### 

### 

### Supplemental Figure 1. Median haplotype frequency across all windows on chromosome 2L for each founder (n=99), calculated with different window sizes and empirical coverages. Haplotype frequencies calculated before imputation (red circles) and after imputation (blue circles) are plotted as a function of the log of the total number of ambiguous genotypes (aka “N-count”). Best fit lines for each dataset were calculated with standard linear regression.

# 

**C**

**B**

**A**

### Supplemental Figure 2. An example of true and predicted allele frequencies at each segregating site on chromosome 2L, where predicted frequencies are calculated either from A) raw mapped reads at 5x empirical coverage, B) HAFs at 5x empirical coverage, C) simulated binomial sampling of reads at 462x coverage. Color represents density of points. RMSE for each set of predictions is indicated in the top left of each panel. Note that RMSE for panels B and C are very similar; this equivalence forms the basis of assigning an ‘effective coverage’ of 462x to the estimated allele frequencies in panel B.

# 

### Supplemental Figure 3. Contribution of DNA from each pooled individual in experimental replicate 1, estimated by average genome-wide allele frequency across all singleton sites. The dashed line represents theoretical expectation for evenly pooled individuals. Error bars represent total expected binomial error, given total read depth at all singleton sites for a given founder.

### 



### 

### Supplemental Figure 4. Effective coverage was calculated for samples simulated at 5x empirical coverage after 5,15, and 50 generations of weak selection, with a founder genotype table missing 1% of calls, using various window sizes for haplotype inference. Colors correspond to the quantiles of the expected exponential distribution of unrecombined fragment lengths that were used as the window size for haplotype inference. Each panel (1-3) represents results from a different simulation round, using a different set of selected sites.

****

### Supplemental Figure 5. Effective coverage for 3 separate simulated long-term experiments each with 5 randomly selected sites under selection (S=0.025), simulated empirical coverage of 5x, and no missing founder genotypes.

### 

### 

### Supplemental Figure 6. True population-wide allele frequencies (grey lines), true sampled chromosome allele frequencies (closed black circles) and HAFs (open circles) calculated at sites under selection (S=0.025) from samples simulated at 5x empirical coverage after 5,10,15,25, and 50 generations of recombination, using founder information with various fractions of missing of founder genotype calls (color).

### 



**C**

**D**

**B**

**A**

### Supplemental Figure 7. Relationship between effective coverage, number of reads per window, and percent of missing genotypes. The plots in the top row (A-B) indicate that the relationships are not linear. The plots in the bottom row (C-D) (where the x- and y-axes have been adjusted to log scale) suggest that the relationships are approximately log-linear.