BIOS 247 Mini-Course, Summer 2024

Whole-Genome Sequencing: From Yeast to Fruit Flies

**Exercise 2: Curating SNPs in IGV**

In this exercise, you’ll look up variants that were called in our WGS pipeline and “curate” them, or decide if they’re really the underlying cause of mutants’ adaptation. You will use IGV to load five mutants and compare them to a reference genome.

*Note: This exercise is meant to simulate the process of curating SNPs from a large group of mutants. If you see mutations in more than one mutant in the same gene, assume this appears in many mutants in the entire group. If you see a mutation in a gene in only one mutant, assume it is the only time this gene is mutated in the entire group.*

Directions:

1. Open the IGV desktop user interface. In the upper left dropdown menu, choose the reference genome: “S. cerevisiae (sacCer3).”
2. To load the five genomes you downloaded from Dropbox, go to File -> Load from File. Upload the five .bam files. (Note: the corresponding .bai files need to be in the same folder, but you do not need to upload them.)
3. Adjust height of sub-windows so that you can see all five genomes.
4. Download the table called\_variants.csv from the data folder in [GitHub](https://github.com/clare-abreu/WGS_Bios247/tree/main/data/called_variants). The table lists variants that were called by the computational pipeline. It is sorted by chromosome location, so that you can search one chromosome at a time, but take note of which genome each line refers to.
5. Go through the variants by looking up their location in the given chromosome. Take note of whether the SNP exists and anything else you notice.

Questions:

1. After going through all of the called variants, did you find any mutations that appear repeatedly in the same gene in more than one mutant? If so, look up the gene in the [SGD](https://www.yeastgenome.org/). Given the source environment (where the mutants evolved), how could these mutations be relevant?
2. In one of the genomes, you’ll find a SNP that appears in about half of the reads. What does this suggest about this organism and whether its adaptive fitness is related to the SNP? (Hint: Our ancestor is a haploid, meaning it has one copy of all chromosomes, but pure auto-diploids have a fitness advantage.)
3. Go back to the barcode region in Benchling and find the gene in which the barcode sits. Look up this gene in the [SGD](https://www.yeastgenome.org/) and find its chromosome location. Do you see any SNPs, and what might be their origin, given that the gene was deleted and replaced in the ancestor? Based on the coverage plots, can you see what part of the sequence was deleted and replaced? (Hint: colored bands indicate the paired read mapped to another chromosome.)
4. Go back to Benchling and find the gene the region displaced (note the annotation at both ends of the region). What does that gene look like in you IGV plots? What do your observations here and above suggest about the limits of identifying structural variants with short-read sequencing?