BIT150 – Fall 2009 – Homework 6

Due on Thursday November 5th by email to TA: astamboliakovach@ucdavis.edu as Hwk6_Lastname BEFORE the Lab

1. **50 points** EST Assembly

Follow the steps described in Lab6 to prepare the directory structure to run Phred::Phrap::Consed using the set of sequences prepared for Homework6. The loblolly pine (*Pinus taeda*) EST sequences you will be working with can be found in your home directory on the ‘plantgenome’ server, in the Hwk6 subdirectory. Show your work below (give me a few sentences on what you did and then give me the fasta sequences of your final contigs, however many there are).

2. **25 points** EST Annotation

Use BLASTX to annotate the resulting consensus sequence of the contig with the largest number of reads that you exported in **Question 1**.

2.1. Record the top 5 significant hits. What are their Scores and the E values?

2.2. Would you affirm they are not just random matches? Why?

2.3. What can you conclude about your EST from the BLASTX results?

3. **25 points** EST resources

3.1. What is the GenBank accession number for the Loblolly Pine (*Pinus taeda*) CCoAOMT (caffeoyl-CoA O-methyltransferase) protein sequence? Include here the protein sequence.

3.2. What type of sequence alignment (BLAST program) should you use to find EST (non-human, non-mouse) hits to this protein sequence?

3.3 What clone number did the most similar Sitka Spruce (*Picea sitchensis*) EST come from? What are its Score and E value?