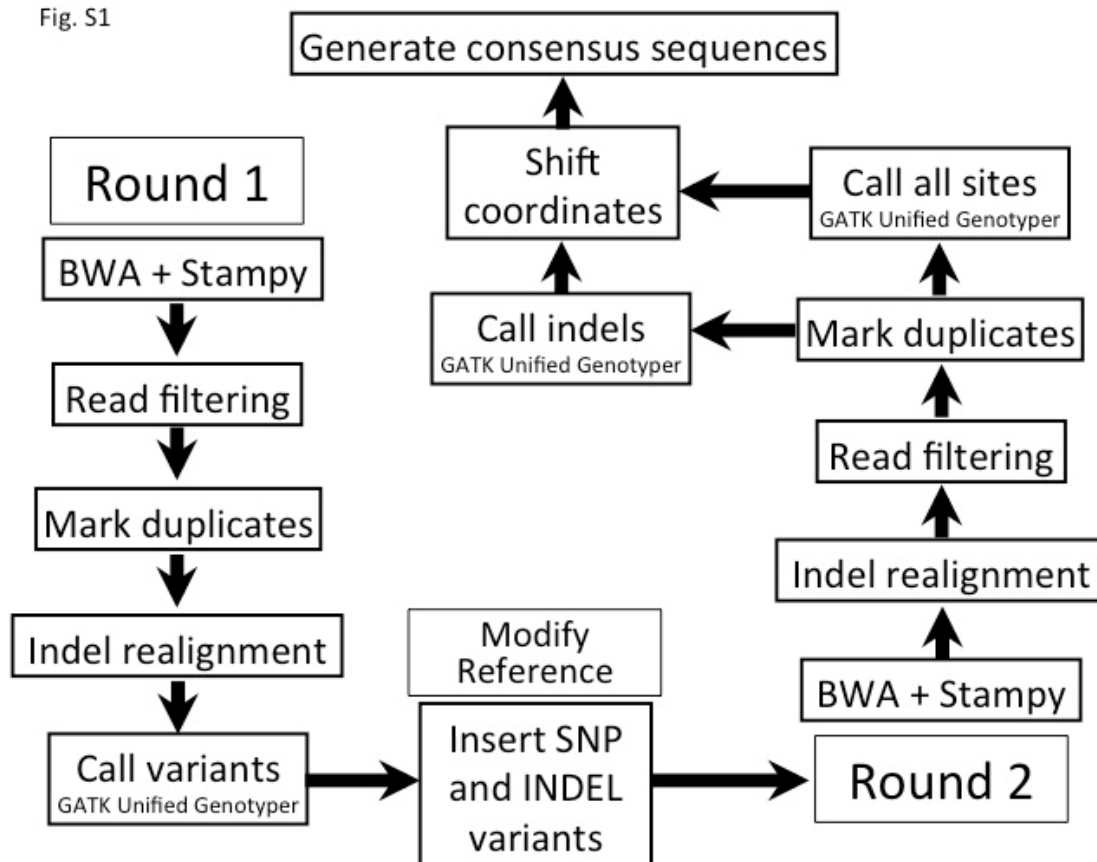
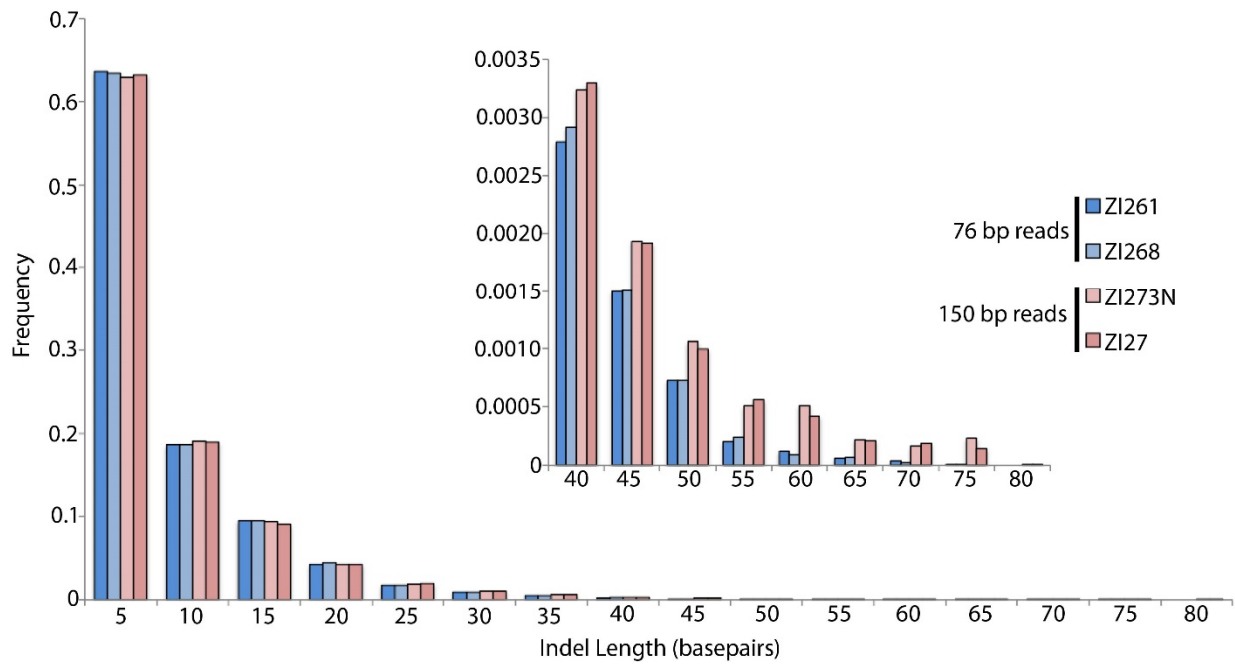


Fig. S1



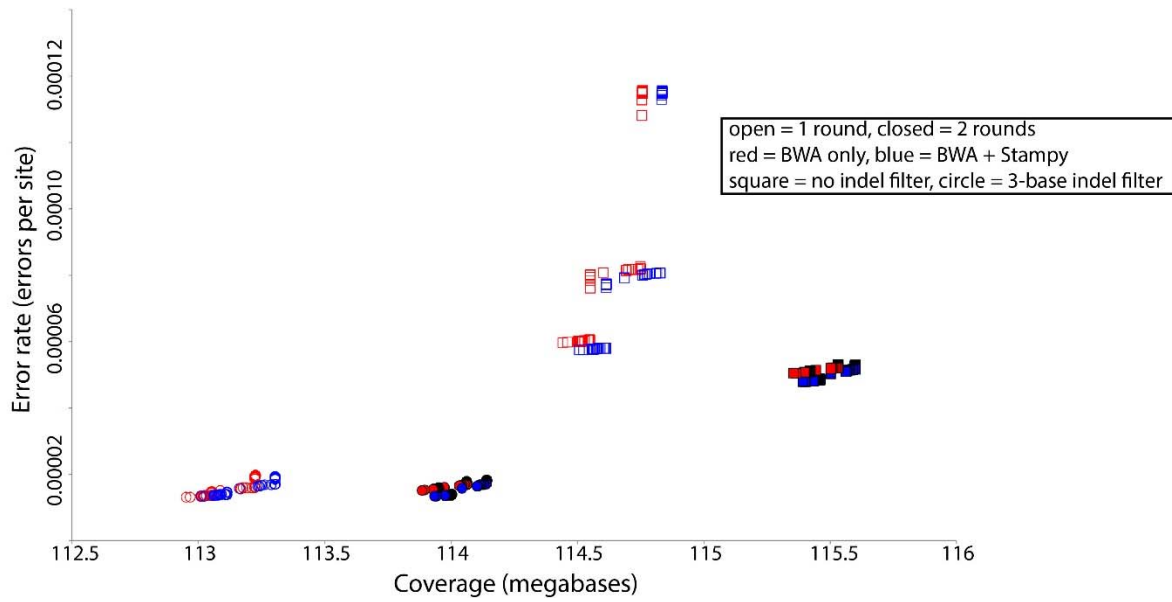
**Figure S1** Graphical depiction of the two round assembly pipeline.

Figure S2.



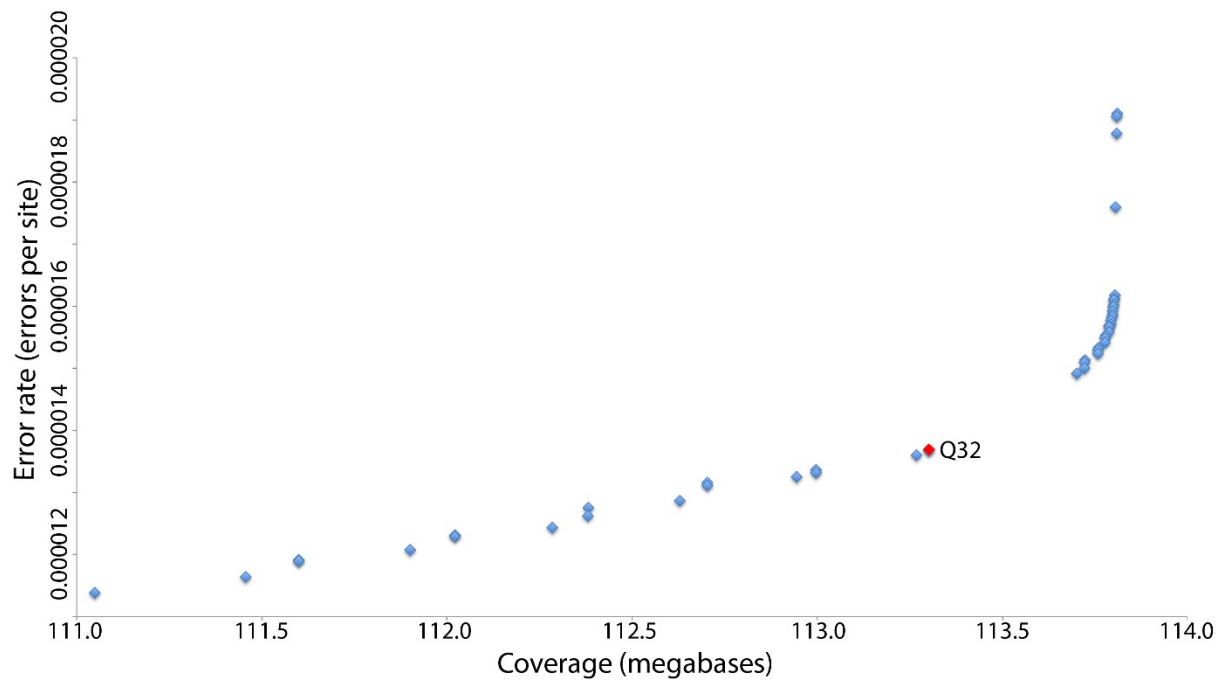
**Figure S2** Length distributions for called indels for 46 bp (blue) and 150 bp (pink) read lengths. The inset zooms in on the frequencies for lengths  $\geq 40$  bp.

Figure S3



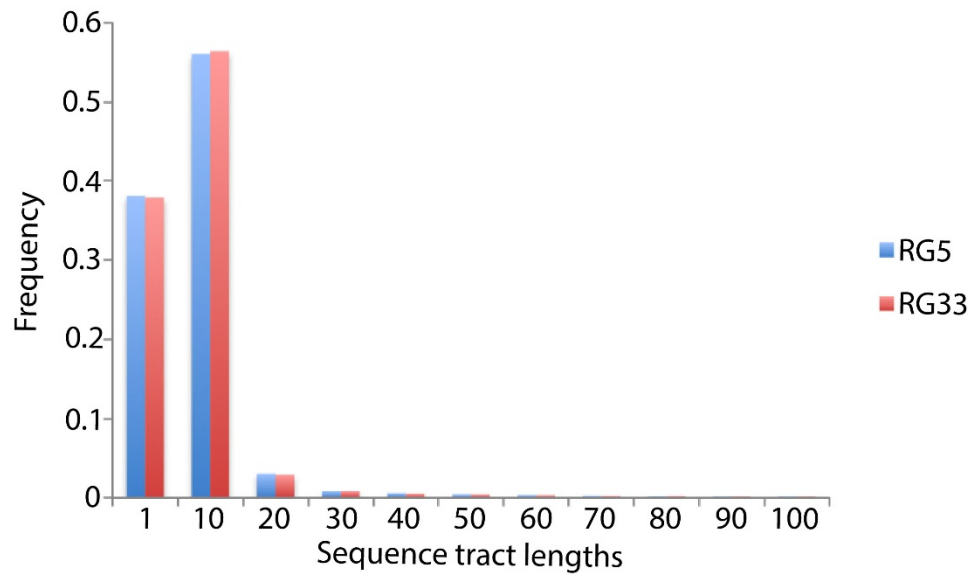
**Figure S3** Comparison of genomic coverage and error rate for various genome assembly pipeline variations, based on resequencing of the *D. melanogaster* reference strain (Pool *et al.* 2012). All quality values from Q10 to Q100 are shown; many gave very similar results. Open symbols indicate a single round of mapping, while closed symbols indicate two rounds of mapping. Red symbols indicate only BWA was used to map, while blue symbols indicate BWA and Stampy were both used. Square symbols indicate no indel filter was applied, while circular symbols indicate a 3-base filter was applied.

Figure S4

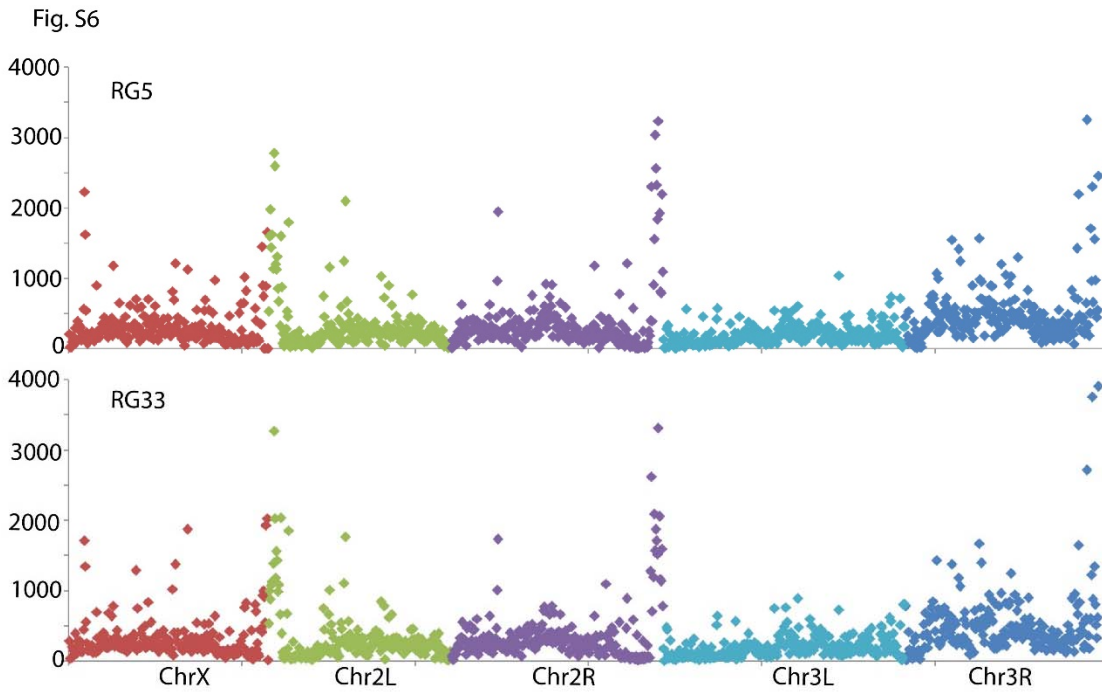


**Figure S4** Evaluation of the tradeoff between genomic coverage and error rate for the diploid caller of the Unified Genotyper; quality values ranged from 10 to 100. Resequenced genomes from the reference strain ( $y^1 cn^1 bw^1 sp^1$ ) were modified to simulate realistic levels of variation. We chose a cutoff of Q32 (red) to maximize coverage and minimize error.

Fig. S5

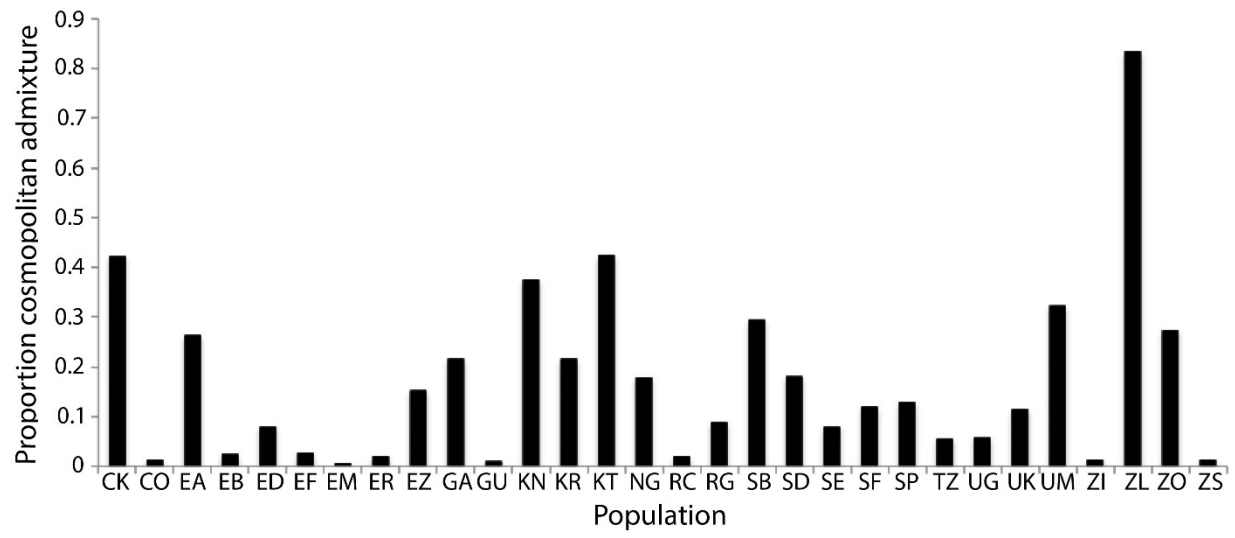


**Figure S5** Histogram of sequence tract lengths for sites added by our two round pipeline for the Rwandan genomes RG5 and RG33.

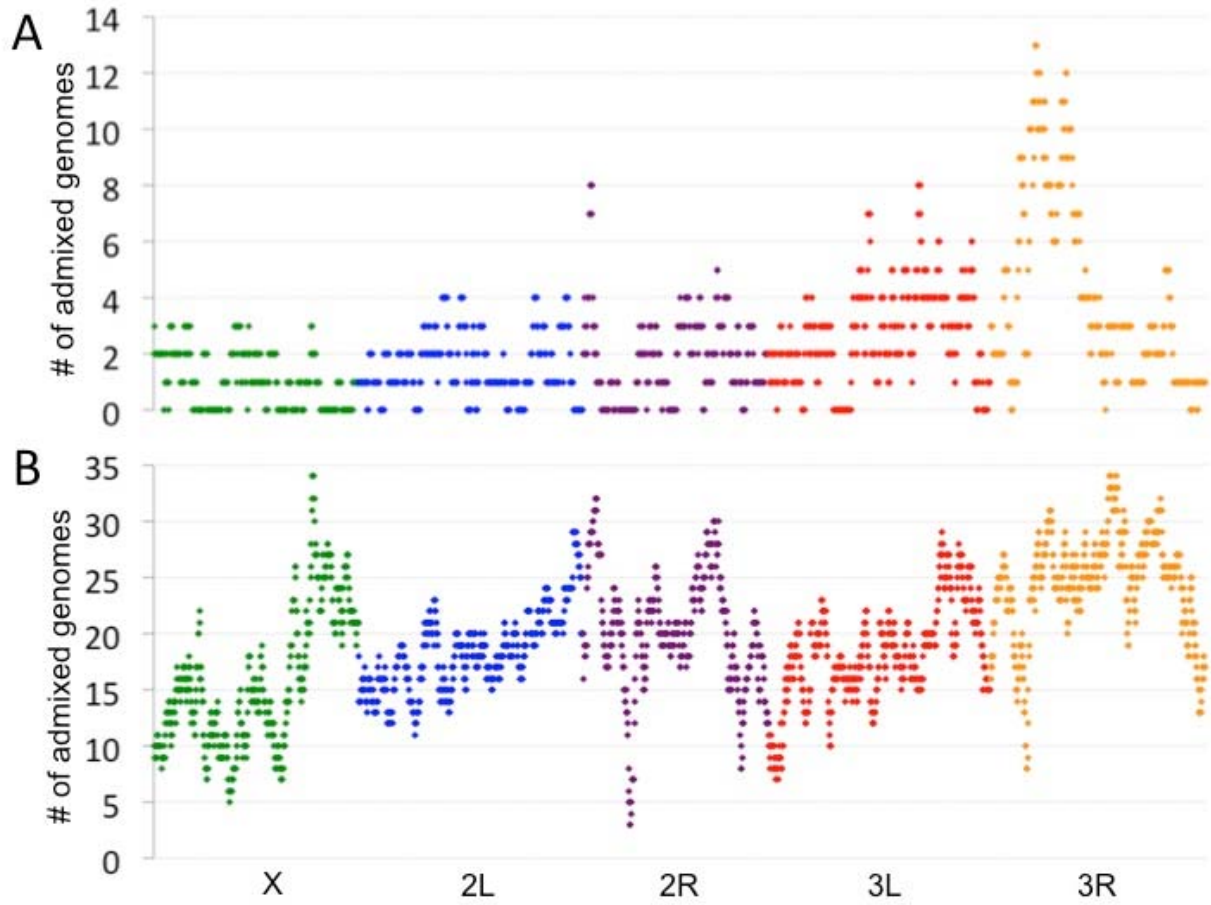


**Figure S6** Number of sites added by our two round assembly pipeline in 100 kb windows across the 5 euchromatic chromosome arms for the Rwandan genomes RG5 and RG33.

Figure S7



**Figure S7** Variation among African populations in estimated cosmopolitan admixture proportions.



**Figure S8** Numbers of sub-Saharan genomes inferred to have cosmopolitan ancestry in each genomic window: (A) For the DPGP3 Zambia ZI sample, and (B) across all other sub-Saharan populations. Windows are depicted for arms X (green), 2L (blue), 2R (purple), 3L (red), and 3R (orange).



**File S1**

**Supplemental Text**

Available for download as a PDF file at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.174664/-/DC1>

## Tables S1-S10

Available for download at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.174664/-/DC1>

**Table S1** Individual sequenced genomes included in this data release, including fly stock ID, genomic library ID and type/source for each library, NIH SRA access numbers, focal chromosome arms, read length, coverage and mean depth of focal chromosome arms, and the data set from which the original sequenced reads originated.

**Table S2** Population samples from which the sequenced genomes originated. The number of sequenced individuals for each focal chromosome arm is given.

**Table S3** Coordinates of residual heterozygosity tracts and pseudoheterozygosity tracts filtered from genomes, and proportions of true heterozygosity and total masked heterozygosity for every genome in the data set. The distinction between the masked proportion and true heterozygosity proportion is due to the presence of artefactual heterozygosity (pseudoheterozygosity) resulting from mismapping or technical issues with individual libraries.

**Table S4** Inversions detected from fly stocks via cytology (DGRP; Huang et al. 2014) or from genomes via bioinformatics (Corbett-Detig and Hartl 2012). Note that in the case of the haploid embryo genomes, live stocks may harbor undetected inversion polymorphism. "INV/ST" indicates known polymorphism. "INV/?" indicates that inverted reads were detected, but the genome was heterozygous in this region. Blank cells indicate inversions that were untested or unreported for this genome/stock.

**Table S5** Recurrent identity-by-descent (IBD) tracts for the each data set. Only IBD tracts outside of these regions were allowed to contribute to individual totals.

**Table S6** Regions of IBD masked from the analyzed genomes, including both individual genomes identified for each tract. See the methods for a detailed description of IBD detection and filtering criteria, and Table S4 for the excluded recurrent IBD regions.

**Table S7** Inversion polymorphism and proportion heterozygosity on each focal chromosome arm for each DGRP genome, illustrating the role of inversions in maintaining heterozygosity in spite of considerable inbreeding effort. "Pseudoheterozygosity" corresponds to the proportion of a chromosome arm prior to normalization, and "Corrected heterozygosity" corresponds to the proportion of a chromosome arm following normalization.

**Table S8** Regions of cosmopolitan admixture masked in Sub-Saharan African genomes.

**Table S9** Genome-wide genetic differentiation and nucleotide diversity for populations with multiple high-coverage focal chromosomes, averaged across the five focal chromosome arms. Values below the diagonal are  $F_{ST}$ , values above the diagonal are  $D_{xy}$ , and bold values on the diagonal are nucleotide diversity. Distance from the *D. melanogaster* reference genome is given in the bottom row.

**Table S10** Individual chromosome arm measures of genetic differentiation and nucleotide diversity for populations with at least two high-coverage sequences. Values below the diagonal are  $F_{ST}$ , values above the diagonal are  $D_{xy}$ , and bold values on the diagonal are nucleotide diversity. Distance ( $D_{xy}$ ) from the *D. melanogaster* reference genome is given in the bottom row.