NAME:
DATE:

HC70AL
SUMMER 2014
PROFESSOR BOB GOLDBERG
Gene Annotation Worksheet

QUESTION ONE
Using primers given to you by your TA, you carried out sequencing reactions to determine the identity of the gene that you'll be uncovering the function of this quarter. You began to process the sequence files using what you've learned from Kelli's presentation on Bioinformatics.

A. Print out copies of your sequence files using the program that Kelli mentioned in her presentation (FINCH TV or 4Peaks). Attach all the sequence print outs to your lab report.

B. Did your sequencing reactions work? If no, what are some things that could have gone wrong?

C. Copy and paste the final sequence obtained with your gene-specific salk primer.

D. Go to the TAIR web site (http://arabidopsis .org) and click on the BLAST link under "Tools." Copy and paste the sequence in Part C in the window. Set the parameters for the BLAST run as follows: BLAST program: BLASTN & Datasets: TAIR10 Genes (+introns, + UTRs) (DNA). Click on Run BLAST. Copy and paste the result with the lowest e-value. This is the gene that you'll be working with this quarter. The gene should have an Arabidopsis Gene Identification number (e.g. AT2G32370). Write down this AGI number.

QUESTION TWO

You are working on an *Arabidopsis* gene that you hypothesized to have an important function in seed development. You decided to study the function of this gene by knockout analysis. You hypothesized that if the gene is important for seed development, knocking out the gene will result in seed lethality. Before you proceed to knockout analysis, you must characterize your gene.

A. What is your Arabidopsis Gene Identification number (ex. AT2G32370)? Briefly explain what the letters and numbers in the gene identification number mean.

B. What is the size of your gene in base-pairs?

C. What is the specific map location of your gene in the genome (chromosome and nucleotide positions)?

D. What is the orientation of your gene in the chromosome?

E. What gene is 5', or "upstream," to your gene (with respect to ATG)? Specify the gene identification number.

F. What protein does this neighbor gene encode?

G. How far is the 5' upstream-gene from your gene in base-pairs?

H. What gene is 3', or "downstream," to your gene? Specify the gene identification number.

I. What protein does this neighbor gene encode?

J. How far is the 3' downstream-gene from your gene in base-pairs?

K. Draw a cartoon map of your gene, and closest neighboring genes 5' and 3' to your gene. Note the correct orientation of your gene and neighboring genes. Label the map using the information obtained in **Parts A-J above**. Note: Draw your gene in the 5' to 3' orientation.

L. Draw a cartoon structure of your gene. Label the 5' end, the 3' end, exons, introns, and untranslated regions (UTR), if any. Include the nucleotide number at the start and stop of each exon.

M. What is the size of the encoded protein (number of amino acids)?

QUESTION THREE

You want to know the function of your gene; however, you first need to determine what protein your gene encodes.

A. Using the TAIR BLAST program, run BLASTX against the predicted proteins in the *Arabidopsis* genome (TAIR10 proteins). Copy and paste the alignment result with the lowest blast e-value.

B. Include one scientific reference in the literature about your gene from the TAIR website. Print that reference out and read it.

QUESTION FOUR

You were given gene-specific salk primers to carry out a PCR reaction. The sequences of your primers are deposited in the STOCK section of the lab Webbook (http://estdb.biology.ucla.edu/webbook). Using your gene-specific salk primer sequences, answer:

A. Where are the gene-specific forward and reverse primers located with respect to your gene? In an exon? Intron? Upstream of the gene? Downstream of the gene?

B. Draw a cartoon map of your gene with the relative positions of the gene-specific salk forward and reverse primers. Note: Draw your gene in the 5' to 3' orientation.

QUESTION FIVE
You genotyped your plants by performing PCR reactions with gene-specific salk primers and a left border T-DNA primer. You then purified and sequenced the T-DNA allele.

A. Based on your PCR results, what is the orientation of the T-DNA with respect to your gene? Does the left border point towards the 5' end of your gene, or the 3' end?

B. You carried out sequencing reactions to verify the T-DNA insertion site. Print out copies of your sequence files using the program that Kelli mentioned in her presentation (FINCH TV or 4Peaks). Attach all the sequence print outs to your lab report.

C. Did your sequencing reactions work? If not, what are some things that could have gone wrong?

D. Copy and paste the final sequence obtained with the LBb1.3 primer.

D. To verify the T-DNA insertion site, go to the TAIR web site (http://arabidopsis.org) and click on the BLAST link under "Tools." Copy and paste the sequence for your T-DNA allele in the window. Set the parameters for the BLAST run as follows: BLAST program: BLASTN & Datasets: TAIR10 Genes (+introns, + UTRs) (DNA). Click on Run BLAST. Note: If there are no BLAST results, try using a different dataset, such as Upstream Sequences.

On the BLAST Results page, scroll down to the alignment of your query sequence and your gene of interest. At which nucleotide in your gene did the alignment start? This is the T-DNA insertion site. Did the alignment start in an intron? exon? UTR? Upstream region?

E. Does your T-DNA insertion site match what is reported on the TAIR website for your SALK line? (Ask Kelli for the identity of your SALK line)

F. Draw a cartoon map of your gene with the T-DNA and the relative positions of the gene-specific forward and reverse primers and the LBb1.3 primer. Label the left border (LB) and right border (RB) of the T-DNA. Note: Draw your gene in the 5' to 3' orientation.

QUESTION SIX

You want to know where your gene is expressed. Browse the *Arabidopsis* gene expression data on the Gene Networks in Seed Development website (http://seedgenenetwork.net/arabidopsis) created by the Goldberg and Harada labs.

Click on "Analyze GeneChip Data." Under the *Arabidopsis* Analysis Tools, click "Browse." (Note: You will need to register with your email address.) You can now browse the *Arabidopsis* mRNAs Profiling Database.

Scroll down to "Browse by Probe Set or Gene ID." Type in your Arabidopsis gene identification number next to "AGI Locus ID." Click "Submit" at the bottom of the page.

Click on the link for your gene in the "Probe Set" column. You can now view the GeneChip Expression Profile for your gene. (Note: The colors in the legend to the right represent the signal intensity for your gene detected on the microarray, i.e. the abundance of mRNA for your gene in each tissue.)

In which seed tissues and developmental stages is your gene detected? What is the mRNA level in these tissues and stages? Is it high or low?