

Making Phylogenetic Trees using a Gene Bank

Muggle's Magic

In this activity, you will create a phylogenetic tree that will show relatedness based on gene composition of the most common protein found in plants Rubisco, also known as Ribulose 1,5 Biphosphate. This protein is an enzyme that helps plants fix carbon dioxide from the air in its effort to make glucose in the light independent reactions of photosynthesis. Yes, even magical plants need to make their own food. Some plants are carnivorous like the venomous tentacula or venus flytrap, but they still do photosynthesis to get a majority of their energy.

Follow this guide to make your tree, but the analysis component is up to you.

1. Go to National Center for Biotechnical Information's website <http://www.ncbi.nlm.nih.gov/>
2. You will need to use the search bar at the top of the page but apply a filter before you look for the information.
3. Change your filter to Nucleotide (since we are looking for gene information on a specific protein).



4. Type in the species that you are looking for and then RBCL after the species name. RBCL stands for Ribulose 1,5 Biphosphate Chloroplast Large subunit. (If your species is not found just use the genus name)
5. You should open up a word document or a notepad file so you can copy (ctrl+C) and paste (ctrl+V) the information you will need from all the species you want to compare. For example, look up 'Delphinium tricorne RBCL'. This would provide you with sequence for larkspur. The other plants you will include in your tree are:
 - a. Vallisneria
 - b. Allium
 - c. Toxicodendron
 - d. Sagittaria
 - e. Tsuga
 - f. Aconitrum
 - g. Prunus
 - h. Ulmus
 - i. Delphinium



You want to use the hyperlink FASTA to get the gene sequence on which our phylogenetic tree will be built.

7. Copy this code exactly to that open word or notepad document

Asclepias tuberosa voucher NMNH Kress 06-8174 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcl) gene, partial cds; chloroplast

GenBank: GQ248553.1

[GenBank](#) [Graphics](#)

```
>gi|240253033|gb|GQ248553.1| Asclepias tuberosa voucher NMNH Kress 06-8174 ribulose-1,5-  
bisphosphate carboxylase/oxygenase large subunit (rbcl) gene, partial cds; chloroplast  
AAGTGTGGATTCAAAGCCGGTGTAAAGAGTACAAATTGACTTATTATACTCCTGAATACGAAACAAAA  
GATACTGATATCTTGGCAGCATTCCGAGTAACCTCTCAACCCGGAGTTCACCTGAAGAAGCAGGGGCCG  
CGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACCTGTTGGACCGATGGACTTACTAGCCTTGATCG  
TTACAAAGGGCGATGCTACCACATCGAGGCCGTTCTCTGGAGAAGAAGATCAATTTATTGCTTATGTAGCT  
TACCTTTAGACCTTTTGAAGAAGGTTCTGTACTAACATGCTTACTTCCATTGAGGTAATGATTTG  
GGTTCAAAGCCCTACGCGCTCTACGCTGGAAGATTGCGAATCCCTCAGGCTTATATAAAACCTCCA  
AGGCCCGCCGATGGCATCCAGTTGAGAGAGATAAATTGAACAAATATGGTCGCTCTGTGGGATGT  
ACTATTAACCAAATTTGGGTTTATCAGCTAAAACTATGGTAGGGCGTTTATGAATGCTCT
```

8. Repeat this procedure for all species that you are going to be working on. If you can't find your exact species, then please just use the genus name. REMEMBER: this is a muggle database and since the knowledge of magical plants is kept secret, you will only be able to locate sequences for the muggle species.
9. **You must also add a gene sequence for an outgroup (a known species that most likely does not share a recent common ancestor with the other species you are analyzing). For your outgroup use 'Tsuga rbcl'. Tsuga are commonly known as hemlock trees.**
10. Save your file of all the FASTA codes of your species you will need to enter it into the next website.
11. Go to the phylogenetic tree maker website <http://phylogeny.lirmm.fr/phylo.cgi/index.cgi>

Phylogeny.fr is a free, simple to use web service dedicated to enhancing and analyzing phylogenetic relationships between molecular sequences. Phylogeny.fr uses and connects various bioinformatics programs to reconstruct a robust phylogenetic tree from a set of sequences.

If you use this site, please cite:
Drepper A., Gagnon C., Blain G., Audin S., Buffet S., Chevret F., Dufourd J.F., Guiselin S., Lohet C., Lortal M., Claverie J.M., Guisard D. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 2008 Jul 1;36(Web Server issue):W459-464. Epub 2008 Apr 29. (PubMed) - just for citation

Phylogeny analysis

- One Click
Paste your set of sequences and let the software make decisions on your behalf (Each step is optimized for your data)
- Advanced
Modify all parameters for the various steps
- A la Carte
Choose your own phylogeny workflow using more programs available

12. Click on the "One click" option.

13. Copy your code from your saved file (all the species) and paste it into the window.

phylogeny.lirmm.fr/phylo_cg/simple_phylogeny.cgi



Home | Phylogeny Analysis | Blast: Sequence Explorer | Online Programs | Your Workspace | Documentation | Downloads | Contacts

"One Click" Mode

Alignment MUSCLE → Curation Gblocks → Phylogeny PhyML → Tree Rendering TreeDyn

1. Overview | 2. Data & Settings

Name of the analysis (optional):

Upload your set of sequences in FASTA, EMBL or NEXUS format from a file:
 No file chosen

Or paste it here (load example of sequences)

Maximum number of sequences is 200 for proteins and 200 for nucleic acids.
 Maximum length of sequences is 2000 for proteins and 6000 for nucleic acids.

▶ [Names association](#)

Use the Gblocks program to eliminate poorly aligned positions and divergent regions

To receive the results by e-mail, enter your address(es):

Do not attach result files

Note: beside sequences count and average length limit for the alignment stage there is also a limitation on the phylogeny stage (sequences_count*sequences_count*aligned_sequence_length=100000)

14. It is highly suggested that you put in an email to get your results. Sometimes the website gets bogged down if it is running a lot of sequences projects from other people at the same time. They will email you a link to your results when they are complete.

15. Your results will look something like this:

"One Click" Mode

Alignment MUSCLE → Curation Gblocks → Phylogeny PhyML → Tree Rendering TreeDyn

Tree Rendering results



gi_818945860_gb_KJ773601.1_Monarda_punctata_ribulose-1_5-biophos
 gi_818945952_gb_KJ773736.1_Phlox_drummondii_ribulose-1_5-biophos
 gi_34765369_gb_AY215124.1_Helianthus_annuus_ribulose-1_5-biophos
 gi_345699021_gb_KF613105.1_Echinacea_purpurea_isolate_HIMS2205-1

0.01

Dynamic Tree Editor

Color Size Label Branch Root to original tree Root to new add node setting Root (outgroup) Tree orientation: B
 Branch Flip colors Swap colors Change leaf name
 Add length for group nodes Add annotations using color

Display:
 Branch annotations Branch support values Branch length values Use color Leaf
 Legend of groups ON OFF Hidden Use color Branches

Show branch length
 Label font:

Tree construction: NeighborJoin Build

Image size: Small Medium Large Full size

16. You have lots of options on this page but here are some explanations for what you will see.
- a. You can make your tree bigger if it looks too small.
 - b. You can download it as a pdf, png, or svg file.
 - c. You can rearrange the branches of the tree if it looks odd. For example, it is highly suggested you REROOT your tree around Tsuga (your outgroup). There is a button to do this below your finished tree.

What do the numbers on the tree mean?

- If you look at the bottom of the sample tree there is a line and a number below (0.03). This number represents the scale of change. This is genetic change between the species that you are analyzing based on the number of substitutions per site. The number of changes divided by the length of the sequence. (This only applies to the horizontal axis. The vertical axis is irrelevant for scaling.)

- If you look on the tree branches there is a red 0.998. This number refers to the statistical strength of the node you are observing. 0 is the lowest support and 1.0 is the highest support.

Nice Job, Wizard! Print or copy/paste your tree for your teacher to see your evidence of your hard work!

