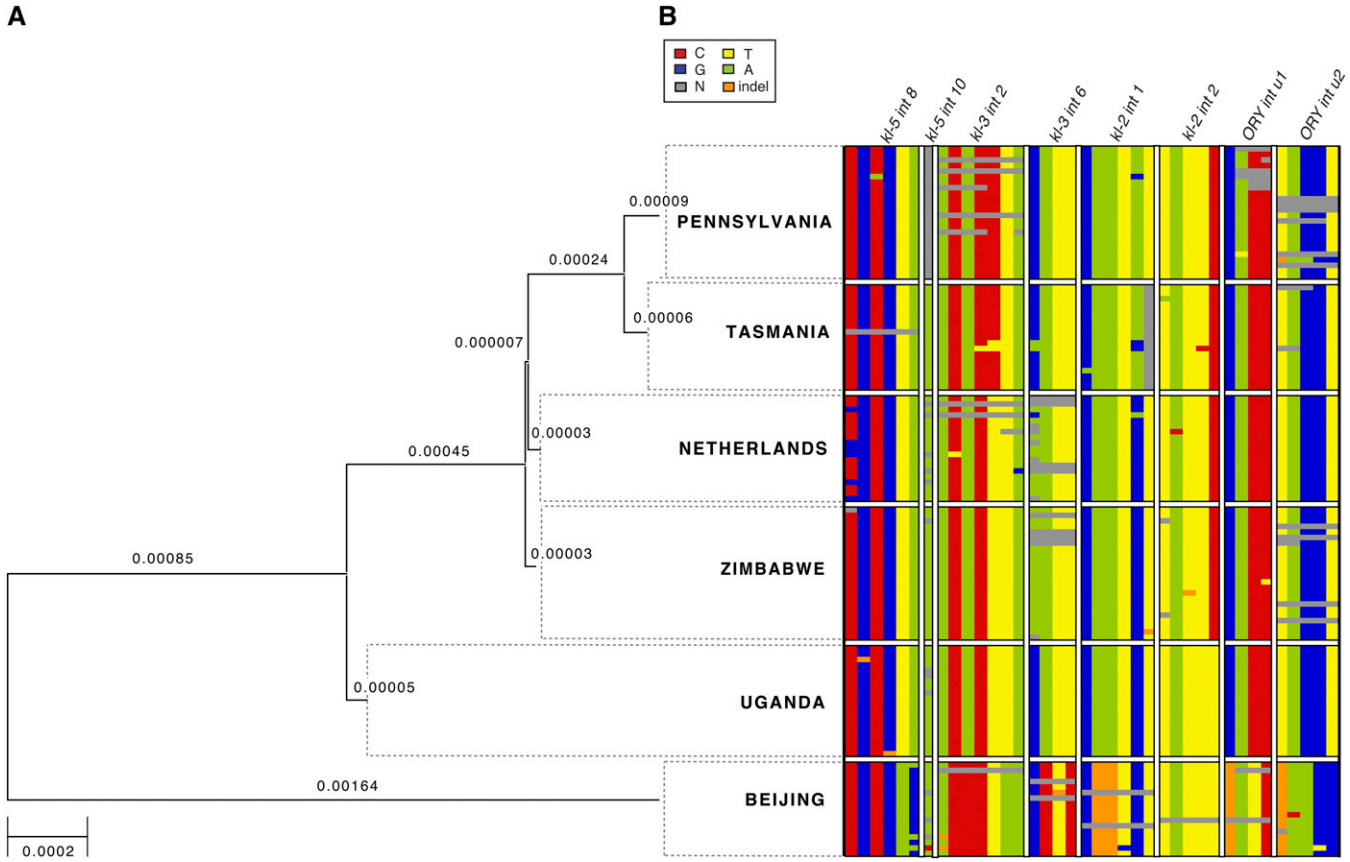

Figure 2 Population structure. (A) Nucleotide Divergence between populations. An unrooted population distance tree generated by a neighbor-joining analysis of the average number of nucleotide substitutions per site between populations (*Dxy*) is diagrammed. (B) Y chromosome polymorphism table. The sequenced fragments of *kl-5*, *kl-3*, *kl-2*, and *ORY* are separated by solid lines. The populations are separated by white space.

expectation (π/π_0) is ~0.7% in African populations and varies between 4.6 and 19.5% outside of Africa (Table 1 and Table S2). We scrutinized each segregating site in our data set: most were sequenced twice on the forward strand and twice on the reverse strand and had independent confirmation with a second DNA isolation and resequencing. Interestingly, there is a significant bias in the frequency of indel polymorphism (usually a single base pair) in African compared to Cosmopolitan populations. Among African Y chromosome variants, 80% (4/5) of the segregating sites are indels, whereas among Cosmopolitan Y chromosome variants only 8.3% (2/24) of the segregating sites are indels (Figure 2B; $P = 0.003$, Fisher's Exact Test). We consider the full data set including indels for the remainder of the article.

Population structure

Our analysis of population structure using the pairwise average number of nucleotide substitutions per site (D_{xy}), F_{st}, K_{st}* (measure of population differentiation analogous to F_{st}) and S_{nn} (a nearest neighbor statistic), yielded surprising results. The African populations do not cluster together: Zimbabwe instead clusters with the Netherlands population. Moreover, Pennsylvania and Tasmania Y chromosomes are nearly indistinguishable (Figure 2 and Table 2). Consistent with several surveys of population differentiation showing considerable population structure in Asia (Hale and Singh 1991; Baudry *et al.* 2004; Pool and Aquadro 2006; Schlötterer *et al.* 2006), the Beijing Y chromosome is highly differentiated from the rest of the populations. There is no evidence of substructuring within the Beijing population on the Y chromosome (Figure 2B). Outside of the clustering of Zimbabwe and Netherlands Y chromosomes, these results are consistent with the population structure of autosomal, X chromosomal, and mitochondrial loci from these or similar populations (Hale and Singh 1991; Baudry *et al.* 2004; Haddrill *et al.* 2005; Pool and Aquadro 2006; Schlötterer *et al.* 2006). Genomic sequences from non-Y-linked loci exclude the possibility that the similarity of Netherlands and Zimbabwe Y chromosomes are explained by contamination in our fly lines (A.G. Clark, unpublished results).

Evidence for intrachromosomal recombination

We used the program RecMin to identify any possible recombination events in our data set (Myers and Griffiths 2003). Three different methods (R_m, R_h, and R_s) yield the inference that at least one recombination event has occurred

Table 2 Genetic differentiation between populations

Populations	Pennsylvania	Beijing	Netherlands	Tasmania	Uganda	Zimbabwe
Pennsylvania	—					
Beijing	0.7286	—				
Netherlands	0.4213	0.5122	—			
Tasmania	0.0088	0.5567	0.3088	—		
Uganda	0.2863	0.4393	0.1389	0.2157	—	
Zimbabwe	0.4055	0.4504	0.0769	0.3106	0.1250	—

Average pairwise *Fst* estimates between populations are shown.

in the sampled regions. This event was confirmed and other possible recombination events were identified by eye (Figure 2B). Although the intronic regions of the Y chromosome that we have amplified are repetitive, the PCR and sequence trace reads indicate that we have amplified single, orthologous regions of the Y chromosome in each fly line. Recombination via crossing-over does not occur in most *Drosophila* species males; however, gene conversion events have been detected on the *D. simulans* Y chromosome between duplicate genes (Kopp *et al.* 2006). While the gene regions surveyed are not known to be duplicated in *D. melanogaster*, the introns are repetitive, and it is possible that there are several similar sequences throughout the Y chromosome offering opportunities for intrachromosomal recombination. These recombination events are likely to reflect mitotic recombination in the male germline.

Population demographic history

We find a reduction in variation on the Y chromosome compared to the X chromosome and this is statistically significant under a standard neutral model, assuming equal effective numbers of breeding males and females (Table 1). However, Cosmopolitan populations of *D. melanogaster* clearly violate the assumption of a constant population size: *D. melanogaster* originated in Africa and populated Europe approximately 16,000 years ago and American and Australian populations were founded from a European ancestor only in the past few hundred years (David and Capy 1988; Caracristi and Schlötterer 2003). Moreover, African populations may have a history of population expansion that began before the out-of-Africa migration events (60,000 years ago; Li and Stephan 2006).

We considered the effect of population demographic history on Y chromosome evolution by simulating data under several different models that have been inferred using empirical data from the Zimbabwe, Netherlands, and Beijing populations (Li and Stephan 2006; Thornton and Andolfatto 2006; Laurent *et al.* 2011; Figure 3). We found that ancestral growth in the Zimbabwe population can account for values of Tajima's *D* similar to what we observe on the Y chromosome; however, the reduction in variation on the Y chromosome compared to the X is actually smaller under growth models than under the standard neutral model (Table 3 and Table S4). A history of population expansion alone therefore cannot explain patterns of variation on the Y chromosome in Zimbabwe. We see similar results when we assume that the Uganda population experienced the same degree of population expansion (Table 4 and Table S5).

All Cosmopolitan populations of *D. melanogaster* have a history of population bottlenecks. We considered several population bottleneck models inferred from empirical data in the literature and asked whether the summary statistics we observe on the Y chromosome could be explained by demographic history alone for the Netherlands and Beijing populations. Population bottleneck models with different bottleneck parameters, with or without population expansion in the ancestral African population, predict different patterns of Y chromosome variation (summarized in Table 3 and Table S4). For example, assuming a severe bottleneck that was fairly short in duration (*i.e.*, B1 models; Table 3 and Table S4), while the observed Tajima's *D* on the Y chromosome is expected under some of these models, rarely generates a reduction in variation (measured as $\Delta\pi$) as large as what is observed in the Netherlands population. In contrast, considering a strong bottleneck with an order of magnitude longer duration (*i.e.*, B2 models; Table 3 and Table S4), so that the population has only recently recovered, makes it likely that one would observe Tajima's *D* as low and $\Delta\pi$ of the magnitude seen in the Netherlands and Beijing populations. These results highlight the utility of considering multiple genomic locations in inferring demographic history. Our Y-linked data from the Netherlands and Beijing can thus be explained purely with demographic models without the need to invoke selection. Although a detailed demographic model has not been fitted to X-linked or autosomal data from the Pennsylvania population, we considered an arbitrary model where a recent bottleneck from a European population occurred in the past few hundred years ($t_b = 0.002$). This model can explain all major aspects of the Y chromosome data set (Table 4 and Table S5); however, D_{TajY} in this model just barely misses the 5% FDR cutoff (*P*-value = 0.0086; *Q*-value = 0.08). Although it is important to note that the timing of the North American bottleneck was not inferred from empirical data, because this bottleneck is so recent in history, the precise timing has little impact on our results (data not shown). All major aspects of the Tasmania population can be explained by demographic history when assuming that Tasmania experienced the same population bottlenecks as each of the Netherlands, Beijing, and Pennsylvania populations (Table 4 and Table S5).

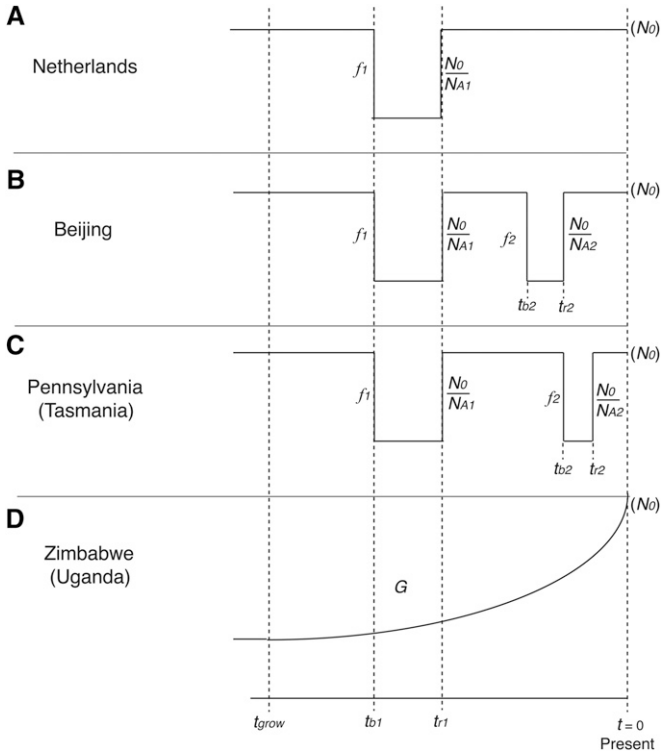

Figure 3 Simple demographic models describing the models of simple population bottlenecks and population expansion and their parameters. For choice of the best demographic model and parameter values, see Tables 3 and 4 and Table S1. (A) The Netherlands population bottleneck describes a population that was reduced in size to $f_1 \times N_0$ at time t_{b1} and recovering to size N_0/N_{A1} at time t_{r1}. (B) The Beijing population bottlenecked twice, where the first was the European bottleneck from an ancestral African population (same as Netherlands bottleneck). (C) The Pennsylvania population experienced a similar bottleneck as did Beijing, except that the second bottleneck occurred more recently. (D) The demographic model for the Zimbabwe population describes an ancestral African population that began expanding at time t_{grow} at rate G until the present time. All time is in scaled units of $4N_e$ generations. We used the Pennsylvania double bottleneck (C) for the Tasmania population and the Zimbabwe expansion model (D) for the Uganda populations. Not shown in the diagram are admixture events (see text and Figure S3 for bottleneck with admixture models).

Because North American populations have more diversity than European populations and may be admixed with African alleles (Caracristi and Schlötterer 2003), we also explored two admixture scenarios: (1) an admixed European population founded North American populations of *D. melanogaster* (model AF-EU; Figure S3A and Table S6) and (2) the Pennsylvania population admixed an African or African-like population after the North American bottleneck (Figure S3B; Table S6; Caracristi and Schlötterer 2003). We simulated the Pennsylvania population under each admixture scenario with 0, 5, 10, 15, and 20% admixture with African alleles at two fixed times corresponding to ~300 years ago for the AF–EU models and ~125 years ago for the AF–PA model. Because the Tasmania and Pennsylvania populations may have originated from a similar European ancestor, we also simulated under the same AF–EU admixture models for the Tasmania population. We get qualitatively similar results under all the admixture models when compared to models with no admixture: observed estimates of D_{Taj} and $\Delta\pi$ are expected under each of these models for both the Pennsylvania and Tasmania populations at a 5% FDR (Table 4 and Table S6). However, the empirical $\Delta\pi$ and D_{Taj} (X and Y linked) in the present data set appear more likely under models with a small degree of admixture between the Pennsylvania population and an African-like population (*e.g.*, AF–PA $p_{AF=}$0.05; *P*-value D_{TajX} =0.6520; *Q*-value D_{TajX} =0.7964; *P*-value D_{TajY} =0.8186; *Q*-value D_{TajY} =0.8396; *P*-value$_{\Delta\pi}$=0.7072; *Q*-value$_{\Delta\pi}$=0.8082; Table 4 and Table S6). While $\Delta\pi$ and D_{TajX} are expected under the models with no admixture, a D_{TajY} as low as we observe in Pennsylvania is not expected under the recent bottleneck model, but this does not survive the 5% FDR cutoff (AF–PA $p_{AF=}$0; *P*-value D_{TajX} =0.5014; *Q*-value D_{TajX} =0.7161; *P*-value D_{TajY} =0.009; *Q*-value D_{TajY} =0.096; *P*-value$_{\Delta\pi}$=0.3418; *Q*-value$_{\Delta\pi}$=0.7933; Table S6, also see the PA–B2 model in Table 4). Natural selection on the Y chromosome may be implicated when the X-linked data fit well to a plausible demographic model, but the Y data clearly reject the same model. We looked for any incongruence between the X- and Y-linked data in our simulations. Although a model with 20% admixture between the European ancestor of North American populations and an African population produced summaries completely consistent with the X-linked data and less consistent with the Y-linked data, none of the summaries could be statistically rejected and survive a correction for the FDR (AF–EU $p_{AF=}$0.20; *P*-value D_{TajX} =0.4988; *Q*-value D_{TajX} =0.7161; *P*-value D_{TajY} =0.0114; *Q*-value D_{TajY} =0.0960; *P*-value$_{\Delta\pi}$=0.0792; *Q*-value$_{\Delta\pi}$=0.2192; Table 4 and Table S6).

Reduction in N_e on the Y chromosome

We estimated the reduction in N_e on the Y chromosome as k (N_{eY}/N_{eX}) using an ABC approach. We ran simulations both under the best demographic model (boldface in Tables 3 and 4) and a standard neutral model. Historical changes in population size affect the ratio of Y to X chromosomes. Under neutrality the expected k is 1/3, but without incorporating demographic history the estimated k is an order of magnitude (Cosmopolitan) or 2 orders of magnitude (African) less than this (Table S7). When we account for the demographic history of each population, the maximum *a posteriori* (MAP) estimate of k is not significantly different from 1/3 in any Cosmopolitan population but is significantly <1/3 in both African populations (MAP k_{ZW} = 0.0041, $P = <10^{-3}$; MAP k_{UG} = 0.0037, $P = <10^{-3}$; Figure 4 and Table S7).

Positive selection in Africa

A history of population expansion in Africa can explain summaries of the frequency spectrum, but cannot explain the ~400-fold reduction in variation on the Zimbabwe and Uganda Y chromosomes compared to the X chromosomes.

Table 3 Performance of different inferred demographic models for the Netherlands, Beijing and Zimbabwe populations

Population	Demographic models[a]	Median D_{TAJ} $(Q)^b$ X	Median D_{TAJ} $(Q)^b$ Y	$\Delta\pi$ $(Q)^b$
Netherlands	B1	−0.4642	−1.4102	*0.7907*
		(0.9257)	(0.9007)	*(<10⁻⁴)*
	B2	**0.5209**	**-1.5108**	**0.9470**
		(0.6966)	**(0.9007)**	**(0.6966)**
	B1–AF	0.349	−0.1418	0.8641
		(0.5778)	(0.4205)	*(0.0378)*
	B2–AF	0.7600	−1.7177	0.9512
		(0.5964)	(0.6966)	(0.6966)
Beijing	B1	−0.2792	−0.5183	*0.6145*
		(0.7961)	(0.1413)	*(0.0053)*
	B2	**0.4050**	**-1.2509**	**0.9469**
		(0.9587)	**(0.5925)**	**(0.9587)**
	B1–AF	−0.2988	−0.5550	*0.6359*
		(0.5925)	(0.1208)	*(<10⁻⁴)*
	B2–AF	0.5572	−1.4466	0.9515
		(0.8599)	(0.7961)	(0.9654)
Zimbabwe	**Exp**	**-0.7658**	**-0.8480**	***0.5378***
		(0.4471)	**(0.0812)**	***(<10⁻⁴)***

The X and Y-linked data from each population were tested under the different types of demographic model (bottlenecks, B1 and B2; bottlenecks with ancestral growth, B1-AF and B2-AF; Exponential growth, Exp).

[a] Results from simulated genealogies for 10 independent X-linked loci, and a single Y-linked locus (assuming $\theta_Y = 1/3\theta_X$) under bottleneck or exponential growth models specified in the Methods and Table S2.

[b] Median D_{Taj} (for X and Y-linked loci) and $\Delta\pi$. Reported are *Q*-values estimated from two-sided *P*-values (see Table S4 for *P*-values). $Q < 0.05$ corresponds to a FDR <5%. The model chosen for use in subsequent simulations is in boldface type.

This drastic reduction in variation could, however, be explained by a history of recent selective sweeps. We simulated selective sweeps using coalescent simulations incorporating population expansion in Africa and assuming constant population size. To estimate the time since the most recent selective sweep in African populations, we used an ABC approach and conditioned on several aspects of the Y chromosome polymorphism data ($\Delta\pi$, π_Y, θ_w, S_1, and D_{Taj}). Under a population expansion model, we estimate the time to the most recent selective sweep to be similar in Zimbabwe (MAP t_{sweep} = 0.00098 [0.00042–0.00226] $4N_{e0}$ generations; Figure 5) and Uganda (MAP t_{sweep} = 0.00082 [0.00033–0.00214] $4N_{e0}$ generations; Figure 5) populations. Assuming that the N_e of African populations is 2.4×10^6 (Thornton and Andolfatto 2006) and 10 generations per year, the last sweep on the Zimbabwe Y chromosome occurred 236 years ago ($0.25 \times (4N_{e0}t/g)$; 101–542 years) and the last sweep on the Uganda Y chromosome occurred 197 years ago (80–515 years). Because population expansion reduces the difference in variation between the X and Y chromosomes and because we condition on $\Delta\pi$, under a model of constant population size, the time of the sweep is pushed back for both populations (Zimbabwe MAP t_{sweep} = 0.00388 [0.00168–0.00956] $4N_{e0}$ generations or 931 (404–2293) years; Uganda MAP t_{sweep} = 0.003122 [0.001271–0.00917] $4N_{e0}$ generations or 749 (305–2226) years; Figure 5). The marginal posterior distribution on t_{sweep} provides 95% credible intervals that overlap between ~400 and 550 years ago for Zimbabwe and ~300–515 years ago for Uganda.

Discussion

While the gene-poor and heterochromatic *D. melanogaster* Y chromosome clearly harbors functional variation, it has very low levels of nucleotide polymorphism. The effect of the Y chromosome on male fitness (Chippindale and Rice 2001), heat tolerance (Rohmer *et al.* 2004), and global (Lemos *et al.* 2008, 2010) or germline-specific (Zhang *et al.* 2000) gene expression may be driven in part by variation in heterochromatin content (Lemos *et al.* 2010), specific *trans*-activators of gene expression (Zhang 2000), or in rDNA copy number (Paredes *et al.* 2011), rather than by variation at the single nucleotide level.

Evidence for admixture between European and African populations?

Surprisingly, the *D. melanogaster* Y chromosome harbors far less polymorphism than expected in African populations compared to Cosmopolitan populations. While overall the Y chromosome recapitulated the history of these populations based on autosomes, the X chromosome, and mitochondria, we found one surprising relationship between Y chromosomes from African and European Y chromosomes: Zimbabwe Y chromosomes are more similar to Y chromosomes from the Netherlands than Uganda. There are two explanations for the similarity between Zimbabwe and Netherlands Y chromosomes:

1. The ancestral African population that founded the European populations could have been similar to Zimbabwe rather than Uganda. However, a shared site between the

Table 4 Performance of demographic models for Pennsylvania, Tasmania and Uganda populations

Demographic models[a]		Median D_{Taj} (Q)[b] X	Y	$\Delta\pi$ (Q)[b]
Pennsylvania	B1	0.6559	−1.159	0.9663
		(0.5138)	(0.1605)	(0.6498)
	B2	0.8364	0.1386	0.9609
		(0.5138)	(0.0803)	(0.5138)
	B1–AF	0.7634	−1.7325	0.9663
		(0.5138)	(0.5965)	(0.5138)
	B2–AF	1.031	0.1390	0.9636
		(0.5138)	(0.1422)	(0.4536)
	AF–EU	0.8321	0.1143	0.9885
	p_{Af} =0.05	(0.7161)	(0.0960)	(0.3299)
	AF–EU	0.9673	0.1334	0.9906
	p_{Af} =0.20	(0.7161)	(0.0960)	(0.2192)
	AF–PA	**−1.5996**	**−1.5455**	**0.8807**
	p_{Af} =0.05	**(0.7964)**	**(0.8396)**	**(0.8082)**
	AF–PA	−0.6201	−1.0113	0.7677
	p_{Af} =0.20	(0.9638)	(0.6785)	(0.1634)
Tasmania	**PA–B2**	—	**0.1386**	**0.9609**
			(0.8400)	**(0.9537)**
	BEI–B2	—	−1.2509	0.9469
			(0.9538)	(0.9537)
	NE–B2	—	−1.5108	0.9470
			(0.9538)	(0.9537)
Uganda	**ZW–Exp**	**−0.7658**	**−0.8480**	***0.5378***
		(0.9350)	**(0.2117)**	***(<10⁻⁴)***

The X- and Y-linked data from each population were tested under the different types of demographic model (bottlenecks, B1 and B2; bottlenecks with ancestral growth, B1–AF and B2–AF; admixture models AF–EU and AF–PA, exponential growth, Exp). The Tasmania and Uganda populations were tested with models based on other populations (PA, Pennsylvania; BEI, Beijing; NE, Netherlands; ZW, Zimbabwe).

[a] Results from simulated genealogies for 10 independent X-linked loci and a single Y-linked locus (assuming $\theta_Y = 1/3\theta_X$) under hypothetical bottleneck or exponential growth models specified in the *Materials and Methods* and Table S2. Three possible demographic models corresponding to the population histories of Pennsylvania, Beijing, and Netherlands populations were simulated for Tasmania.

[b] Median D_{Taj} (for X- and Y-linked loci) and $\Delta\pi$. An empirical estimate of D_{Taj} for the X chromosome in Tasmania has not been published. Reported are Q-values estimated from two-sided P-values. $Q < 0.05$ corresponds to a FDR <5%. The model chosen for use in subsequent simulations is in boldface type.

Uganda and Beijing Y chromosomes is discordant with this interpretation. Moreover, Pool and Aquadro (2006) hypothesized that the ancestral source population for the Out-of-Africa migration to Europe was similar to a Uganda population based on X-linked variation data.

2. The similarity of the Y chromosomes from Zimbabwe and Netherlands may be evidence of admixture between these populations. Several studies have documented recent admixture between Cosmopolitan populations and populations from Congo (Capy *et al.* 2000) and Zimbabwe (Kauer *et al.* 2003), Eritrea, Gabon, and South Africa (Pool and Aquadro 2006) and between Caribbean/South American populations and West African populations (Caracristi and Schlötterer 2003).

The present data set has too few informative polymorphisms to distinguish between these possibilities. An ancestral European population colonized the Americas and Australia in the past few hundred years (Caracristi and Schlötterer 2003; David and Capy 1988). The similarity between the Tasmanian and Pennsylvanian Y chromosomes suggests that these populations originated from the same, or a similar, ancestral population in Europe.

Recent selection on African Y chromosomes

Variation is reduced a staggering 400-fold on the Y chromosome compared to the X chromosome in our African population samples, whereas fold reduction varies between 6- and 66-fold outside of Africa. This is in stark contrast with variation across the autosomes, the X chromosome, and mitochondrial loci, which all harbor more variation in African than outside of Africa. While a demographic history of population expansion in Africa can explain some aspects Y chromosome frequency spectra in Zimbabwe and Uganda, it cannot explain the drastic reduction in variation on the Y chromosome. There are three possible explanations for the dearth of Y-linked variation on African Y chromosomes:

1. A higher variance in male reproductive success in Africa relative to outside of Africa reduces the effective population size of the Y chromosome and can lead to lower levels of variability (Nunney 1993; Charlesworth 2001). Indeed, Hutter *et al.* (2007) found evidence for skewed sex ratios in African populations of *D. melanogaster*. While this is likely to make some contribution to this pattern, the maximum reduction in N_e on the Y compared to the X resulting from sexual selection is about ninefold (Caballero 1995; Charlesworth 2001), whereas the estimated reduction in N_e on the Y compared to the X in Africa is 256-fold (k = 0.0039) and 370-fold (k = 0.0027) in Zimbabwe and Uganda, respectively (Figure 1; Figure 4; Table S7). A higher variance in male reproductive success in Africa is therefore not likely to be sufficient to explain this pattern.
2. Selection against deleterious mutations could reduce levels of variation on the Y chromosome relative to the X chromosome, but rarely to the degree observed in African populations (Kaiser and Charlesworth 2009). The relatively small number of genes on the Y chromosome offers few targets of purifying selection: we estimate that there are ~19 kb of nonsynonymous nucleotide sites on the *D. melanogaster* Y chromosome. For a target of this size in a nonrecombining region of the genome, we expect to see a reduction in variation an order of magnitude smaller (Kaiser and Charlesworth 2009) than we observe if background selection were the only force shaping patterns of polymorphism on the Y chromosome. Moreover, the strength of background selection is not expected to vary among these populations and, therefore, background selection is unlikely to explain differences in the degree of >reduced diversity on the Y chromosome between populations.
3. African Y chromosomes may have experienced selective sweeps in the recent past. We estimate that the time to the most recent selective sweep in Zimbabwe and Uganda is similar—in the past ~200–500 years (Figure

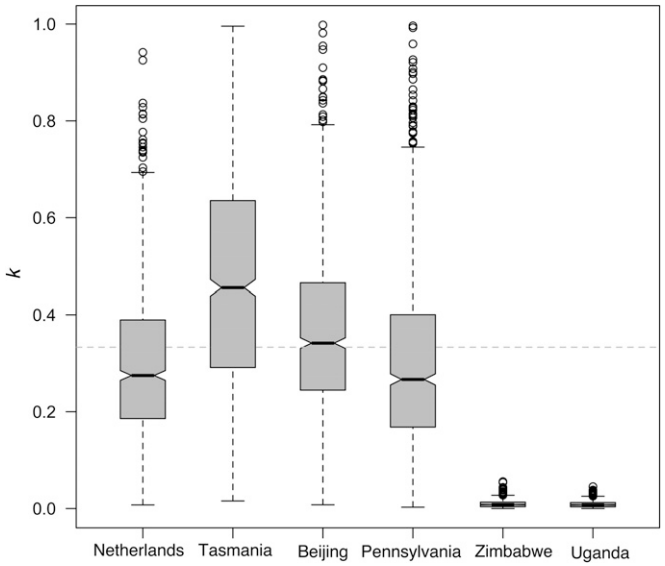

Figure 4 The estimated reduction in effective population size of the Y chromosome compared to the X chromosome. The marginal posterior distributions of k (N_{eY}/N_{eX}) under the appropriate demographic model (in boldface type in Tables 3 and 4) for each population are plotted. The dotted line represents the neutral expectation of k assuming equal numbers of breeding males and females (1/3). Zimbabwe and Uganda populations have a k significantly lower than 1/3. *P*-values were calculated from the empirical cumulative distribution functions of the posterior distribution for k.

5), assuming an N_e of 2.4×10^6 (Thornton and Andolfatto 2006). It is unlikely that a single Y chromosome has swept across Africa in the recent past because there are a number of fixed differences between Zimbabwe and Uganda Y chromosomes (Figure 2B). Instead, it seems more likely that the Y chromosomes in Africa experienced local adaptation and, thus, independent selective sweeps in the recent history of each population. X-linked polymorphism data show some structure among sub-Saharan African populations, suggesting that there may be reduced gene flow between Uganda and Zimbabwe populations (Pool and Aquadro 2006). It is possible that a European Y chromosome swept through the Zimbabwe population as a result of recent admixture.

What could be driving local Y-linked adaptations? The presence of X-linked meiotic drivers inside Africa could cause the replacement of Y chromosomes and shape patterns of local adaptation. Although the frequency of X-linked male meiotic drive is unknown in *D. melanogaster* (see Hanks 1968; Hurst 1996; Reed *et al.* 2005 for possible evidence of such drive in this species), cryptic X-linked meiotic drive is common in other *Drosophila* species (Gershenson 1928; James and Jaenike 1990; Orr and Irving 2005; Montchamp-Moreau *et al.* 2006; Tao *et al.* 2007). Another possibility is that effects of Y chromosome polymorphism on gene expression (Lemos *et al.* 2008, 2010) may influence the adaptive evolution of the African Y chromosome. African and Cosmopolitan populations differ in patterns of gene expression (Meiklejohn *et al.* 2003; Hutter *et al.* 2008). The *D. melanogaster* Y chromosome affects the expression of genes across the genome (Lemos *et al.* 2008, 2010), and a subset of these genes are similar to genes that differ in expression between populations (Hutter *et al.* 2008). For example, local Y-linked adaptation in Africa could be a result of the Y chromosome's effect on thermotolerance (Rohmer *et al.* 2004) or the Y chromosome may have an effect on wing flight muscle architecture, for which there is expression divergence between African and Cosmopolitan populations (Hutter *et al.* 2008).

Patterns of variation in Cosmopolitan Y chromosomes is consistent with a recent, long bottleneck

The frequency spectra and the reduction in variation on the Y *vs.* the X chromosome is more consistent with a history of recent, long population bottlenecks (B2) rather than shorter and more severe bottlenecks (B1). Pool and Nielsen (2007) showed that different demographic scenarios can produce sharply different patterns of variation on the X chromosome compared to the autosomes and that their inference can be extended to effectively haploid regions of the genome such as the Y chromosome and mitochondria. Our results largely agree with their simulations: strong/severe bottlenecks with a longer duration can widen the difference between the X and the autosomes, or the Y and the X chromosomes. Our results also agree with the contrasting scenario where population expansion can cause smaller differences in X and autosomal or Y and X chromosomal variability (Pool and Nielsen 2007).

While natural selection is not necessarily required to explain the reduction of diversity and skew in the frequency spectra of the Cosmopolitan Y chromosome, the data do not exclude the possibility that the *D. melanogaster* Y chromosome may have an adaptive history. First, several differences exist between Pennsylvanian/Tasmanian Y chromosomes and the Netherlands Y chromosomes. Because there is little population structure in Europe, this suggests that the Netherlands Y chromosome may have experienced recent selection. This Y chromosome may have experienced positive selection in the European ancestor prior to sweeping through the Zimbabwe population. Moreover, the reduction in variation on the Pennsylvania Y chromosome compared to the X chromosome is less likely if there is a moderate amount of admixture between Pennsylvania and an African-like population (20% African alleles), although the observed reduction in variation ($\Delta\pi$) was not statistically significant after controlling for the false positive rate (Tables 4; S6). Therefore, the combination of a lack of recombination and recent population bottlenecks may compromise our ability to detect positive selection on Cosmopolitan Y chromosomes. Additionally, if shorter bottlenecks (similar to the B1 models described here) are truly more representative of the demographic history of these populations, then the observed magnitude of reduced variation on the Y chromosome is not expected under neutrality and may be better explained by a history of selection. A more complete understanding of the demographic history of Cosmopolitan

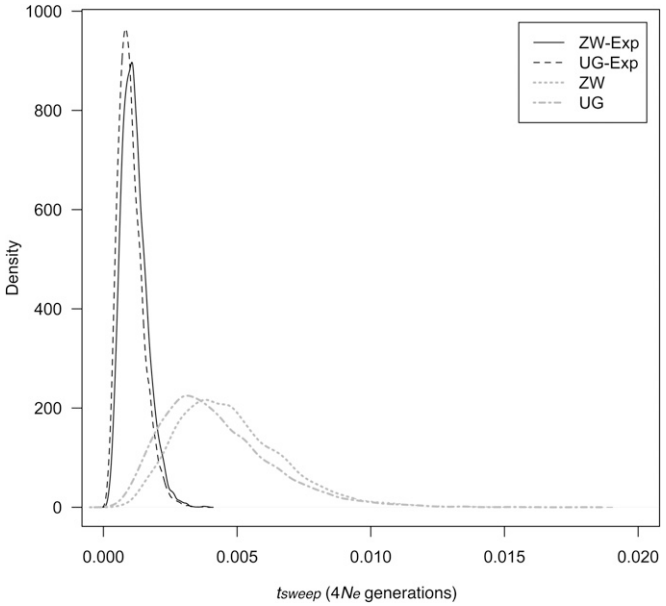

Figure 5 Inference of the time since the most recent selective sweep in African populations. The marginal posterior distribution of the time since the most recent selective sweep (t_{sweep}) shows evidence for a recent selective sweep in both Zimbabwe and Uganda populations under both models of population expansion (black) and constant population size (gray).

populations is necessary to tease apart adaptive and demographic forces shaping the Y chromosome.

Acknowledgments

The authors are grateful for the generous donation of flies by C. F. Aquadro, W. Ballard, J. Pool, Z. Bochdanovits, and A. Hoffman. The authors thank D. Bachtrog for help early on in the project, K. Thornton for help with modeling using libsequence, and anonymous reviewers and David Begun for suggestions that improved the manuscript. This work was supported by a National Institutes of Health grant R01 GM64590 to A.G.C. and A.B. Carvalho.

Note added in proof: After the advanced online publication of this article, Singh *et al.* (2013) published their inference of new demographic model parameters for a Ugandan population of *D. melanogaster*. We implemented their best-fitting model and it does not qualitatively change our conclusions: both Zimbabwe and Uganda Y chromosome data reject the bottleneck followed by exponential expansion model (Singh *et al.* 2013) at *P*-value $< 1\times10^{-4}$ (median simulated $\Delta\pi = 0.6965$).

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.

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 C. F. Aquadro, 1995 Molecular variation at the vermilion locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. Genetics 140: 1019–1032.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57: 289–300.

Bochdanovits, Z., and A. W. de Jong, 2003 Experimental evolution in *Drosophila melanogaster*: interaction of temperature and food quality selection regimes. Evolution 57: 1829–1836.

Caballero, A., 1995 On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics 139: 1007–1011.

Capy, P., M. Veuille, M. Paillette, J. M. Jallon, J. Vouidibio *et al.*, 2000 Sexual isolation of genetically differentiated sympatric populations of *Drosophila melanogaster* in Brazzaville, Congo: The first step towards speciation? Heredity 84(4): 468–475.

Caracristi, G., and C. Schlötterer, 2003 Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. Mol. Biol. Evol. 20: 792–799.

Carvalho, A. B., B. P. Lazzaro, and A. G. Clark, 2000 Y chromosomal fertility factors kl-2 and kl-3 of *Drosophila melanogaster* encode dynein heavy chain polypeptides. Proc. Natl. Acad. Sci. USA 97: 13239–13244.

Carvalho, A. B., B. A. Dobo, M. D. Vibranovski, and A. G. Clark, 2001 Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 98: 13225–13230.

Charlesworth, B., 2001 The effect of life-history and mode of inheritance on neutral genetic variability. Genet. Res. 77: 153–166.

Chippindale, A. K., and W. R. Rice, 2001 Y chromosome polymorphism is a strong determinant of male fitness in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 98: 5677–5682.

Clark, A. G., 1987 Variation in Y chromosome segregation in natural populations of *Drosophila melanogaster*. Genetics 115: 143–151.

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

Gershenson, S., 1928 A new sex-ratio abnormality in *Drosophila obscura*. Genetics 13: 488–507.

Goldstein, L. S., R. W. Hardy, and D. L. Lindsley, 1982 Structural genes on the Y chromosome of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 79: 7405–7409.

Haddrill, P. R., K. R. Thornton, B. Charlesworth, and P. Andolfatto, 2005 Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Genome Res. 15: 790–799.

Hale, L. R., and R. S. Singh, 1991 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history *vs.* selection in the genetic structure of geographic populations. Genetics 129: 103–117.

Hanks, G. D., 1968 *RD* (*Recovery Disrupter*) activity associated with various wild type X chromosomes. Drosoph. Inf. Serv. 43: 98.

Hardy, R. W., K. T. Tokuyasu, and D. L. Lindsley, 1981 Analysis of spermatogenesis in *Drosophila melanogaster* bearing deletions for Y-chromosome fertility genes. Chromosoma 83: 593–617.

Hoskins, R. A., C. D. Smith, J. W. Carlson, A. B. Carvalho, A. Halpern *et al.*, 2002 Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. Genome Biol. 3: research0085–0085.

Hudson, R. R., 2002 Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics 18: 337–338.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111: 147–164.

Hurst, L. D., 1996 Further evidence consistent with *Stellate*'s involvement in meiotic drive. Genetics 142: 641–643.

Hutter, S., H. Li, S. Beisswanger, D. DeLorenzo, and W. Stephan, 2007 Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosome-wide single nucleotide polymorphism data. Genetics 177: 469–480.

Hutter, S., S. S. Saminadin-Peter, W. Stephan, and J. Parsch, 2008 Gene expression variation in African and European populations of *Drosophila melanogaster*. Genome Biol. 9: R12.

James, A. C., and J. Jaenike, 1990 "Sex ratio" meiotic drive in *Drosophila testacea*. Genetics 126: 651–656.

Jiang, P. P., D. L. Hartl, and B. Lemos, 2010 Y not a dead end: epistatic interactions between Y-linked regulatory polymorphisms and genetic background affect global gene expression in *Drosophila melanogaster*. Genetics 186: 109–118.

Kaiser, V. B., and B. Charlesworth, 2009 The effects of deleterious mutations on evolution in non-recombining genomes. Trends Genet. 25: 9–12.

Kauer, M., D. Dieringer,, and C. Schlötterer, 2003 Nonneutral admixture of immigrant genotypes in African *Drosophila melanogaster* populations from Zimbabwe. Mol. Biol. Evol. 20: 1329–1337.

Kennison, J. A., 1981 The genetic and cytological organization of the Y chromosome of *Drosophila melanogaster*. Genetics 98: 529–548.

Kolaczkowski, B., A. D. Kern, A. K. Holloway, and D. J. Begun, 2011 Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. Genetics 187: 245–260.

Kopp, A., A. K. Frank, and O. Barmina, 2006 Interspecific divergence, intrachromosomal recombination, and phylogenetic utility of Y-chromosomal genes in Drosophila. Mol. Phylogenet. Evol. 38: 731–741.

Krsticevic, F. J., H. L. Santos, S. Januario, C. G. Schrago, and A. B. Carvalho, 2010 Functional copies of the Mst77F gene on the Y chromosome of *Drosophila melanogaster*. Genetics 184: 295–307.

Laurent, S. J., A. Werzner, L. Excoffier, and W. Stephan, 2011 Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia. Mol. Biol. Evol. 24: 2041–2051.

Lemos, B., L. O. Araripe, and D. L. Hartl, 2008 Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. Science 319: 91–93.

Lemos, B., A. T. Branco, and D. L. Hartl, 2010 Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. Proc. Natl. Acad. Sci. USA 107: 15826–15831.

Li, H., and W. Stephan, 2006 Inferring the demographic history and rate of adaptive substitution in Drosophila. PLoS Genet. 2: e166.

Lyckegaard, E. M., and A. G. Clark, 1989 Ribosomal DNA and *Stellate* gene copy number variation on the Y chromosome of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 86: 1944–1948.

Meiklejohn, C. D., J. Parsch, J. M. Ranz, and D. L. Hartl, 2003 Rapid evolution of male-biased gene expression in Drosophila. Proc. Natl. Acad. Sci. USA 100: 9894–9899.

Montchamp-Moreau, C., D. Ogereau, N. Chaminade, A. Colard, and S. Aulard, 2006 Organization of the sex-ratio meiotic drive region in *Drosophila simulans*. Genetics 174: 1365–1371.

Myers, S. R., and R. C. Griffiths, 2003 Bounds on the minimum number of recombination events in a sample history. Genetics 163: 375–394.

Nunney, L., 1993 The influence of mating system and overlapping generations on effective population size. Evolution 47: 1329–1341.

Orr, H. A., and S. Irving, 2005 Segregation distortion in hybrids between the Bogota and USA subspecies of *Drosophila pseudoobscura*. Genetics 169: 671–682.

Paredes, S., A. T. Branco, D. L. Hartl, K. A. Maggert, and B. Lemos, 2011 Ribosomal DNA deletions modulate genome-wide gene expression: "rDNA-sensitive" genes and natural variation. PLoS Genet. 7: e1001376.

Pool, J. E., and C. F. Aquadro, 2006 History and structure of sub-Saharan populations of *Drosophila melanogaster*. Genetics 174: 915–929.

Pool, J. E., and R. Nielsen, 2007 Population size changes reshape genomic patterns of diversity. Evolution 61: 3001–3006.

Przeworski, M., 2003 Estimating the time since the fixation of a beneficial allele. Genetics 164: 1667–1676.

Reed, F. A., R. G. Reeves, and C. F. Aquadro, 2005 Evidence of susceptibility and resistance to cryptic X-linked meiotic drive in natural populations of *Drosophila melanogaster*. Evolution 59: 1280–1291.

Rohmer, C., J. R. David, B. Moreteau, and D. Joly, 2004 Heat induced male sterility in *Drosophila melanogaster*: adaptive genetic variations among geographic populations and role of the Y chromosome. J. Exp. Biol. 207: 2735–2743.

Schlötterer, C, H. Neumeier, C. Sousa, and V. Nolte, 2006 Highly structured Asian *Drosophila melanogaster* populations: A new tool for hitchhiking mapping? Genetics 172: 287–292.

Singh, N. D., J. D. Jensen, A. G. Clark, and C. F. Aquadro, 2013 Inferences of demography and selection in an African population of *D. melanogaster*. Genetics 193: 215–228.

Tao, Y., L. Araripe, S. B. Kingan, Y. Ke, H. Xiao *et al.*, 2007 A sex-ratio meiotic drive system in *Drosophila simulans*. II: an X-linked distorter. PLoS Biol. 5: e293.

Thornton, K., 2003 Libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics 19: 2325–2327.

Thornton, K., and P. Andolfatto, 2006 Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics 172: 1607–1619.

Vibranovski, M. D., L. B. Koerich, and A. B. Carvalho, 2008 Two new Y-linked genes in *Drosophila melanogaster*. Genetics 179: 2325–2327.

Zhang, P., 2000 A trans-activator on the Drosophila Y chromosome regulates gene expression in the male germ line. Genetica 109: 141–150.

Zurovcova, M., and W. F. Eanes, 1999 Lack of nucleotide polymorphism in the Y-linked sperm flagellar dynein gene *Dhc-Yh3* of *Drosophila melanogaster* and *D. simulans*. Genetics 153: 1709–1715.

Communicating editor: D. Begun

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.146167/-/DC1

Surprising Differences in the Variability of Y Chromosomes in African and Cosmopolitan Populations of *Drosophila melanogaster*

Amanda M. Larracuente and Andrew G. Clark

DOI: 10.1534/genetics.112.146167

Figure S1 Population differentiation. An unrooted population distance tree generated by a neighbor-joining analysis of *Fst* is diagrammed.

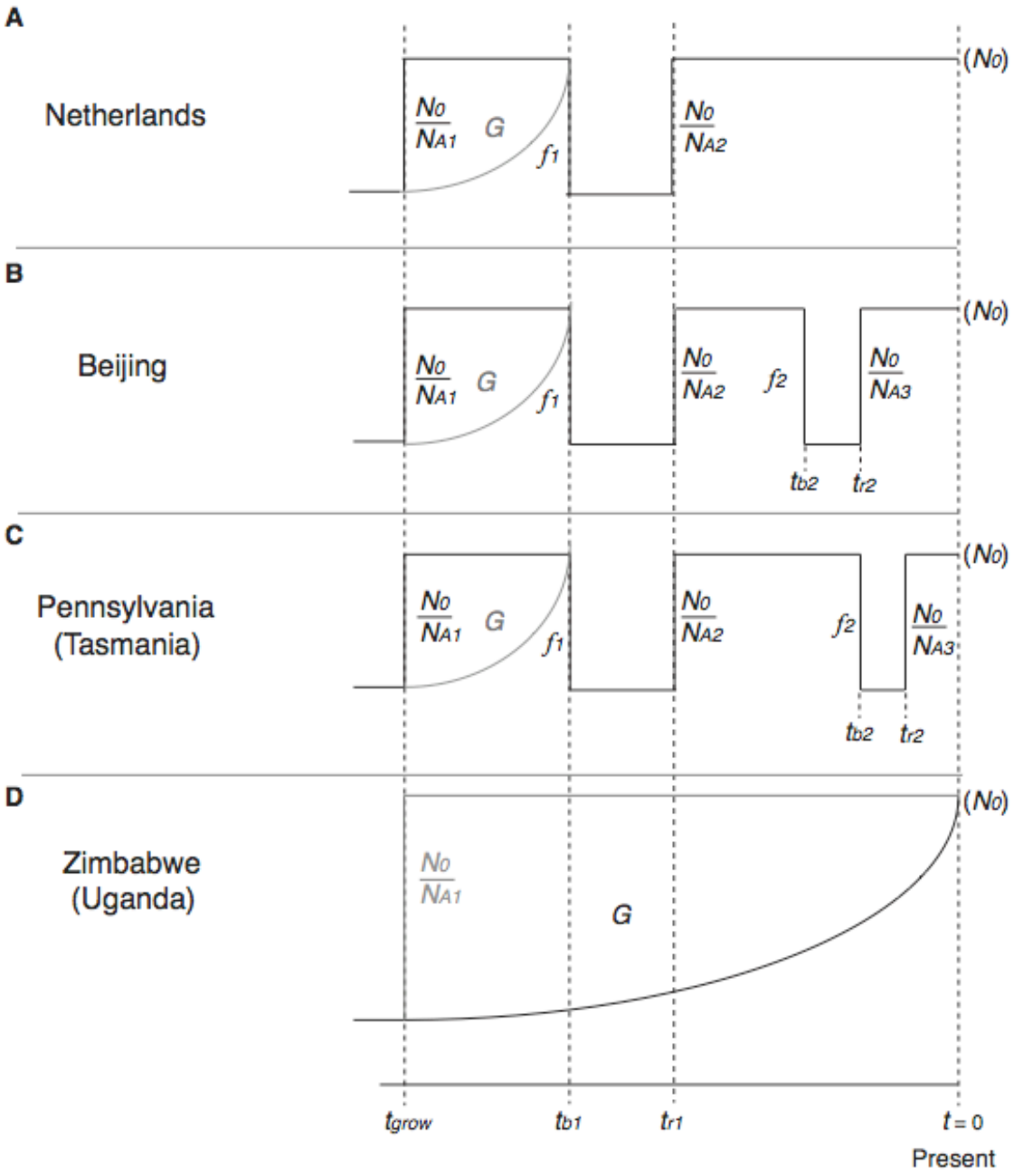

Figure S2 Alternate demographic models explored. Diagrammed are schematics describing the models of simple population bottlenecks with ancestral growth and population expansion and their parameters. For choice of best demographic model, and parameter values, see Tables 3 and 4, and S1. **A**. The Netherlands population bottleneck describes an ancestral African population that experienced stepwise growth (to size N_0/N_{A1}) or exponential growth (in grey; at rate G) at time t_{grow}. At time t_{b1}, the population reduced in size to f_1*N_0 and recovered to size N_0/N_{A1} at time t_{r1}. **B**. The Beijing population arose from an expanding ancestral population (as in A), and bottlenecked twice, the first describes the European bottleneck from an ancestral African population (same as Netherlands bottleneck). **C**. The Pennsylvania population experienced a similar ancestral population growth and bottlenecks except that the second bottleneck occurred more recently. **D**. The demographic model for the Zimbabwe population describes an ancestral African population that began expanding at time t_{grow} at rate G until the present time (or stepwise to size N_0/N_{A1}; in gray). All time is in scaled units of $4N_e$ generations. The best demographic models for Tasmania is the Pennsylvania double bottleneck (C) and the Uganda population is the Zimbabwe expansion model (D).

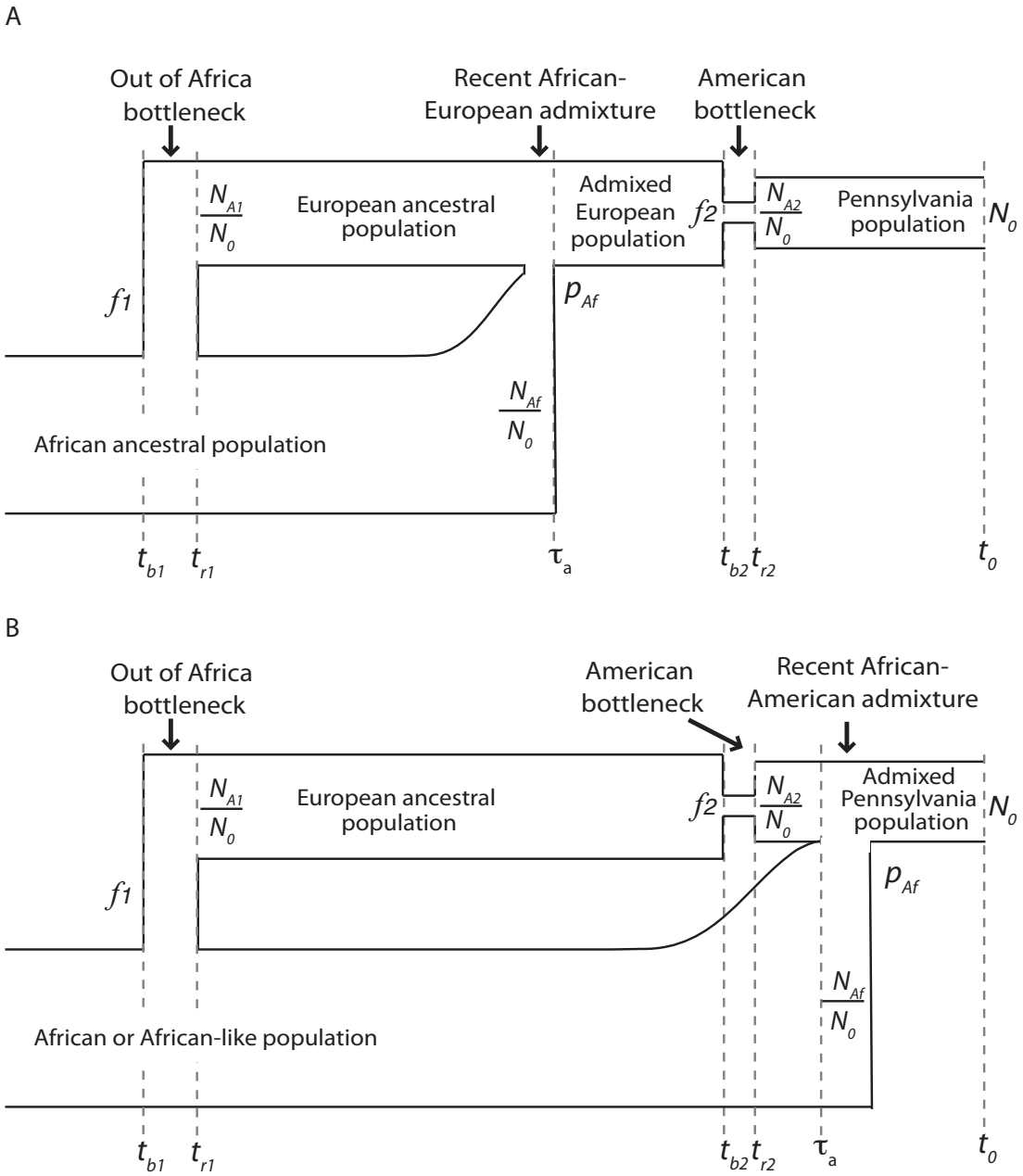

Figure S3 Model describing the founding of the Pennsylvania population from an admixed European/African ancestor. The American bottleneck describes an ancestral African population that founded a European population as a bottleneck at time t_{b1}, the population reduced in size to f_1*N_0 and recovered to size N_{A1}/N_0 at time t_{r1} (the Out-of-Africa bottleneck). A. The AF-EU admixture model: At time t, the admixture between the European and African population (where size of African population was N_{Af}/N_0) occurred where p_{Af} is the fraction of African alleles entering the European population. The Pennsylvanian population was founded from this ancestral admixed population through a bottleneck at time t_{b2} (the American bottleneck; the population reduced in size to f_2*N_0 and recovered to size N_{A2}/N_0 at time t_{r2}). B. The AF-PA model: The Pennsylvanian population was founded from a bottleneck at time t_{b2} (the American bottleneck; the population reduced in size to f_2*N_0 and recovered to size N_{A2}/N_0 at time t_{r2}). Following the bottleneck, the admixture occurred between the Pennsylvania population and an African, or African-like population. At time t, the admixture between the Pennsylvania and African-like population (size N_{Af}/N_0) occurred where p_{Af} is the fraction of African alleles entering the Pennsylvania population.

Table S1 Summary of X-linked datasets from literature

POPULATION	*n*	θ	π	D_{Taj}	Reference
Zimbabwe	24	0.43895	1.325	-0.38	Haddrill *et al.* 2005
Uganda	10	1.210	1.020	-0.8	Pool and Aquadro 2006
Beijing	10	0.360	0.340	-0.58	Pool and Aquadro 2006
Tasmania	19	-	0.212	-	Kolaczkowski *et al.* 2011
Netherlands	24	0.571	0.414	-0.66	Haddrill *et al.* 2005
Pennsylvania	24	0.664	0.557	-0.65	Haddrill *et al.* 2005

Table S2 Estimates of π/π_0 on the Y chromosome in each population

POPULATION	π_{obs}	π_{sim}	π/π_0	95% CI
Zimbabwe	0.00323	0.43895	0.0074	0.0200-0.0032
Uganda	0.00266	0.39647	0.0067	0.0184-0.0029
Beijing	0.01696	0.11876	0.1428	0.4749-0.0602
Tasmania	0.01370	0.07022	0.1951	0.7364-0.0800
Netherlands	0.01505	0.13643	0.1103	0.3493-0.0466
Pennsylvania	0.00849	0.18446	0.0460	0.1364-0.0198

Estimates of π_{sim} are based on 10^5 neutral coalescent simulations.

Table S3 Model descriptions and parameter values

	TYPE	N_0	t_{b1}	d_1	t_{r1}	f_1	N_0/N_{A1}	t_{b2}	d_2	t_{r2}	f_2	N_0/N_{A2}
NE	B1	1.075×10^6	0.049	0.001	0.048	0.002	0.125	-	-	-	-	-
	B2	2.4×10^6	0.022	0.018	0.004	0.029	1.0	-	-	-	-	-
BEI	B1	4.14×10^5	0.127	0.003	0.124	0.002	0.125	0.020	0.001	0.019	0.020	0.254
	B2	2.4×10^6	0.022	0.015	0.004	0.029	1.0	0.0021	0.0001	0.002	0.020	*1*
PA	B1	1.075×10^6	0.049	0.001	0.048	0.002	0.125	*0.0009*	*0.0003*	*0.0006*	*0.001*	*1*
	B2	2.4×10^6	0.022	0.015	0.004	0.029	1.0	*0.0004*	*0.0001*	*0.0003*	*0.001*	*1*

Model descriptions and parameter values. For each Cosmopolitan population, parameter values for the "B1" (more severe bottleneck of shorter duration) and "B2" (longer bottleneck with more recent recovery) bottleneck types are listed. Each of these models were run considering the case of stepwise or exponential growth in an ancestral African population beginning at time t_{grow} (0.0833 for B1 and 0.186, 0.124 and 0.479 for NE, BEI and PA B2 models respectively) and a stable ancestral population that was not expanding (See Figure S2). All times are in terms of $4Ne_0$ generations and were scaled appropriately for the Y-chromosomes for simulations ($t_Y=3t_X$). Italicized numbers are arbitrarily assigned values consistent with population history but were not inferred from empirical data (in the literature). The time descriptors "d_i" were not parameters of the model, and simply reflect t_{bi}-t_{ri} where *i* corresponds to 1 or 2 for the first and second bottlenecks, respectively.

Table S4 Performance of different inferred demographic models for the Netherlands, Beijing and Zimbabwe populations.

	DEMOGRAPHIC[a] MODEL	MEDIAN D_{TAJ} (P) [b,c]		MEDIAN Π (P) [b,c]		MEDIAN θ_W (P) [b,c]		$\Delta\pi$ [b,c]
		X	Y	X	Y	X	Y	
Netherlands	B1	-0.4642 (0.8244)	-1.4102 (0.7710)	0.4667 (0.9019)	*0.1066 (<10^{-4})*	0.5318 (0.9256)	*0.1636 (<10^{-4})*	*0.7907 (<10^{-4})*
	B2	**0.5209 (0.4686)**	**-1.5108 (0.7414)**	**0.0033 (0.8375)**	**0.0207 (0.5164)**	**0.3191 (0.3642)**	**0.0381 (0.3196)**	**0.9470 (0.4976)**
	B1-AF-Exp	-0.2060 (0.4216)	-0.1408 (0.1602)	0.4846 (0.7720)	*0.1806 (<10^{-4})*	0.5318 (0.8856)	*0.1903 (<10^{-4})*	*0.6581 (<10^{-4})*
	B2-AF-Exp	-0.1498 (0.5464)	-1.7112 (0.4558)	0.4201 (0.9836)	*0.0835 (<10^{-4})*	0.4255 (0.6448)	*0.1446 (<10^{-4})*	*0.8171 (<10^{-4})*
	B1-AF-SW	0.349 (0.2614)	-0.1418 (0.1556)	1.240 (0.3864)	*0.1792 (<10^{-4})*	1.1170 (0.4842)	*0.1903 (<10^{-4})*	*0.8641 (0.0126)*
	B2-AF-SW	0.7600 (0.2840)	-1.7177 (0.4412)	1.702 (0.2020)	*0.0834 (<10^{-4})*	1.489 (0.234)	*0.1446 (<10^{-4})*	0.9512 (0.4082)
Beijing	B1	-0.2792 (0.5766)	-0.5183 (0.0572)	0.1467 (0.0912)	*0.0591 (0.0154)*	*0.1414 (0.0626)*	0.0691 (0.0402)	*0.6145 (0.0014)*
	B2	**0.4050 (0.8902)**	**-1.2509 (0.3950)**	**0.3578 (0.9654)**	**0.0201 (0.7534)**	**0.3535 (0.9160)**	**0.0311 (0.8574)**	**0.9469 (0.8514)**
	B1-AF-Exp	-0.3339 (0.3142)	-0.5535 (0.0396)	0.5911 (0.0950)	*0.2374 (<10^{-4})*	0.6363 (0.0764)	*0.2796 (<10^{-4})*	*0.6019 (<10^{-4})*
	B2-AF-Exp	0.0463 (0.8792)	-1.4499 (0.5324)	0.4378 (0.6712)	*0.0805 (<10^{-4})*	0.4241 (0.7384)	*0.1281 (<10^{-4})*	*0.8209 (0.0026)*
	B1-AF-SW	-0.2988 (0.3822)	-0.5550 (0.046)	0.620 (0.0878)	0.2382 *(<10^{-4})*	0.6716 (0.0742)	*0.2796 (<10^{-4})*	*0.6359 (<10^{-4})*
	B2-AF-SW	0.5572 (0.6756)	-1.4466 (0.5876)	1.693 (0.0826)	*0.0813 (<10^{-4})*	1.556 (0.0770)	*0.1281 (<10^{-4})*	0.9515 (0.9466)
Zimbabwe	**AF-Exp**	**-0.7658 (0.3832)**	**-0.8480 (0.0522)**	**1.320 (0.9841)**	***0.6126 (<10^{-4})***	**1.620 (0.6699)**	***0.7830 (<10^{-4})***	***0.5378 (<10^{-4})***
	AF-SW	-0.930 (0.0781)	-1.2914 (0.3258)	*6.4900 (<10^{-4})*	*2.880 (<10^{-4})*	*8.314 (<10^{-4})*	*4.393 (<10^{-4})*	*0.5553 (<10^{-4})*

[a] Results from simulated genealogies for 10 independent X-linked loci, and a single Y-linked locus (assuming θ_Y=1/3θ_X) under hypothetical bottleneck or exponential growth models specified in the Methods and Table S1.

[b] Median D_{Taj} , π (per 100 sites), θ (per 100 sites) for X and Y-linked loci and $\Delta\pi$.

[c] *P-values* for D_{Taj} , π (per 100 sites), θ (per 100 sites), and $\Delta\pi$ are two-sided and were calculated using the ecdf function in R. *P-values* that are below a FDR of 5% are indicated with italics.

The model chosen for use in subsequent simulations is in bold print.

Table S5 Performance of different demographic models for the Pennsylvania, Tasmania and Uganda populations.

DEMOGRAPHIC[a] MODEL		MEDIAN D_{TAJ} (P)[b,c]		MEDIAN Π (P)[b,c]		MEDIAN θ_W (P)[b,c]		$\Delta\pi$[b,c]
		X	Y	X	Y	X	Y	
PA	B1	0.6559 (0.3538)	-1.159 (0.0344)	0.0611 (0.0078)	0.0023 (0.5202)	*0.0524* *($4x10^{-4}$)*	0.0075 (0.2422)	0.9663 (0.5802)
	B2	**0.8364 (0.3612)**	**0.1386 (0.0086)**	**0.2659 (0.4900)**	**0.0023 (0.6306)**	**0.2622 (0.1320)**	**0.0149 (0.2868)**	**0.9609 (0.3574)**
	B1-AF-SW	0.7634 (0.3060)	-1.7325 (0.4830)	0.2875 (0.2990)	0.0104 (0.8494)	0.2622 (0.0500)	0.0261 (0.8744)	0.9663 (0.333)
	B2-AF-SW	1.031 (0.2376)	0.1390 (0.0254)	1.443 (0.3670)	0.0545 (0.0186)	1.210 (0.4690)	0.0523 (0.1050)	0.9636 (0.1620)
TAS	PA-B2	**-**	**0.1386 (0.0560)**	**0.2659 (0.8902)**	**0.0023 (0.9700)**	**-**	**0.0149 (0.1916)**	**0.9609 (0.5384)**
	BEI-B2	-	-1.2509 (0.5240)	0.3578 (0.6682)	0.0201 (0.4828)	-	0.0311 (0.6166)	0.9469 (0.7654)
	NE-B2	-	-1.5108 (0.8376)	0.0033 (0.7586)	0.0207 (0.4212)	-	0.0381 (0.5190)	0.9470 (0.7300)
UG	ZW-Exp	**-0.7658 (0.9350)**	**-0.8480 (0.1512)**	**1.320 (0.3932)**	***0.6126*** ***($<10^{-4}$)***	***1.620*** ***($6x10^{-4}$)***	***0.783*** ***($<10^{-4}$)***	***0.5378*** ***($<10^{-4}$)***

[a]Results from simulated genealogies for 10 independent X-linked loci, and a single Y-linked locus (assuming θ_Y=1/3θ_X) under hypothetical bottleneck or exponential growth models specified in the Methods and Table S1.

[b] Median D_{Taj} , π (per 100 sites), θ (per 100 sites) for X and Y-linked loci and $\Delta\pi$.

[c]*P-values* for D_{Taj} , π (per 100 sites), θ (per 100 sites), and $\Delta\pi$ are two-sided and were calculated using the ecdf function in R. *P-values* that are below a FDR of 5% are indicated with italics.

The model chosen for use in subsequent simulations is in bold print.

Table S6 ***P*-values for Tajima's *D* and Δp under admixture models for the Pennsylvania and Tasmania populations.**

model	τ_a[a]	P_{Af}[b]	D_{Taj}[c]			$\Delta\pi$[c]	
			X (PA)	Y		PA	TAS
				PA	TAS		
AF-EU Admixture	0.0022	0	0.7912 (0.5012)	0.1165 (0.0116)	0.1165 (0.0634)	0.9610 (0.3326)	0.9610 (0.5192)
		0.05	0.8321 (0.5048)	0.1143 (0.0144)	0.1143 (0.0708)	0.9885 (0.8136)	0.9885 (0.1402)
		0.10	0.8985 (0.5024)	0.1038 (0.0138)	0.1038 (0.0604)	0.9896 (0.6850)	0.9896 (0.0972)
		0.15	0.9365 (0.5000)	0.1334 (0.0108)	0.1334 (0.0658)	0.9903 (0.6232)	0.9903 (0.076)
		0.20	0.9673 (0.4988)	0.1334 (0.0114)	0.1334 (0.0610)	0.9906 (0.5800)	0.9906 (0.0792)
AF-PA Admixture	0.0007	0	0.7936 (0.5014)	0.0992 (0.009)	NA	0.9614 (0.3418)	NA
		0.05	**-1.5996 (0.6520)**	**-1.5455 (0.8186)**	**NA**	**0.8807 (0.7072)**	**NA**
		0.10	-1.3428 (0.6570)	-1.568 (0.7330)	NA	0.8041 (0.2418)	NA
		0.15	-0.9511 (0.8022)	-1.448 (0.5132)	NA	0.7842 (0.0822)	NA
		0.20	-0.6201 (0.9638)	-1.0113 (0.3562)	NA	0.7677 (0.0286)	NA

[a]Results from simulated genealogies for 10 independent X-linked loci, and a single Y-linked locus (assuming θ_Y=1/3θ_X) under hypothetical bottleneck models including varying amounts of admixture between African and European populations at time t_a (in 4*Ne* generations—time was scaled appropriately for the X and Y chromosomes in simulations). In the AF-EU admixture models, the European ancestor was admixed with African alleles and admixture was immediately followed by the American bottleneck. In the AF-PA admixture models, the Pennsylvania population admixed with African alleles after the American bottleneck.

[b]p_{Af} reflects the degree of admixture (the proportion of African alleles in the admixed population)

[c] Two-sided *P*-values for D_{Taj} and $\Delta\pi$ were calculated using the ecdf function in R.

No *P*-values fall below an FDR <5%.

The AF-EU models were used to obtain *P*-values for the Tasmanian population. The AF-EU and AF-PA P_{Af}=0 models in bold print indicate the equivalent to model PA-B1 in Table 4. The model chosen for use in subsequent simulations is in bold print.

Table S7 Estimates of the reduction in effective population size on the Y chromosome

POPULATION	*k*-NEUT[a] *k*-DEMO	95% CI	*P*-VALUE
Zimbabwe	0.0070	0.0060-0.0080	$<1x10^{-4}$
	0.0041	0.0005-0.0279	$<1x10^{-3}$
Uganda	0.0062	0.0053-0.0072	$<1x10^{-4}$
	0.0037	0.0004-0.0256	$<1x10^{-3}$
Beijing	0.0670	0.0430-.0772	$<1x10^{-4}$
	0.2925	**0.0912-0.7530**	**0.9500**
Tasmania	0.03998	0.0344-0.0469	$<1x10^{-4}$
	0.4397	**0.1044-0.9443**	**0.6140**
Netherlands	0.0425	0.0364-0.0497	$<1x10^{-4}$
	0.2240	**0.0633-0.6723**	**0.7300**
Pennsylvania	0.0356	0.0305-0.0415	$<1x10^{-4}$
	0.1858	**0.0514 -0.7782**	**0.7240**

[a] The M.A.P. estimate of scaling factor *k,* which represents Ne_Y/Ne_X, under a standard neutral model (top) and demographic model (bottom). Significance based on an empirical cumulative distribution function of the posterior distribution of *k* and the expectation that *k* should be 1/3 if there are equal numbers of breeding males and females (two-sided *P*-values).