PHYLOGENY PROJECT –HANDS ON

Project Goal:
The main goal of this project is to use *rbcL* gene sequences to study phylogeny of green plants. We are going to gather DNA sequences of *rbcL* gene from Genbank, and compare them to build a phylogenetic tree.

Learning Objectives:
By the end of this project, SWAT:
1. Recognize the availability of vast pool of molecular data at the NCBI Genbank website (http://www.ncbi.nlm.nih.gov/genbank/).
2. Learn mutations among DNA sequences.
3. Reconstruct tree using freely available software such as ClustalX and MEGA.
4. Understand the meaning of ancestral vs. recent species, clade, and interpret evolutionary relationships among species.

Requirements:
A. Internet access for data mining.
B. Installation of free software
   1. **ClustalX (download from** http://www.clustal.org/download/current/ by choosing clustalx-2.1-win.msi (http://www.clustal.org/download/current/clustalx-2.1-win.msi) for windows and clustalx-2.1-macosx.dmg (http://www.clustal.org/download/current/clustalx-2.1-macosx.dmg) for MAC. This software allows you to align DNA or protein sequences.

   2. **MEGA5 (download from** http://www.megasoftware.net/mega.php for windows and http://www.megasoftware.net/megamac.php for MAC)

Getting Started:

SECTION 1
1. Google "Genbank and click on the link (**GenBank** Home - National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/**genbank**/)
2. Use the dropdown next to the word "GenBank" to change from defaulted "Nucleotide" and select "Gene"

   - <u>Type</u> "rbcl" in the entry box to the right of dropdown and <u>search</u>

3. Select the link: **rbc**L – RuBisCO large subunit [Chlamydomonas reinhardtii] (http://www.ncbi.nlm.nih.gov/gene/2717040)
   - Click on "Genomic regions, transcripts, and products" in the table on contents.
   - Select "FASTA".
   Your new screen should look like this.

4. Copy the sequence of the Chlamydomonas.
    *   Paste into *a NotePad file(PC) or Texteditor(Mac)*
    *   To find *Notepad* go to *Start* menu, *All programs* then click *Accessories* you should see *Notepad*
5. After pasting into *NotePad,* leave the prompt sign">" and delete text before the DNA sequence, Replace deleted text with "Algae"

    Your NotePad should appear as shown below.

Now save this file as "tree.fasta". In the drop down choose "All files". We Now have the *rbcL* gene sequence for one species. Now we will get similar sequences from other plant species.

1. Go to NCBI BLAST website http://blast.ncbi.nlm.nih.gov/
   a. Select "nucleotide blast" under the Basic Blast section.
      b. It will bring you to this window.

Copy the DNA sequence from *NotePad*
   b. 1st arrow. "Enter accession numbers……FASTA sequences"- Paste your algae sequence in the area indicated by the
   c. 2nd arrow. Type *moss* in the organism box. In the dropdown choose "Mosses(taxid:3208)"
   d. 3rd arrow. Choose somewhat similar sequences(blastn)
   e. 4th arrow. Click "Show results in a new window". Click BLAST
            *Wait until program is finished, a new Tab should appear like the one below.
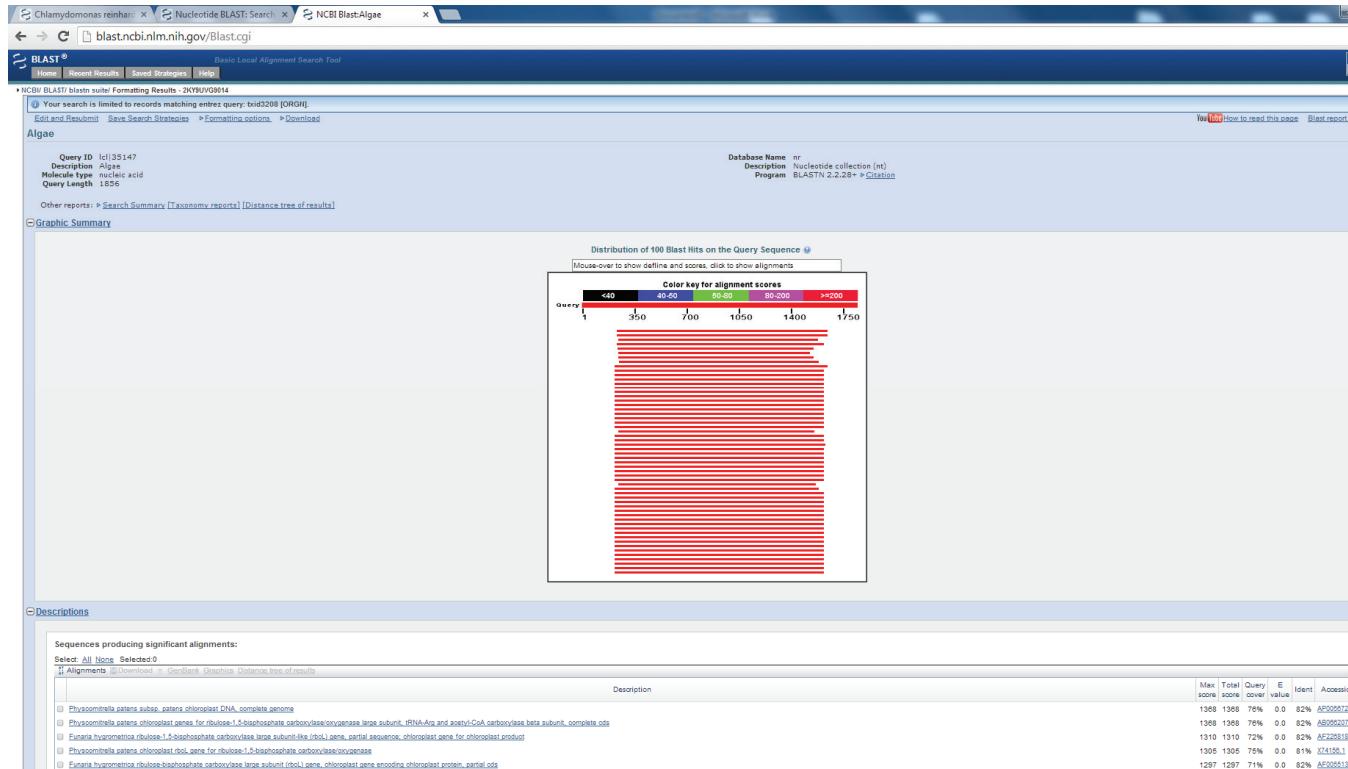
*The first red line on your page represents your reference sequence "Algae". Below that line the other lines contain DNA sequences similar to your initial reference sequence.

7.2- Find the description "Pohlia drummondii chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, complete cds, specimen_voucher: personal:H. Akiyama:21292"

    a. Put a check mark to the left of the text and double click the description.

    b. It will appear as shown below.
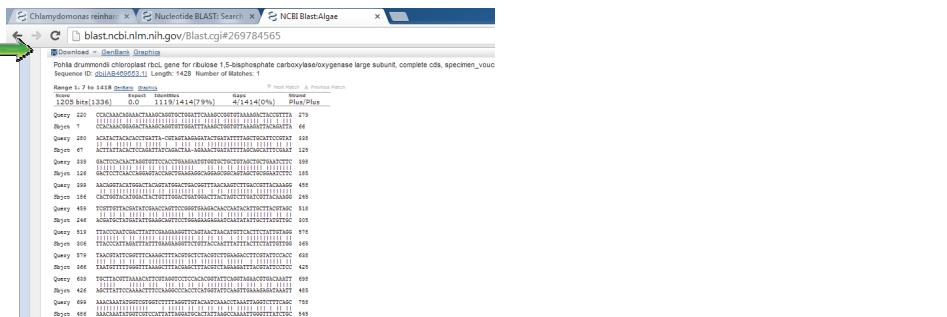
    c. Click *download* and press *Continue*

2.   A new sequence is downloaded. DoubleClick to open the sequence.

3. When sequence is open delete text after ">" but before the beginning of the gene sequence *example below*
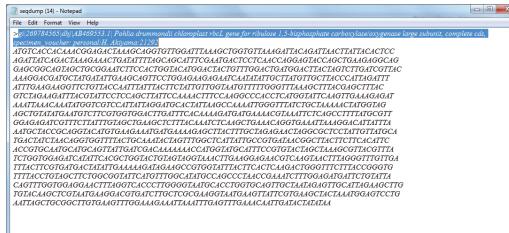
4. Name this new sequence "moss". Hit enter if the sequence and the name came together.
   a. Your window should look like this.

5. Copy and paste this new sequence "moss" into your tree.fasta file
   a. Your tree.fasta file should look like this.
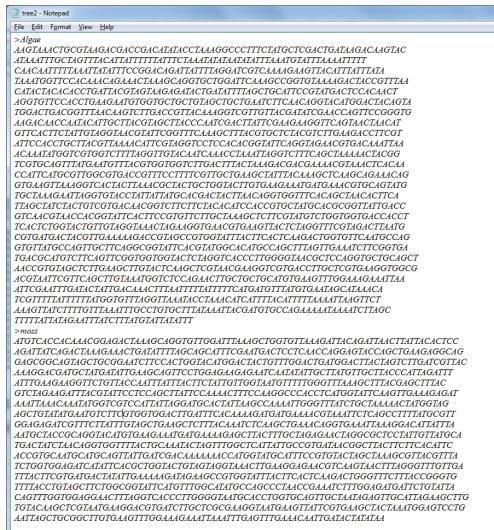
SECTION 3

You now have an *Algae* and *Moss* sequence in your tree.fasta file.

PHYLOGENY PROJECT –HANDS ON

1. You take your Algae sequence and compare to ferns.
    a. Open BLAST window(http://blast.ncvi.nlm.nih.gov/)
    b. Copy your Algae sequence into blast box.
    c. <u>Type</u> *fern* in organisms place use the dropdown to select "ferns (taxid:241806)", and *optimize* for "somewhat similar sequences".
    d. <u>BLAST</u>
2. In new result window check the first result in the "Description" section. You can locate this section after the "red lines".
    a. After checking your selected sequence click Download
    b. Choose "FASTA (aligned sequences).
    c. Click Continue.
    d. Open downloaded sequence and copy it to your "tree.fasta" file give it the name *fern.*

       *Your tree.fasta file should look like this-

SECTION 4

Using the same steps as above you will now BLAST pine, soybean and wheat:

1. Organism: pine family (taxid:3318), optimize for somewhat similar sequences.
    a. After checking your selected sequence (1ˢᵗ option) click Download
    b. Choose "FASTA (aligned sequences).
    c. Click Continue.
    d. Open downloaded sequence and copy it to your "tree.fasta" file, give it the name *pine*
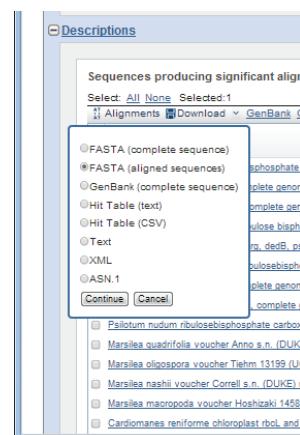
2. Organism: wheat (taxid:4565), optimize for somewhat similar sequences.
   a. After checking your selected sequence (1st option) click Download
   b. Choose "FASTA (aligned sequences).
   c. Click Continue.
   d. Open downloaded sequence and copy it to your "tree.fasta" file, give it the name *wheat*

3. Organism: Soybean (taxid:3847), optimize for somewhat similar sequences.
   a. After checking your selected sequence (1st option) click Download
   b. Choose "FASTA (aligned sequences).
   c. Click Continue.
   d. Open downloaded sequence and copy it to your "tree.fasta" file, give it the name *soybean*
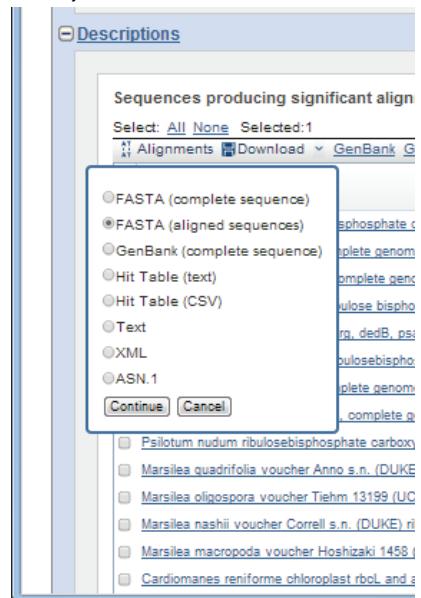
***You now should have rcbL sequences for algae, moss, fern, pine, wheat and soybean in your "tree.fasta" file. **Double check** your work to make sure every organism has the respective label.

SECTION 5

1. Open program **clustalX2** on your computer.
   a. When clustalX2 is open, go to "file", "load sequences"
   b. Find your tree.fasta file.
   c. A portion of your clustalX2 file should look like this:
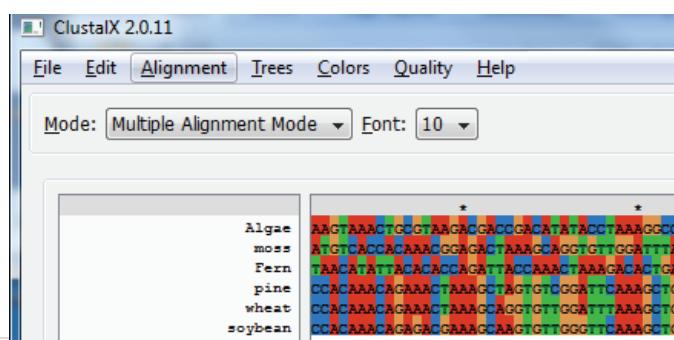
d. Click the "alignment" tab, and click the option "do complete alignment"- a small window will come up, click "OK".
e. After alignment is complete, you can scroll over in the screen and see the mutations among sequences.
f. Go to "file", select "save sequence as", a new window will show up, uncheck the "clustalXformat" and check "FASTA format" and click "OK". *remember where you saved that file too.

   a. We have now created the input file to run in the program MEGA

SECTION 6

Constructing a tree using MEGA5

1. Open MEGA by double-clicking on the program.
2. Click on the "data" tab and "open a file/session"
3. Open the "tree.fasta" file.
    Hit "analyze".
4. Choose "nucleotide sequence", "Ok" and then "Yes" in the next steps, then hit "OK"
    *Your Mega window should look like this:
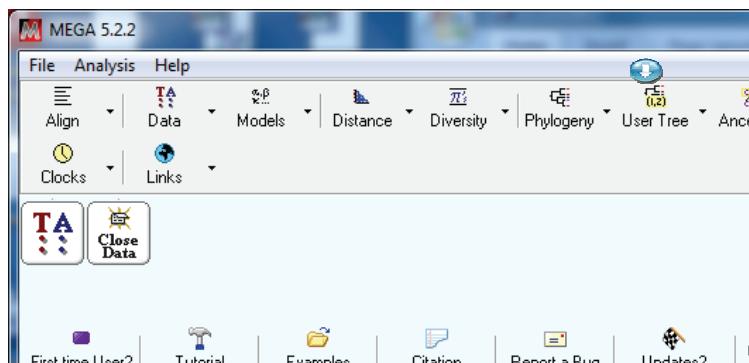
a. Click on the "phylogeny" tab, choose the first option in the dropdown. Click "okay" and finally "compute"
b. After your data has been analyzed a tree will pop up and will look similar to this:
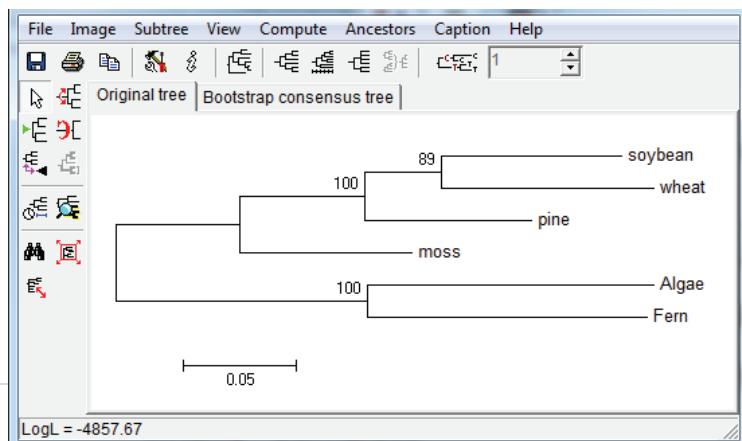
c. Now click algae on your tree and click on the second option on your left sidebar (one with the green arrow). This will root the tree to show evolution of organisms. Your tree will now look like this:

```
                                                      89 ┌────────── soybean
                                             100 ┌───────┤
                                    100 ┌─────────┤        └───────── wheat
                           ┌────────────┤         └──────────── pine
                           │            └────────────── moss
              ┌────────────┤
              │            └──────── Fern
    ──────────┤
              └──────── Algae

         ├───────┤
          0.05
```
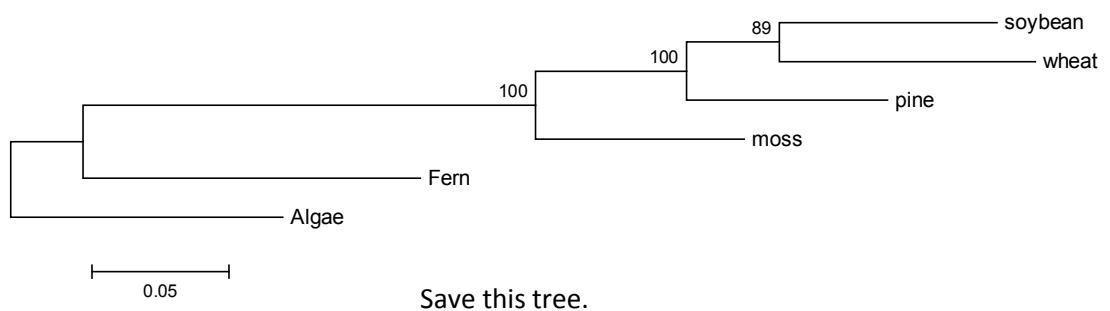
Save this tree.

Now you have a beautiful evolutionary tree as shown below.

Reading tree (Example)
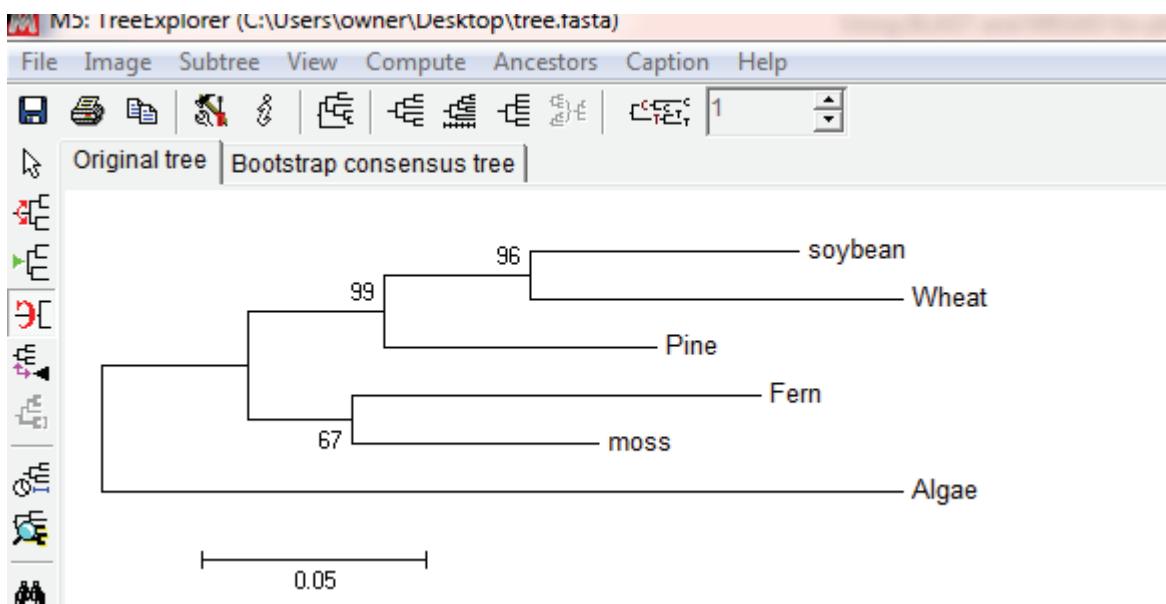
You can play with various options about the cosmetic of the tree. You can save the tree in different format by clicking the second tab "image" on the menu bar or directly print as pdf from the file menu.

How to read the above tree (five of many messages are included)

1. All green plants had ancestor similar to the Alga
2. Fern and Moss form a group (=clade) that is supported by a bootstrap support of 67. This means out of 100 replications, 67% of the time they are together. They are more closely relagted each other than to other plants sampled in the present data set.
3. Pine is ancestral to both wheat and soybean (BS = 99).
4. Soybean and wheat are the most recent plants among the six sampled. They are more closely related to each other (BS=96) than any other plants sampled here.
5. Soybean, wheat and pine form a clade (a group of ancestror and all of its descendants) and fern and moss form another clade. All five together form a bigger clade. The inner lines represent ancestor for respective members, and those outer lines on the right are the descendants.