MEGA tutorial on Linux

Answer the questions in red to make sure you truly understand what you are doing. Feel free to ask us questions during the workshop!

1. Clone the directory from GitHub. Cd into the directory where the fasta sequences are.
2. Concatenate all the .fasta files, and add a new line in between each file. Use grep to ensure the right number of total sequences in the merged file.

```
[patricia@biol-dwalsh-cluster Scenario1_final]$ for f in *.fasta; do (cat "${f}"; echo) >> Scenario1_merged.fasta; done
[patricia@biol-dwalsh-cluster Scenario1_final]$ grep -c '>' Scenario1_merged.fasta
96
```

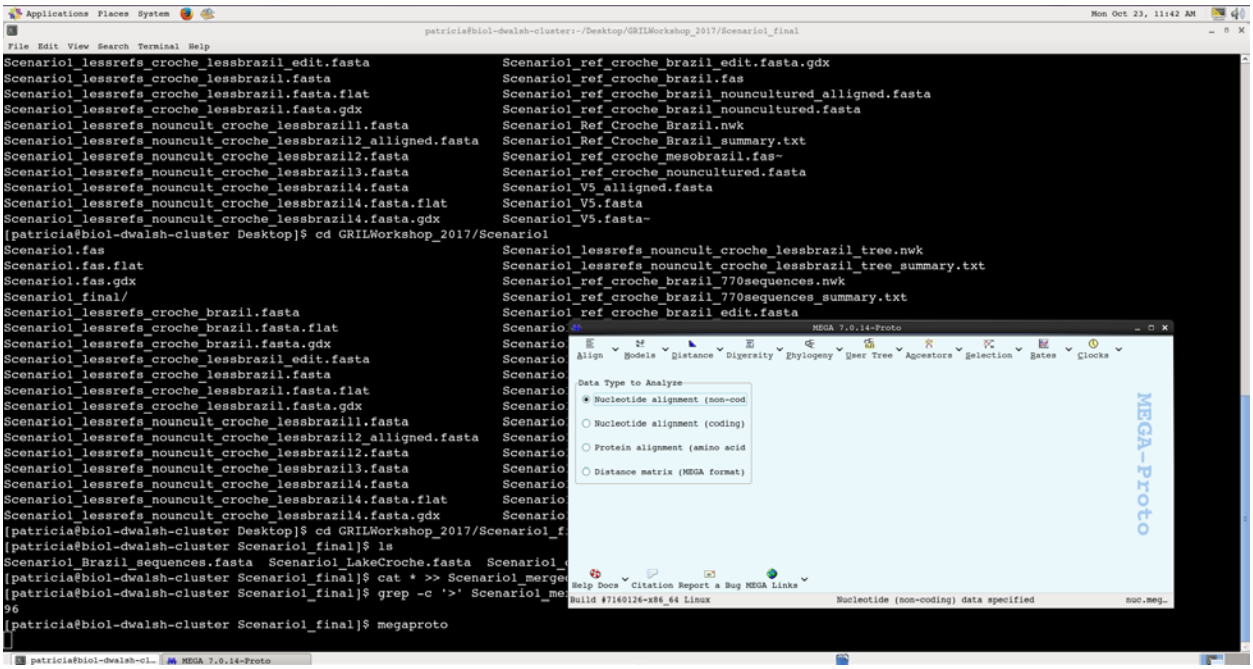
What does the grep function do?

3. Use **Muscle** to align the sequences. –in for the input file, -out for the output file. You can specify max memory used if necessary (if you have thousands of sequences for example)

```
[patricia@biol-dwalsh-cluster Scenario1_final]$ muscle -in Scenario1_merged.fasta -out Scenario1_merged_alligned.fasta

MUSCLE v3.8.31 by Robert C. Edgar

http://www.drive5.com/muscle
This software is donated to the public domain.
Please cite: Edgar, R.C. Nucleic Acids Res 32(5), 1792-97.

Scenario1_merged 96 seqs, max length 2528, avg  length 1036
00:00:00     22 MB(5%)  Iter   1  100.00%  K-mer dist pass 1
00:00:00     22 MB(5%)  Iter   1  100.00%  K-mer dist pass 2
00:00:04    116 MB(28%)  Iter   1  100.00%  Align node
00:00:04    116 MB(28%)  Iter   1  100.00%  Root alignment
00:00:04    117 MB(28%)  Iter   2   10.64%  Refine tree
```

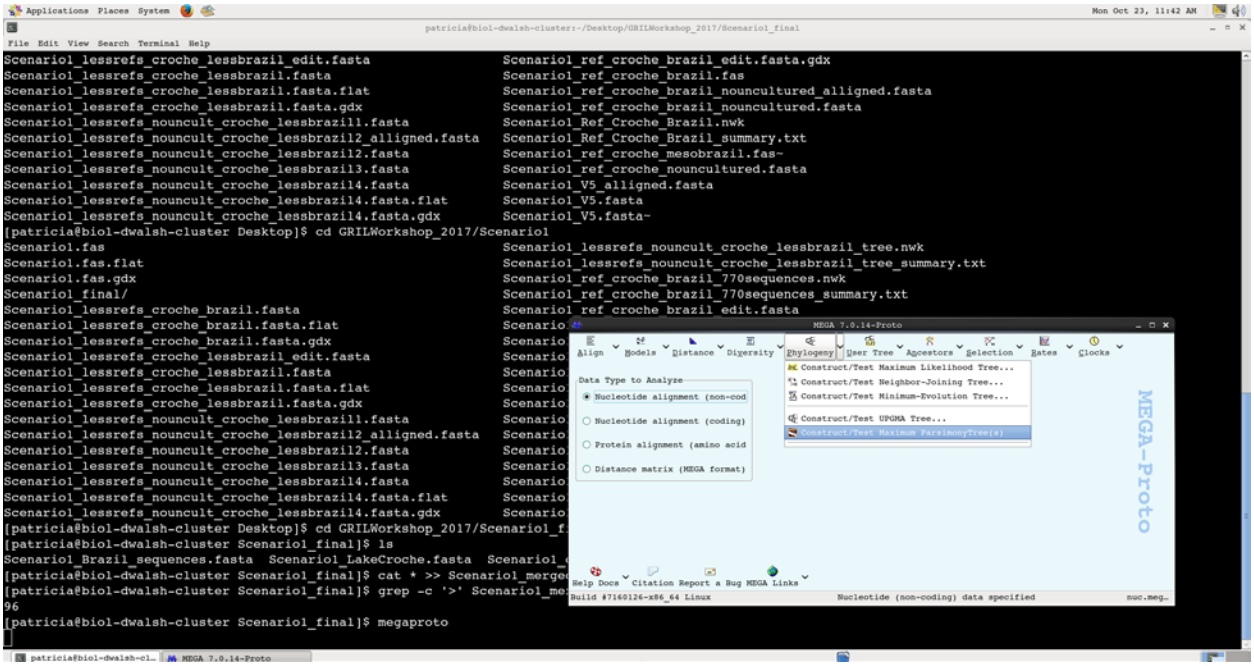
How many sequences did you give it?
What is the average length of the sequences?
What is the name of the output file?

4. Now set the parameters to make the tree. Use **megaproto**. A pop-up window will open. Select "Nucleotide alignment (non-coding)

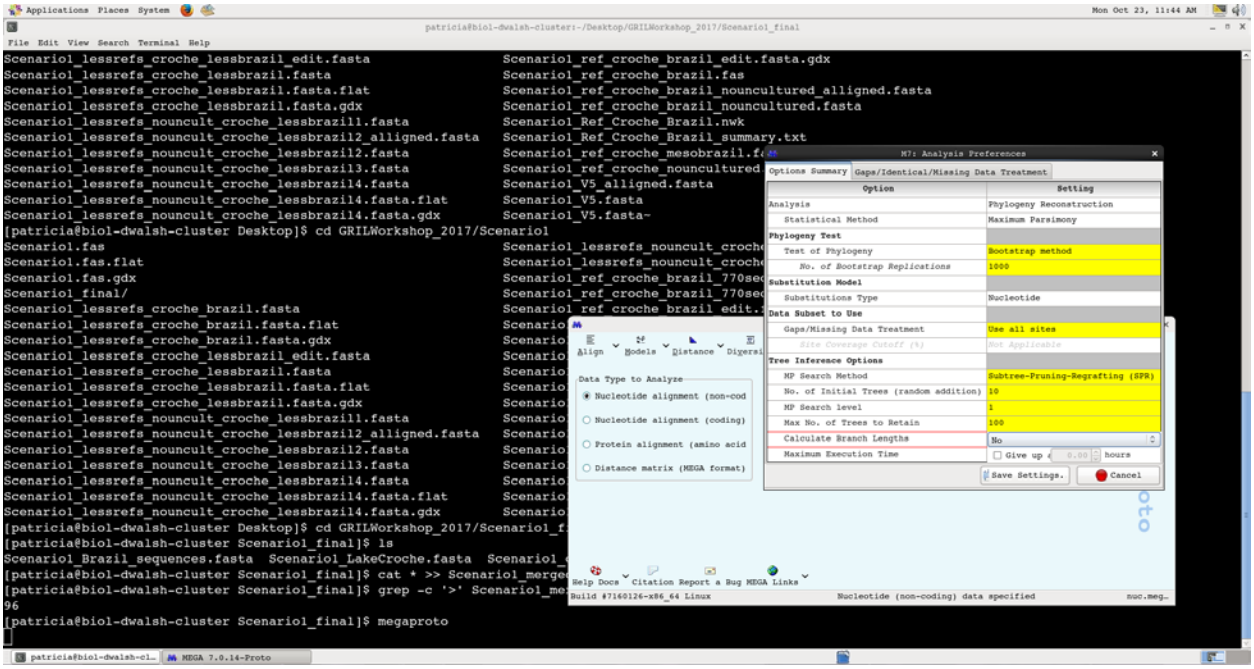


Can we make phylogenies using proteins? If so, which option would we select?

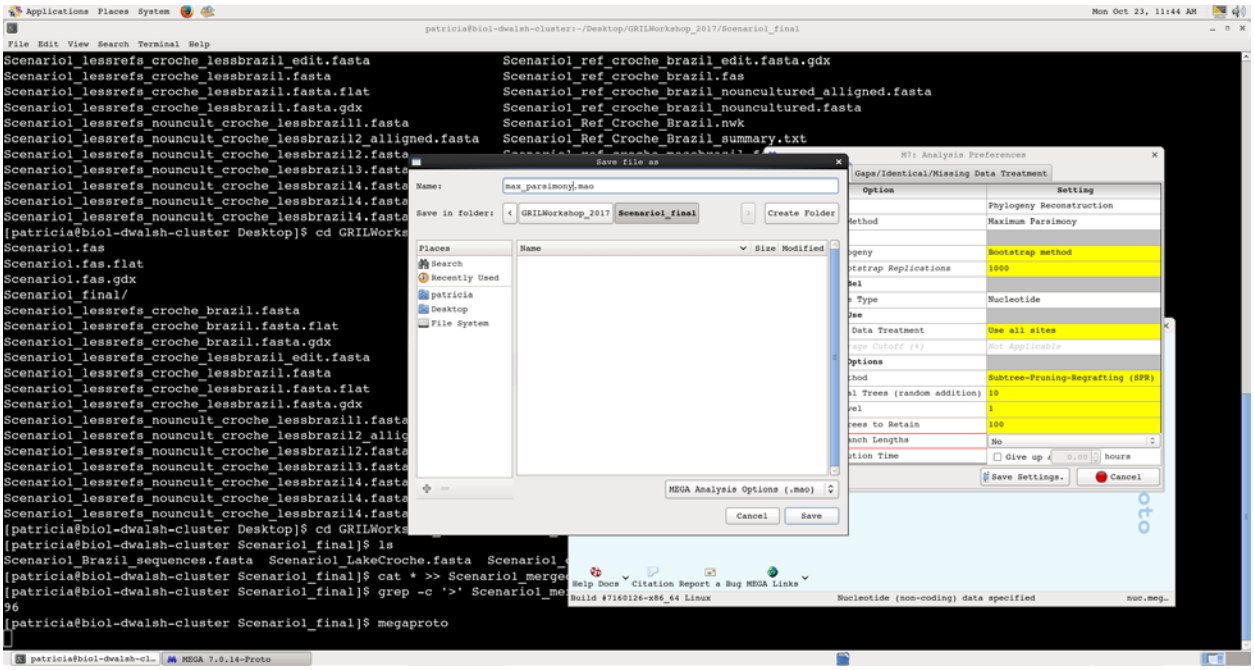
Then, click on the Phylogeny drop-down menu, then select the Maximum Parsimony option.



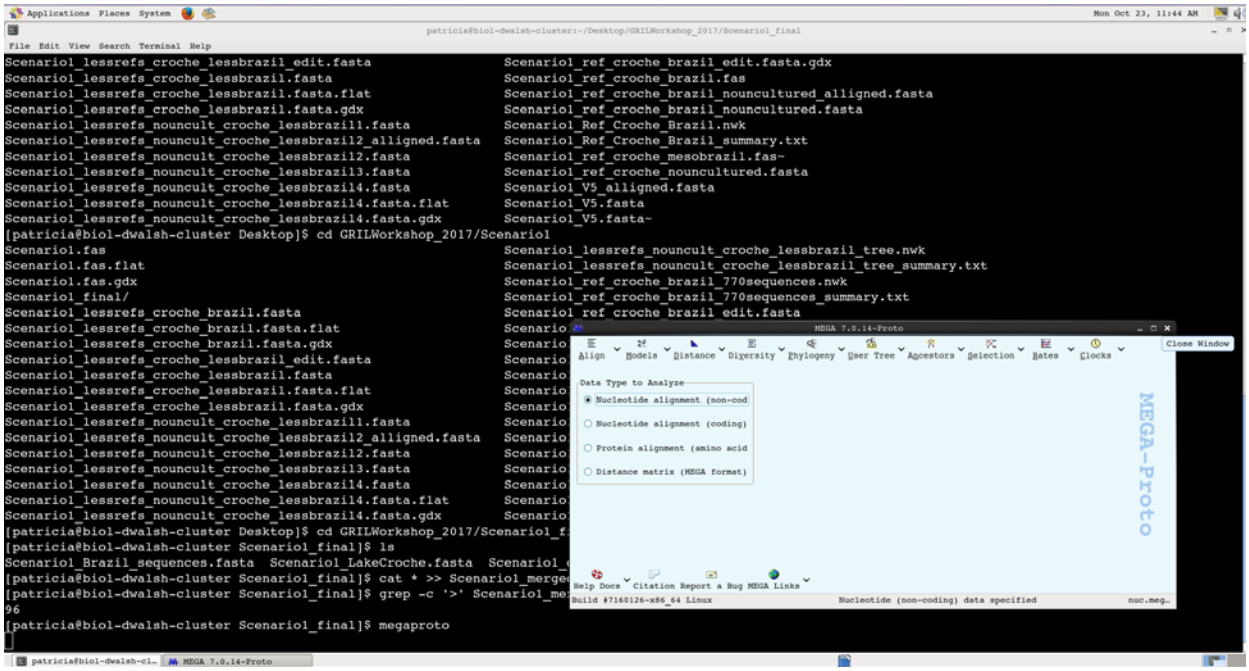
Use the following settings.



Click “Save Setting” and rename it. Save as a .mao (MEGA alignment option) file.



Close the popup window.

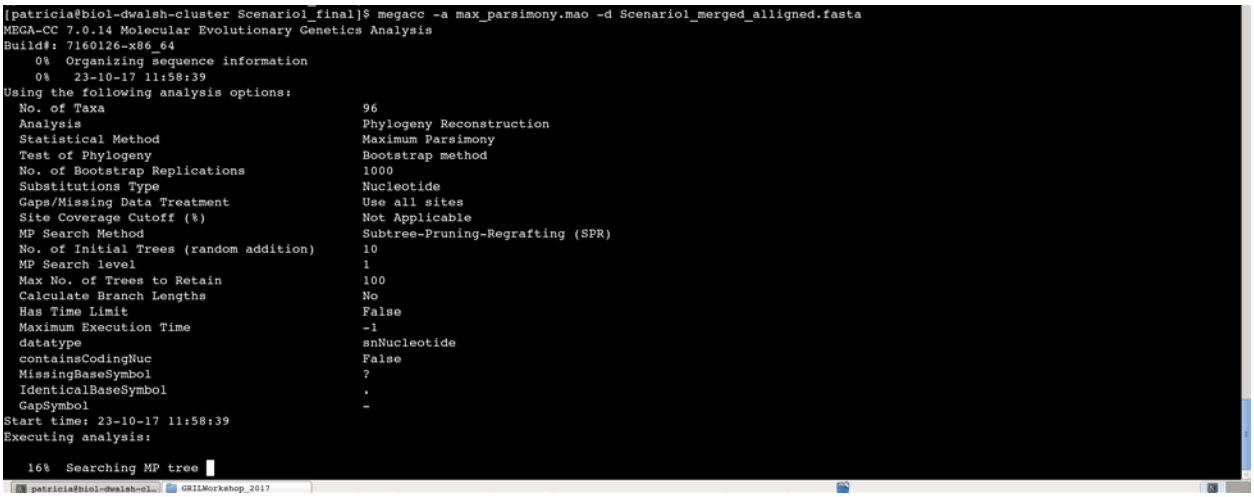


5. Check the contents of your folder using ls –lh

```
[patricia@biol-dwalsh-cluster Scenario1_final]$ ls -lh
total 500K
-rw-rw-r-- 1 patricia patricia 2.6K Oct 23 11:44 max_parsimony.mao
-rw-rw-r-- 1 patricia patricia 5.4K Oct 23 11:40 Scenario1_Brazil_sequences.fasta
-rw-rw-r-- 1 patricia patricia  47K Oct 23 11:40 Scenario1_LakeCroche.fasta
-rw-rw-r-- 1 patricia patricia 252K Oct 23 11:49 Scenario1_merged_alligned.fasta
-rw-rw-r-- 1 patricia patricia 118K Oct 23 11:47 Scenario1_merged.fasta
-rw-rw-r-- 1 patricia patricia 3.0K Oct 23 11:40 Scenario1_outgroup.fasta
-rw-rw-r-- 1 patricia patricia  63K Oct 23 11:40 Scenario1_References_seqs.fasta
```

What does the –lh option mean?

6. To make the tree, check the options for **megacc** (mega compute core). Type "megacc –h" in the terminal.
7. Run the tree by providing the .mao file (-a option), the data (-d option). It's important to use the "aligned" fasta sequences. The steps and percentage values describing progress will be shown on the Terminal. Wait.



What does bootstrap mean?

Why would we use 100? 1000? None?

8. When the tree is finished check the contents of the directory again. Can you find the new files created?

   What are the file formats of the files created?

   If someone asked you how you made your tree, which file could you give them?
9. The Newick tree (.nwk) has been created, you can open it using FigTree and follow the steps described in Part II of the tutorial.