BIO472: Problem set #3. Due 4/21/14

Background:
Meiotic drive describes the biased transmission of a genetic element during gametogenesis. One of the best-studied meiotic drive systems is the autosomal sperm killer of Drosophila melanogaster—Segregation Distorter (SD)—a meiotic drive complex on chromosome 2 (Reviewed in Larracuente and Presgraves 2012). Males heterozygous for SD transmit the SD chromosome to >95% of their progeny. SD targets a satellite DNA repeat in the pericentromeric region of chromosome 2R called Responder (Rsp). Different alleles of Rsp segregate in natural D. melanogaster populations—those that are sensitive to segregation distortion (Rsp\(^s\)) have more repeats than alleles that are insensitive to segregation distortion (Rsp\(^i\)). SD chromosomes carry Rsp\(^i\) alleles that have few, if any, Rsp repeats. Several upward modifiers of SD exist on SD chromosomes that strengthen drive. Many SD chromosomes pick up local inversions to reduce recombination between the driver and the target (prevents suicidal genotype formation) and presumably to keep modifiers of SD linked to the driver. Though beneficial to SD chromosomes for drive strength, the suppression of recombination is predicted to come at a cost: the efficacy of natural selection is reduced in regions of low recombination and these regions should be prone to deleterious mutation accumulation. Consistent with this prediction, SD chromosomes are frequently sterile or lethal when homozygous. SD chromosomes exist at a low but stable frequency of 1-5% worldwide, but show surprisingly dynamic evolution: in Africa, SD chromosomes underwent a recent selective sweep where one SD chromosome has replaced another (Presgraves et al. 2009).

References:

Online references:
https://github.com/srobb1/RelocaTE/tree/master/sample_relocaTE_run

Assignment:
You have been given directories (in /scratch/bio472_2014/data/TEs ) that contain de novo TE annotations from the program relocaTE for an SD or wild-type genome. Please complete the analyses below, answer the questions below, and place a tarball containing your responses, code and output files in /scratch/bio472_2014/Users/completed_assignments. Please refer to the online reference for RelocaTE for details on how the TE calls were made.

Assigned samples:
PS1 (SD-Roma) –Mary
PS2 (SD-NK04)—Purba
PS3 (SD-Madison)—Emerson
1. Count the number of TEs found on chromosomes 2L and 2R. (Hint: Parse the gff file called ALL_INSERTS.tename.gff found in each directory)

2. How many are unique to the SD (or Raleigh) chromosome?

3. How many are unique to the reference chromosome?

4. How many are shared between the reference the SD (or Raleigh) chromosome?

5. What is the breakdown of TE types on 2L and 2R for the SD (or Raleigh) chromosome (i.e. counts per TE type)? Create a plot in R to represent this information.

Discussion questions:
Can you draw any conclusions about the genetics/population genetic history of the SD (or Raleigh) chromosome by comparing it to the reference chromosome? What did you expect to see? Can you make any more sense of the pattern if you integrate the SNP calls from problem set 2 with the TE insertions?