Supplementary Methods 1: Joint estimates of θ and ρ using an approximate Bayesian method.

We employ a new multi-locus estimator of ρ (Thornton, unpublished) that is similar in spirit to the summary likelihood estimator described by Wall (2000). Here, posterior distributions of ρ and θ were jointly estimated by an approximate Bayesian method based on summary statistics of the data and rejection sampling (see Pritchard *et al*. 1999 and Algorithm D of Marjoram *et al*. 2003). For each population, each locus was summarized by its sample size (n), alignment length (L), nucleotide diversity (π), number of segregating sites (S), the number of haplotypes (K), and the minimum number of recombination events in the sample (R_m). R_m was calculated as the maximum of either Hudson and Kaplan's (1985) estimator or *K-S-1* (Myers and Griffiths 2003). The summary statistics used for the entire data set were the mean π, mean S, and mean R_m across loci. We assume that mutation and recombination rates are constant per site (both within and amongst loci) and that loci are independent. For each population, the joint posterior distribution of θ, ρ, and $f(\theta,\rho)$ was obtained according the following rejection algorithm:

1. Summarize the data by mean π, mean S, and mean R_m
2. Draw random θ and ρ from a uniform prior distribution.
3. Simulate 10 loci with parameters n_i, L_i, θ, and ρ, where the subscript i refers to the parameter for the i-th locus, and parameters without subscripts are constant across loci.
4. Summarize the 10 simulated loci as in step 1.
5. Accept θ and ρ if all three summaries of the simulated data are within ε of the observed data (where ε is an arbitrary constant).

6. Return to step 2 until k draws from $f(\theta,\rho)$ are obtained.

As the acceptance rate for these rejection algorithms can be quite low (Marjoram *et al.* 2003; Pritchard *et al.* 1999), we first generated $k=10^3$ draws from the joint posterior using wide priors on θ and ρ for each population with a relatively liberal ε. We then took the 1% and the 99% quantiles of the marginal distributions of θ and ρ from the resulting $f(\theta,\rho)$ as the ranges for new uniform priors on the parameters. Using the new prior, we obtained $k=10^4$ samples from $f(\theta,\rho)$ using a smaller ε than in the initial phase. For all populations, the acceptance rates were approximately 10^{-3}. The analysis took roughly two weeks on 14 G5 processors.

Supplementary Methods 2: Evaluating alternatives to the standard neutral model.

We use the combined 115 locus dataset to evaluate three types of alternative models: an exponential growth model; a bottleneck model and a recurrent hitchhiking model. For consistency with our 10 surveyed loci, alignments for the 105 Glinka et al. (2003) loci were culled to remove all multiply hit sites and sites overlapping deletions. We summarise the combined data set using various statistics including the average nucleotide diversity (π), average pairwise divergence to *D. simulans* (D_{xy}), the HKA χ^2 statistic, and the means and variances of Tajima's *D* and Fay and Wu's *H*. In each case, simulation parameters were scaled to mimic the observed data (*i.e.* levels of variability, π, and average pairwise divergence, D_{xy}) as closely as possible. In simulations with recombination, we assume $\rho/\theta=7$ which is close to the lower 95% confidence bound for the 115 locus Zimbabwe dataset. *P*-values are based on 10,000 replicates of the coalescent.

Population growth model: To see how compatible the Zimbabwe data are with population growth models (Table 4), we used the program *ms* (Hudson 2002) to illustrate some features of these models. We assume an exponential growth model $N_t = N_o * e^{rt}$, where r is the rate of growth and t is the time (in units of $4N_e$ generations) that growth began. We modeled 5-fold and 10-fold growth at two different growth rates. The time that growth began, t, was calculated by setting $t = \ln(N_t / N_o)/r$. For $r = 10$ and 100 these t were: 0.05756 and 0.005756, respectively for 10-fold growth and 0.04024 and 0.004024 for 5-fold growth. Using the HKA framework (Hudson *et al*. 1987), we estimate the expected θ parameter for each locus and the divergence time, T, (in units of N_e) to *D. simulans* using the 115 loci surveyed in Zimbabwe and a single *D. simulans*. Using this framework, the divergence time to *D. simulans* was estimated to be $8N_e$ generations ago, where N_e is the effective population size of the Zimbabwe population. θ estimates for each locus are available on request to P.A. For each model, the parameters θ and T were scaled to produce a similar average level of diversity as observed in the Zimbabwe population (*i.e*. average $\pi \sim 0.01$ per site).

Population bottleneck model: As for growth models, we used the program *ms* to illustrate some features of bottleneck models. We model a bottleneck as occurring at time T_b in the past of severity f. The population recovers instantly to its former size T_d generations later. In Table 4, we model a bottleneck occurring in the past of the Zimbabwe population. Simulated data for each locus was generated using the sample size, the length in base pairs, and θ. In this case, we assumed that θ of the ancestral population is two times higher than the average π observed in the Zimbabwe population. In models BN1-3 of Table 5, we model a single bottleneck associated with the colonization of non-African

populations. The duration of the bottleneck, T_d, was set to reduce diversity in the derived population to that observed in the Netherlands relative to the Zimbabwe population ($\pi\_Neth/\ \pi\_Zim \sim 0.4$). For each replicate of 115 simulated loci, we then performed the same multi-locus statistical tests of neutrality as performed on the real data.

Recurrent hitchhiking model: We used a program (Przeworski 2002) that models the effect of recurrent selective sweeps on a randomly chosen neutral locus (kindly provided by M. Przeworski). The parameters of this model include the strength of selection, s, the effective population size of the species (N_e), and λ, expected number of sweeps in $4N_e$ generations. For models HH1 and HH2 (Zimbabwe population, Table 4), we assume that recurrent sweeps have reduced levels of diversity (π) to 0.5θ. We assume $N_e = 5$ million and set s, λ and θ to result in an average level of diversity (π) that is close to that average observed diversity in Zimbabwe (*i.e.* average $\pi \sim 0.01$ per site). For simulations of the Netherlands population (Table 5), we assumed θ was the average θ estimated in the Zimbabwe population (using the HKA framework) and that $N_e = 2.5$ million. We then chose s and λ to reduce diversity to the mean observed in the Netherlands relative to Zimbabwe ($\pi\_Neth/\ \pi\_Zim \sim 0.4$).

Supplementary Methods 3: Is it appropriate to combine our 10-locus data set and the 105 locus dataset of Glinka et al. 2002?

Our population structure analysis failed to find evidence for structure between Zimbabwe and Kenya; this makes it less likely that we would expect to find structure between two different Zimbabwe population samples. We used exactly the same Netherlands population sample used by Glinka et al. (i.e. collected by Andrew Davis in

2000) so these samples are unlikely to be different. Nonetheless, we compared several summary statistics for the 10 locus dataset (sub-samples of 12 alleles) to bootstrap replicate sub-samples of 10 loci from the Glinka et al. dataset. The results are as follows:

Summary used	Two-tailed probabilities of observing the Haddrill et al. mean values in subsets of the Glinka et al. data.	
	Zimbabwe	Netherlands
θ	0.12	0.31
HKA $\chi 2$	0.23	0.88
Tajima's D	0.21	0.05
FayWu H	0.09	0.04

Note that these *P*-values have not been corrected for performing multiple tests. The expected number of $p<0.05$ events in 8 independent stats is 0.4; the Poisson prob. of observing two of these is 0.062, thus there is no strong support for differences. Regardless, the largest departures are in fact observed in the Netherlands sample, for which we have sequenced alleles from the same population sample as Glinka et al. Thus, our 10-locus dataset is not demonstrably different than the Glinka et al. sample and it is appropriate to combine them. Differences in the mean values of summary statistics in Table 3 vs Tables 4 and 5 probably reflect sample size differences (which we account for explicitly in simulations) and the large variance among loci due to departures from the neutral panmictic population model (part of the point of this paper).

Supplementary Methods 4: Alignment of *simulans* sequences.

The *simulans* sequence obtained for four of the loci contained short regions where it was not possible to align to the *melanogaster* sequences. These regions were effectively

removed from the alignment by changing all the residues to N. The loci and regions are detailed below:

Locus	Size of region (bp)	Position in alignment
spaghetti squash	38	268-305
cut	14	458-471
vermilion	57	225-281
licorne	12	71-82

In addition, at the *licorne* locus, the sequenced *simulans* allele contained a 57 bp deletion after the site of the section of 12 base pairs that could not be aligned. We therefore inserted a 65bp section of an allele from the *simulans* genome sequencing project (Contig7752.1 sim6), which covered bases 75 to 163 of the alignment (including a 23 bp alignment gap), to create a 'composite' *D. simulans* allele.