**File S1.** Additional Materials and Methods

***Quality filtering of the Raleigh and Zambian data***

The Raleigh strains used for the analysis were generated by MACKAY *et al*. 2012 and belong to the *Drosophila* Genetic Reference Panel. The DGRP data set consists of fully sequenced genomes of 205 inbred *D. melanogaster* lines collected from Raleigh, North Carolina. The Zambian strains belong to the *Drosophila* Genome Nexus (LACK *et al*. 2015) and consist of 197 fully sequenced strains.

We masked regions in the data inferred to have high levels of identify by descent (IBD), admixture, and inversions because these are all factors that can contribute to high LD. Scripts for masking were provided by John Pool (www.johnpool.net) along with regions inferred to have high IBD and admixture (LACK *et al*. 2015). Only African strains were masked for admixture between African and European strains. North American strains have a high level of genome-wide admixture (DUCHEN *et al*. 2013), and so, admixture in these strains was controlled for by performing neutral simulations under the admixture model by DUCHEN *et al*. 2013. All heterozygous sites and inferred tracts of residual heterozygosity (LACK *et al*. 2015) were masked as well. Inversions were masked by excluding the intervals of breakpoints identified in CORBETT-DETIG AND HARTL 2012; HUANG *et al*. 2014; LACK *et al*. 2015.

The extensive masking resulted in large tracts of 'Ns' in the data with an average length of 4 MB. In some cases, entire chromosomes were masked. However, strains with high levels of masking on one chromosome did not necessarily have high levels of masking on other chromosomes. To account for this heterogeneity in the strains with masking on each chromosome, we down sampled to 100 strains with the least number of Ns per chromosome. Thus, each chromosome consisted of different sets of 100 strains. The strain IDs used for each chromosome are given in Table S3.

To have the best resolution and power to detect individual selective sweeps with H12, we applied less stringent filters to the Raleigh and Zambian data sets and did not mask regions of IBD, admixture, and inversions. Instead, we excluded eight Zambian and 27 Raleigh strains with genome-wide IBD levels with at least one other strain > 20% as calculated by LACK *et al*. 2015. The excluded strain IDs are as follows: ZI397, ZI530, ZI269, ZI240, ZI218, ZI207, ZI523, ZI86,

RAL-385, RAL-358, RAL-712, RAL-399, RAL-879, RAL-355, RAL-810, RAL-350, RAL-832, RAL-882, RAL-306, RAL-799, RAL-801, RAL-859, RAL-907, RAL-790, RAL-748, RAL-336, RAL-850, RAL-365, RAL-786, RAL-730, RAL-861, RAL-59, RAL-646, RAL-812, and RAL-787. In this version of the data set, we masked all heterozygous sites and inferred tracts of residual heterozygosity as in the previous data set.

H12 was applied to the entire 178 Raleigh strains and 188 Zambian strains. To reduce any remaining heterogeneity in sample size due to missing data, we down-sampled to the 145 strains with the least amount of missing data in each analysis window for inclusion in the H12 calculation.

***Calculation of the critical value ($H12_o$)***

To calculate the critical value of H12, or, 1-per-genome false discovery rate value, we used the method described in GARUD *et al*. 2015 whereby we simulated 150,000 analysis windows under a neutral constant $N_e$ demographic model, which represents approximately 10x the number of analysis windows in the data. We assigned the $10^{th}$ highest H12 value in the simulated distribution to be $H12_o$.

***Calculation of the probability of overlap of peaks in the Zambian and Raleigh scans***

To calculate the probability of 3 or greater overlapping peaks in the Zambian and Raleigh scans in Figure 2, we calculated the fraction of the autosomal arms in each genome covered by H12 peaks. In Raleigh 1.51% of the autosomal arms are covered by H12 peaks (corresponding to 4,252,692 bps) and in Zambia 4.38% of the autosomal arms are covered by H12 peaks (corresponding to 1,461,431bps). The probabilities of observing three out of 25 peaks in Raleigh overlapping three out of 25 peaks in Zambia can be computed with a right-tailed binomial probability, where the probability of an overlap in Zambian or Raleigh data corresponds to either 1.51% or 4.38%, respectively. Thus the probability of three or more overlaps in the Raleigh data is 0.0224 and the probability of 3 or more overlaps in the Zambian data is 0.0005.

***Test for structure, admixture and backflow generating H12 peaks in Zambian data***

We tested for an enrichment of strains among the top three haplotypes across all the peaks in the Zambian H12 scan. To calculate significance, we performed a permutation analysis. We shuffled

the labels of strains comprising all the haplotypes at each peak and counted the number of strains appearing among the three most common haplotypes in *n* or more peaks, where *n* ranged from 0 to 25 peaks. We repeated this procedure 10,000 times. Then, comparing against the distribution of the number of repeated strains among the top three haplotypes for *n* or more peaks, we calculated an empirical P-value for observing 15 strains appearing among nine or more peaks' top haplotypes. We did the same for all other *n*s and found that for lower *ns* we did not get significant P-values.

We tested for the possibility of backflow of strains from North America to Zambia generating the H12 peaks in the Zambian data set. First we tested for a dip in diversity in the Raleigh data at positions corresponding to the top 25 peaks in the Zambian data set. We computed $\pi$ in 10Kb non-overlapping windows on all autosomal arms in the Raleigh data. We then calculated the mean $\pi$ in any window overlapping the start and end coordinates of each of the top 25 peaks identified in the Zambian data set and compared this mean with the distribution of $\pi$ calculated in all remaining windows in the Raleigh data set.

***Test for enrichment of peaks in ancestry disequilibrium hubs***

To test whether the top 25 peaks in the Raleigh data set are enriched in ancestry disequilibrium hubs identified by POOL 2015, we first identified that the end coordinates of the first and last windows of two peaks overlap the coordinates of the hubs listed in Table S6 of POOL 2015. We then calculated the fraction of the autosomes that the hubs occupy (8.1%) and then calculated a binomial probability of observing two out of 25 peaks overlapping any hub at random.