

## Scripts for Quantifying Allele-Specific Expression

This collection of scripts was used to quantify allele-specific expression in *Drosophila*, as described in Stevenson et al. 2013 *BMC Genomics* 14: 536. They require the following commonly used and publicly available software: Bowtie, SAMtools, BEDtools, the GEM library, and Fast Statistical Alignment (FSA). These programs and custom Perl and R scripts were used to simulate and analyze sequence reads from two strains of *Drosophila melanogaster* and between *D. melanogaster* and *D. simulans*, but can be adapted for use with other genomes. Real sequencing data can be used in place of the simulated sequence reads. Included are scripts to determine the density of differentiating sites and mappability, two important factors that affect accuracy of inferred measures of allele-specific expression.

We used these scripts in Stevenson et al. 2013 *BMC Genomics* 14:536 **Highly accessed** as follows:

Within-*D. melanogaster* comparison:

- 1a). For a particular line of the *Drosophila* Genetic Reference Panel (DGRP), find all homozygous genotypes (compile\_genos.pl)
- 2a). Compare these genotypes to those in the *Drosophila melanogaster* reference genome, looking for sites that differentiate the reference and DGRP (or "alternative") alleles, find these sites, and create an edited version of the genome to match these genotypes (edit\_genome.pl)
- 3a). Tile sequence reads across regions of the genome (constitutive exons, in our case, but could be any set of regions). This is done for each allele and on each strand (forward strand and reverse-complementary strand). (tiled\_reads.pl)
- 4a). Align sequence reads to genome using Bowtie and get site-specific abundance measurements for each allele using SAMtools (bowtiePipe.pl, samtoolsPipe.pl, pileup\_SNP\_ASE.pl)
- 5a). Determine the density of differentiating sites, i.e. for each differentiating site, the maximum number of differentiating sites within the length of a sequence read (count\_neighbors.pl, create\_windows.pl)
- 6a). Determine site-specific measurements of genome mappability, i.e. ability to uniquely align sequence reads from a particular location to the genome, using the GEM library (get\_mappability.pl)

Between *D. melanogaster* and *D. simulans* comparison:

- 1b). Using the genomes of resequenced lines of *D. melanogaster* and *D. simulans* from the McIntyre laboratory at the University of Florida, tile sequence reads across all orthologous exons (generate\_reads.pl)
- 2b). Using FSA, perform pairwise alignment of each orthologous exon to determine location of differentiating sites as well as indels (pairwise\_aln\_FSA.pl, compare\_pairwise.pl, find\_indels.pl)
- 3b). Align sequence reads to genomes, as above in (4a).
- 4b). Determine density of differentiating sites, as above in (5a).
- 5b). Determine site-specific mappability, as above in (6a).

Questions about using these scripts or adapting them for other uses can be sent to Patricia J. Wittkopp <wittkopp@umich.edu> or Kraig Stevenson <kraigrs@umich.com>.