7.003 Spring 2022
Day 13 BLAST Gene Identification Worksheet

This worksheet will help you identify the mutated genes in your verified candidate mutants from your transposon mutagenesis screen. Please follow the instructions below and answer the three questions at the end in your Day 13 Post-Lab.

Prepare sequence result:

1) Go to the "Day 13 Class Sequencing Results" file posted on Lab Day 13. Find your three sequences. Did you get a good contiguous read (without too many Ns) on all of them? Note that if you have too many N's, you might not get good results in the BLAST search.

2) Use your mTn3 + pRSQ2 SnapGene .dna file to find the sequence of the terminal repeat (TR) at the end of the transposon. Scan the beginning of your sequencing result for this TR sequence (make sure you're using the correct strand sequence in the correct orientation!). The TR sequence should appear early on in the sequencing result, since the M13(-40) sequencing primer is located pretty close to the end of the mTn3 transposon. The TR sequence will mark the transposon-genomic DNA junction. The sequences immediately DOWNSTREAM or AFTER the TR sequence are the genomic DNA sequence to be used in your analysis for all the subsequent steps.

3) From the sequences immediately AFTER the transposon TR sequence, try to find a stretch, which does not have any Ns (ideally at least several hundred nucleotides long). It does not necessarily have to be long though – sometimes, a minimum of 20 nucleotides may be enough to provide a significant match.

4) Use SnapGene to determine if there are any EcoRI sites in the sequence stretch you selected. You may trust your sequence only up to the first EcoRI site. (*Why?*) If there are no EcoRI sites, you may now search your sequence as is in the BLAST database. If there is an EcoRI site, copy and paste only the sequences upstream or before the EcoRI site to perform your BLAST search.

BLAST search: BLAST (Basic Local Alignment Search Tool) compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.

5) The web site for the BLAST database can be found below (a link is also available under 7.003 Lab Resources):

http://blast.ncbi.nlm.nih.gov/Blast.cgi

6) Once you are on the BLAST site, click on the *Nucleotide BLAST* option (also known as a "nucleotide → nucleotide" or "blastn" search).
7) Copy and paste your sequence into the "Enter Query Sequence" window.

8) Make sure "Standard databases (nr etc)" is selected in the Database options. Enter "Saccharomyces cerevisiae" into the Organism bar (use the "(taxid:4932)" option that appears in the drop-down menu). Click BLAST at the bottom of the page. Wait until your query is run (this may take up to a minute) and results are returned. If your query does not turn up any matches, then you may leave the organism bar blank to see where the sequence came from (e.g. sequences from the pRSQ2 plasmid which has a bacterial origin may show up).

9) When you get the search results, look for names of genes in the search results. Typically some results will turn up in the form of 'Chromosome I complete sequence.' What you are looking for are three-letter gene names (e.g. *CDC5, HIS3, STE18,* etc). If you cannot find any gene names but only sequences of chromosomes, go to Step 10.

10) If you do not get any specific gene names from the blastn search (or if you want to confirm the results from your blastn search), go back to the BLAST homepage (as shown in Step 5) and click on the *blastx* option where you can search protein databases using a nucleotide query. The blastx program will translate the nucleotide sequence to protein sequences in all possible different frames and look for any matching protein sequences in the database.

11) To use blastx, copy and paste your DNA sequence into the "Enter Query Sequence" window, enter "Saccharomyces cerevisiae" into the Organism bar (use the "(taxid:4932)" option again), and click BLAST at the bottom of the page.

12) If both blastn and blastx fail to produce an actual gene name result (or if you want to definitively verify your BLAST results), you can use the Saccharomyces Genome Database (SGD) website to analyze your sequences. SGD provides extensive curated information on genes and gene products, as well as tools for homology searches and other functional analyses. The SGD website can be found at the link below (a link is also available under 7.003 Lab Resources):

http://www.yeastgenome.org/

13) To perform a sequence homology search on SGD, go to the "Sequence" drop-down menu at the top of the SGD homepage and click on "BLAST," which will take you to the *S. cerevisiae* NCBI-BLAST Search page.

14) Copy and paste your DNA sequence into the "Type or Paste a Query Sequence" window. Click on the "Run NCBI-BLAST" button.

15) The SGD BLAST search may come up with several different sequence matches. Look for the best sequence match result that has the longest stretch with the greatest

level of sequence homology (highest percentage of nucleotide Identities). Click on the "Genome Browser" link for that best sequence match result.

16) The Genome Browser view will show you a diagram of where exactly in the yeast genome your sample sequence is located (i.e. where your transposon is inserted). The query sequence you entered will be highlighted in yellow in the Genome Browser view. If needed, you can use the Zoom tools (magnifying glass icons near the top center of the page) to zoom in or out to give you a better view of the genomic region where your transposon is inserted. The Genome Browser diagram will indicate any annotated genomic elements (e.g. genes, open-reading frames, etc) present in the region you are viewing. Observe where your transposon inserted (it will be at one of the edges of the yellow highlighted region (*how will you know which edge the transposon inserted at?*)) and look for any gene sequences that it may have disrupted. (*Note that a transposon doesn't necessarily have to insert directly within a gene's coding sequence to disrupt that gene's function – how might a transposon that inserted upstream or downstream of gene still possibly affect that gene's function?*).

17) Once you have identified all the genes for your αF-resistant mutants, you can search for information on each of your identified genes by entering the gene name in the search box in the upper right corner of the SGD homepage. This will take you to the Summary page for your gene. Here, you can find a brief overview description of your gene and its function. There are also multiple tabs/sections (e.g. Sequence, Protein, Phenotype, etc.) at the top of the page which you may browse through to obtain more information about your gene.

For your Day 13 Post-Lab, please include the following information:

1) List all of the genes that came up in your screen. If one of your sequences did not produce a valid gene result, you should indicate as such and then use a sample back-up sequence in its place (open the "Day 13 Back-up sample sequences for BLAST" file posted on Lab Day 13 and choose any of the sample genes provided to analyze in its place). Regardless of whether or not a student's sequence results "worked," all students should be able to write about three genes total for their Post-Lab (and in their SciComm paper!).

2) Using information from the SGD database (or from a PubMed literature search), compile a one- or two-sentence summary about the function of each gene, if known.

3) For each gene, explain how you think this gene came up in your screen. Try to elaborate a molecular mechanism by which the disruption of this gene could have potentially caused α-factor resistance in yeast.

7.003 Applied Molecular Biology Lab
Spring 2022