Prior to the CQUEST lab:

1.
a) Go to www.cquest.utoronto.ca to read about CQUEST and obtain a login ID and set up your password. Test to make sure this works before the lab. We will be working in room RW211.
b) If you already have a CQUEST login, I still recommend testing it before the lab.

2. If the programs and analyses listed below are unfamiliar to you, it would be wise to briefly review these prior to the CQUEST lab. This will allow us to more efficiently assist you during class time.

Phylogenetic analyses – Summary of data flow:

1. Obtain your DNA sequence(s) in Fasta format from the course web site: http://courses.eeb.utoronto.ca/eeb331/

2. Text editor: paste your DNA sequence and save as plain text .txt file type

3. GenBank: identify similar sequences and an outgroup sequence from BLAST searches (See below under "During the Cquest lab")

4. Text editor: append sequences from GenBank to your .txt file from above

5. Open your file in BioEdit (from "Botany" folder in Start menu), go to "Accesory applications", run "ClustalW multiple alignment"

6. Use BioEdit to manually optimize your automatically generated sequence alignment. Double click on the sequence names and rename them all with unique, short names (10 characters) such as the GenBank accession numbers. Save your final alignment as a "Phylip 4" format file. Drag this file to the directory Phylip/exes and rename it "infile" (lower case, no file extension)

7. Open the DNApars module from within Phylip/exes to perform a phylogenetic analysis of your sequences (use settings provided by JM). This will automatically generate two output files: output and outtree within the same folder

8. Open the outtree file in a text editor and add the genus and species names beside each of the GenBank accession numbers (no spaces, separate words with underscores "_")

9. Open your outtree file using FigTree to view the tree

During the CQUEST lab:

Create a FASTA file:

In a text editor, create two FASTA files, one for your ITS and one for your LSU sequences.

Fasta format:

>Name1 {type return}
Sequence… {type return}
>Name2 {type return}
Sequence… {type return}
Etc.

Example of FASTA format:

>Craterellus_tubaeformis
ACTG…
>AY1234567 (this is the unique GenBank Accession number)
ACTG
>AY2345678
ACTG

...etc

We will change all the names to unique, 10 characters entries in BioEdit.

Save as Plain Text format
(ex. ITS_Fasta.txt)

Use GenBank nBLAST tool to obtain similar sequences from the database:

Go to http://blast.ncbi.nlm.nih.gov/Blast.cgi
- Select "nucleotide blast" from BASIC BLAST;
- Copy/paste your seq in Fasta format in the "Enter Query Sequence" window;
- Select "others" (nr/nt) for the database to use (under "Choose Search Set");
- Program Selection: choose " Highly similar sequences (megablast)"
- Click BLAST

Scroll past the picture and briefly review your list of top hits.

> *A good bit score will be approximately twice the value of the length of your sequence.*
> *(ex. ITS sequence 600bp, ideal bit score = ~1200bp)*
>
> *A good e-value will be 0.0*
> *(this is the probability of finding this sequence in the database as a match to yours, by chance alone)*

Select the first 20 alignments (or less) with good bit scores and e-values.

> *Ideally, you would like to have about 20 top hits, in addition to your own sequence, in the FASTA file, but…*
>
> *Skip sequences that are "not identified"/"uncultured"*
>
> *Skip sequences with an e-value less than e-80*
>
> *If you are still unsure, ask Jean-Marc or TA for assistance.*

Also select an outgroup sequence.

> *Pick one from the bottom of your alignment list, it should be a sequence closely related to your ingroup sequences but outside the group. Ask Jean-Marc or TA for assistance if you are unsure what to select.*

Scroll back to top of alignments

Click on button "Get selected sequences"

Click on "CoreNucleotide records"

Change Display option to FASTA from the drop-down list

Change Show option to a number equal to or greater than the number of sequences you selected (ex. 50)

Change "Send to" option to "text" from the drop-down list

Copy these results and append/paste to your FASTA file from above. Save this file.

> *Edit the name for each of the downloaded sequences to show the Accession number followed by the Genus and species, as follows: AY566725_Amanita_muscaria (no space, hyphen etc…)*

Use BioEdit to manually optimize your alignment:

If you wish to continue this work on your own PC (does not work on OSX for mac)*, you may download this program and documentation for free from:* http://www.mbio.ncsu.edu/BioEdit/bioedit.html
If you use a Mac, you should use Se-Al instead of BioEdit. Download Se-Al from http://tree.bio.ed.ac.uk/software/seal/

Open BioEdit

File → Open the alignment file in Fasta format

> *Be sure to select the appropriate editing "Mode:" there are 3 options available.*
> *In the Edit mode, you can highlight a column, and hit "delete" to remove a column*

Use the "edit", "select/slide", or "grab/drag" modes to continue editing. Press "ctrl+z" to undo a change.

After aligning, trim the start and end of your alignment so you have a single rectangular block of aligned sequences.

> *Tips:*
> *For ITS: sequences usually begin with the motif CATTA*
> *sequences usually end with the motif GACCT*
> *Basidiomycete sequences are usually ~600 bp*
> *For LSU: sequences usually begins with the motif TTCCCT*
> *Sequences are usually ~900-1000 bp*

Save this file as "ITS_trimmed.phy (Phylip 4 file format)."

Delete ambiguously aligned regions if necessary.

> *Common problem:*
> *Only the first 10 characters of the sequence name will be carried over to Phylip in the next step so make sure they are unique, ex. GenBank Accession number. Double-click on the sequence name in BioEdit to open editing window.*

Save this file with a new name, ex. "ITS_edited.phy (Phylip 4 file format)."

Use Phylip to conduct a phylogenetic analysis:

If you wish to continue this work on your own PC, you may download this program and documentation for free from: http://evolution.genetics.washington.edu/phylip.html

Open the Phylip folder.

Move your edited BioEdit alignment file (.phy) into the Phylip folder. Rename it “infile” (lowercase, no file extension). Rename the “font1” file as “fontfile”, this is a bug in the program.

Open DNAPars and follow instructions. This module will create a file called “outtree.”

Rename “outtree” to something else so that it is not overwritten the next time you do this.

Open your renamed “outtree” file in a text editor (ex. Right click, open with, Word).

Edit the tree definition file by adding the Genus and species names to the appropriate accession numbers in your original fasta file.

For example,

Original tree definition:
(Accession4 : 20, (Accession3 : 10, (Accession1: 5, Accession2: 10)))

change to:
(Accession4_Genus_species : 20, (Accession3_Genus_species : 10, (Accession1_Genus_species: 5, Accession2_Genus_species: 10)))

This can be done relatively easily using the ctrl+h “search and replace” function.

Save as Plain Text file type.

Open your file in FigTree.
Follow instructions.
Save/print tree.

An alternative program to FigTree (freeware, http://tree.bio.ed.ac.uk/software/figtree/) is called TreeView and can be downloaded for free from http://taxonomy.zoology.gla.ac.uk/rod/treeview.html . In this program, you can select your outgroup taxon (Tree→Define outgroup), and root the tree using the outgroup method (Tree→Root with outgroup).

*Multiple most parsimonious trees may be saved. For this project, you only need to consider one of these.

Tip:
Before you logout of CQUEST, do not leave files you want to keep on the desktop, move them onto the shared drive into “My Documents” folder instead.

After the CQUEST lab:

If needed, it is possible to complete these analyses on your own time either at the CQUEST lab or on your own PC if you download the free programs.

You will need to send Jean-Marc your alignment files (“trimmed” and “edited”) for his review by email, along with a hard copy of your Lab Report 1.