BIOS 247 Mini-Course, Summer 2024

Whole-Genome Sequencing: From Yeast to Fruit Flies

**Exercise 1: Aligning Sanger sequencing data to a reference**

In this exercise, you’ll align Sanger sequencing data to a reference sequence. We obtained this data by amplifying the barcode region of the barcoded yeast used in the Petrov Lab. You will align several different yeast mutant sequences to the reference in order to see their exact DNA barcodes.

Directions:

1. Load the reference sequence (barcode-region-lb.gb in the data/Sanger\_data folder in Github) in Benchling by clicking on the upper left **+** then **DNA/RNA sequence** then **Import DNA/RNA sequences**.
2. On the right, click **Alignments**A black and white text box

   Description automatically generated with medium confidencethen **Create New Alignment**.
3. Click **Choose Files** and upload the .ab1 files in the Sanger\_data folder.
4. Complete the alignment (default settings).
5. Click **Open** in the lower left to view the alignment. Scroll to the aligned region along the bottom of the screen.

Questions:

1. The primers SK45 and SK49 were used to amplify this DNA. Is there asymmetry in the accuracy of the alignment near the two of these primers?
2. Only the forward primer (SK45) was included with the DNA for Sanger sequencing, meaning that only forward reads were generated. How might this explain the asymmetry in accuracy? (Hint: molecular dye adds weight to DNA fragments.)
3. Why would Sanger sequencing become inaccurate for very long fragments of DNA?