

## File S1

### Supporting Methods

#### Predicting Physical Resolution of Mapping

How much resolution can one expect from PSiSeq? The following parameters influence choice of bin size: recombination rate, gene frequency, genome-wide sequence divergence, quality of reference assemblies (or transcriptomes), and level of replication.

For example, *D. simulans* and *D. sechellia* are roughly 2% divergent (KLIMAN *et al.* 2000), and given the quality of each reference genome assembly we captured half this divergence in our SNP map (1%). At 1,000 SNPs/bin this provides 100kb resolution. We chose 1,000 SNPs/bin given *Drosophila* gene frequency (~1 gene per 12kb or ~8 genes per 100kb; calculation based on annotation from FlyBase *D. melanogaster* r5.37).

$$S = \frac{B}{(x/2)}$$

Where,  $B$  is bin size, or #SNPs/bin;  $x$  is average sequence divergence, and  $S$  is expected physical size resolution. Thus, for populations 1% divergent with SNP markers capturing half this divergence, bin sizes of 1,000 SNPs/bin will provide 500kb resolution. With an average of one crossover per chromosome per generation, fifteen generations of introgression will isolate 0-2% of the mappable genome (see Figure 2). A resolution of 500kb in *Drosophila* easily captures 2% of the genome (~3.6mb).

#### PSiSeq Guide

Included in a separate .pdf is a step-by-step guide for our informatic pipeline.

This guide, File S2, is available for download as a compressed folder at <http://www.genetics.org/content/suppl/2011/09/21/genetics.111.129445.DC1>.