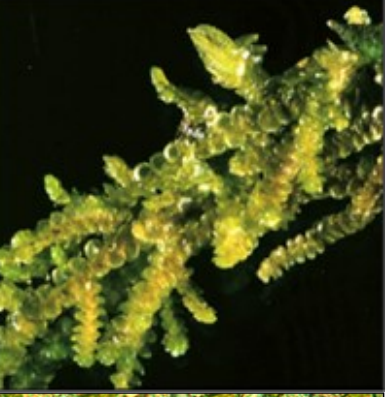
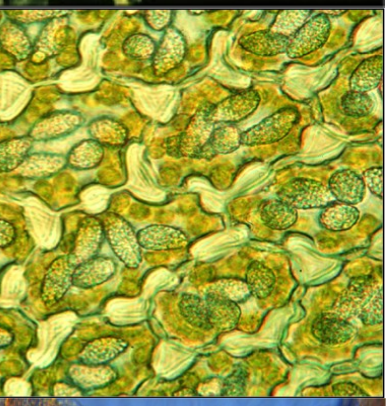
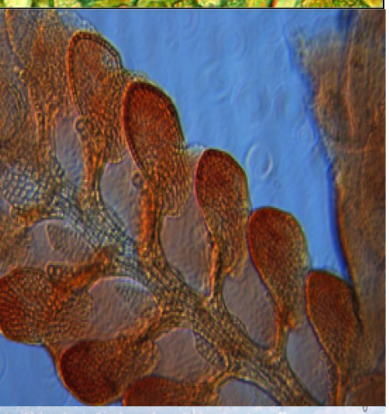
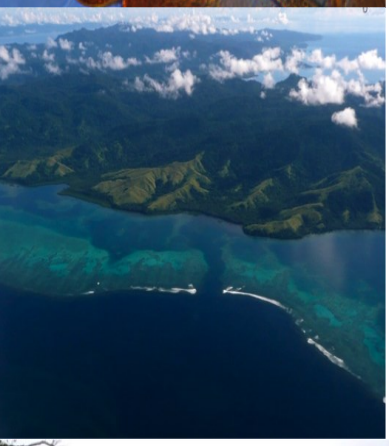








Phylogeny Activity

MicroPlants.fieldmuseum.org

A Citizen Science Project

Collection Connection

MicroPlants

microplants.fieldmuseum.org

The Field Museum

ZOONIVERSE

Creating a Phylogenetic Tree

A phylogenetic tree depicts the evolutionary relationships within a group of organisms. The tree structure is also known as a phylogeny or a cladogram. Each branch represents a lineage, the nodes represent the last common ancestor, and the tips of each branch represent groups of descendant taxa. Two groups that split from the same node are referred to as sister groups because they are each other's closest relatives (Figure 1). Phylogenetic trees usually include an out-group taxon that is least closely related to the rest of the descendent taxa, which can be useful in understanding better the relationships between several groups of organisms over a long period of time (Understanding Evolution, 2014).

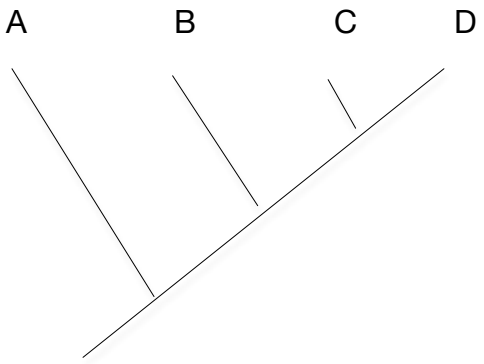


Figure 1 A phylogeny showing the evolutionary relationship between species A, B, C, and D. Species C and D are referred to as sister taxa because they evolved more recently in comparison to species A and B.

Biologists will typically refer to these groups as clades, which are groups of organisms that include the last common ancestor and all the descendants of that ancestor. Phylogenies are a useful tool for organizing information about biological diversity, structuring classifications, and providing insight into evolutionary events (Baum, 2008). Evolutionary trees can be drawn in several ways and still depict the same information. The length of each branch is irrelevant. Trees are chosen by how neatly and clearly they represent information. Phylogenetic mapping is a useful skill to develop for aspiring biologists in order to better understand modern evolutionary theory and the relationships between different organisms.

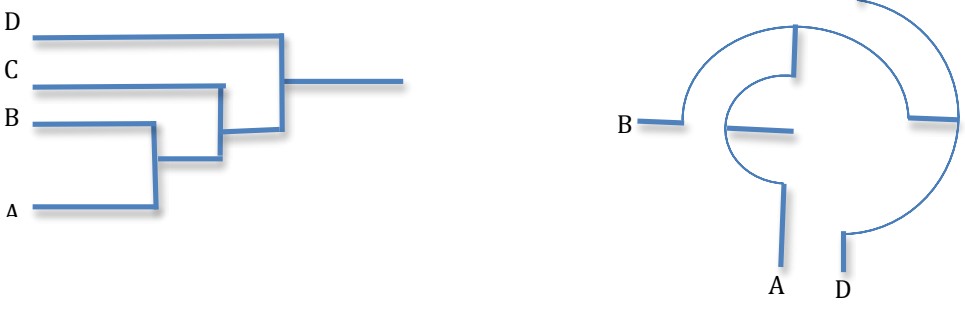


Figure 1 Different types of phylogenetic trees showing equivalent relationships.

Researchers are conducting a taxonomic treatment of a diverse group of early land plants in the liverwort genus Frullania, which is a taxonomically diverse genus that includes over two thousand published names. This taxonomic research includes documenting, describing, and discovering new species to science. Liverworts (Marchantiophyta) are pivotal in our understanding of early land plant evolution and exist as important components of the vegetation in many regions of the world. Because of their small size, liverworts cannot withstand changes in the environment as well as large plants, such as trees; therefore they are good environmental indicators of climate change. Bryophytes can acquire great biomass under the canopy of forests, help regulate soil moisture, sequester nutrients, and provide a habitat for small invertebrates and fungi. Taxonomic conclusions will be drawn from a multi-faceted data set by, including studying plant morphology, conducting fieldwork and, experimental growth studies, as well as DNA sequence data, and population studies using DNA microsatellite markers. One way we can run analyses on DNA sequences is by understanding genetic codingBy comparing genetic sequences from different species we can further understand how different species of Frullania have evolved. Below is an example of how scientists classify the genus Frullania.

- Kingdom: *Plantae*
- Phylum: *Marchantiophyta*
- Class: *Jungermanniopsida*
- Order: *Jungermanniales*
- Family: *Jubulaceae*
- Genus: *Frullania*
- Species: *pycnantha*

Frullania pycnantha

You will create a phylogenetic tree that compares multiple species of *Frullania*. Follow the directions below to perform a nucleotide search for each isolate using Genbank. You will then use these sequences to build a phylogenic tree of known species on MABL. After creating your phylogenetic tree, your instructor will provide you with a set of unknown sequences for comparison. You will add the unknowns to the saved set of sequences and generate a second tree for comparison and analysis.

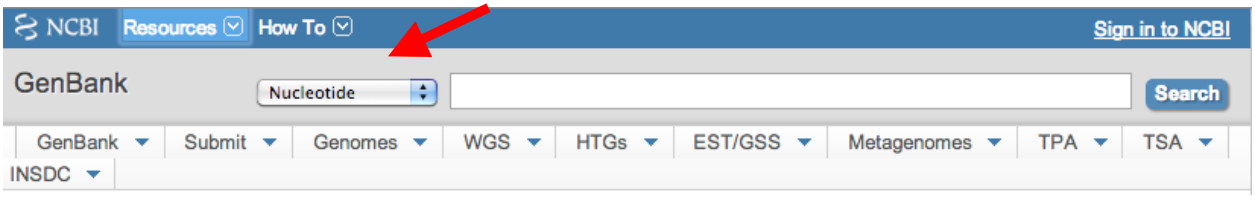
To run this lab you will need:

GenBank: http://www.ncbi.nlm.nih.gov/genbank
MABL: http://www.phylogeny.fr/version2_cgi/phylogeny.cgi

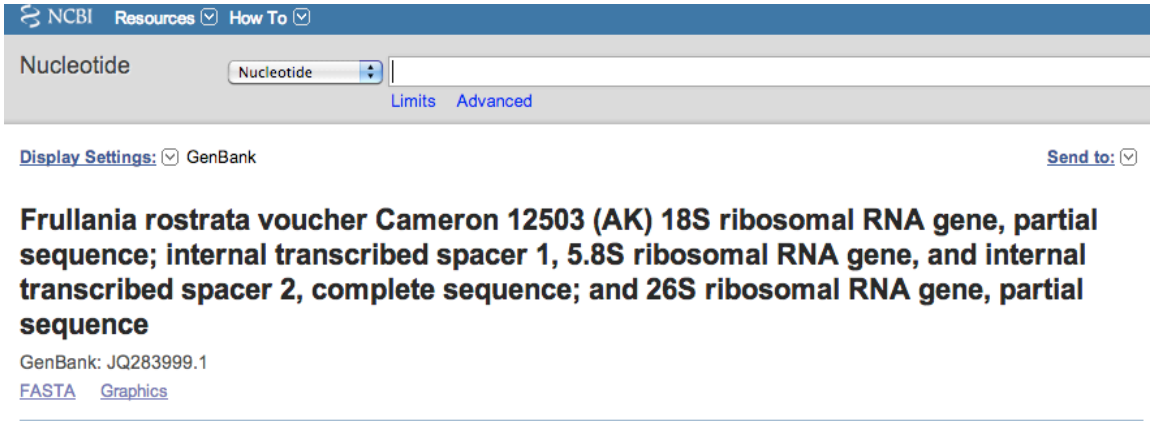
GenBank is a database created by the National Institute of Health (NIH) that holds a reserve of DNA sequences used by the public and medical professionals alike. MABL stands for **M**ethods and **A**lgorithms for **B**ioinformatics designed by **L**aboratoire Informatique Robotique Microelectrnique Montpellier. These websites were designed for the purpose of providing a free and simple web tool that can analyze phylogenetic relationships.

1. Go to GenBank: http://www.ncbi.nlm.nih.gov/genbank

2. Make sure that the search bar is set for **Nucleotide**.

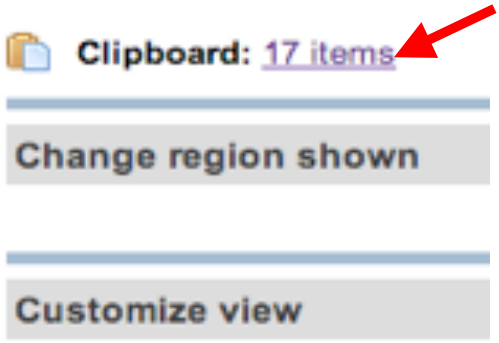


3. Type in or copy and paste each accession number from Table 1 into the search box making sure to include a comma and a space after each accession number and hit enter. An accession number is the identifier of a specific record. Whenever something is catalogued it is given a number so that it can be easily accessed. Hence the name, accession number. After the search is complete you will see the species accession information. This information describes the organism and provides references.



The figure above lists useful information about the catalogued species provided by the accession number. This unique identifier is given to a DNA or protein sequence record, which allows researchers to track different versions of a particular sequence record and the associated sequencs over time in a single data repository. This particular gene sequence can better identify poorly described, rarely isolated, or phenotypically aberrant strains and can lead to the recognition of unknown species.

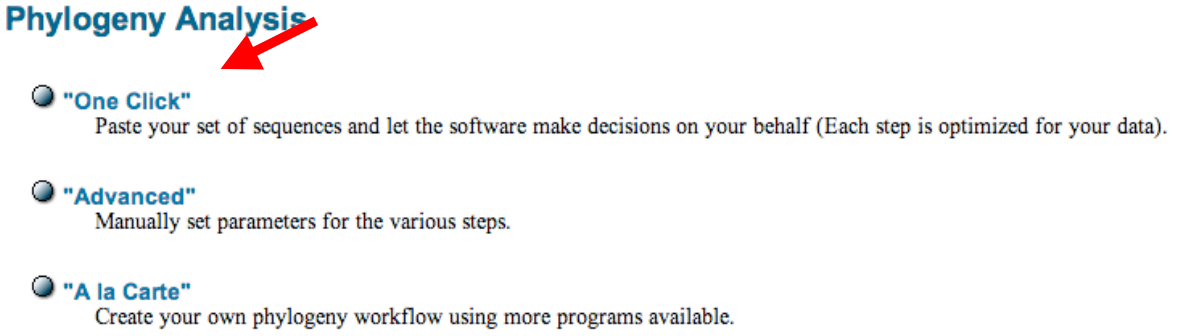
4. Next, **select** all of the sequences and click on **Send to** (upper right corner of the screen), select **Clipboard,** and click on **Add to Clipboard**. You will use these sequences to generate your phylogenetic tree.



5. Next to the clipboard icon you should see that 17 items have been saved. Click on Clipboard, select all 17 Accession numbers→ Send to→ File → Format =00 vq FASTA → Click on Create File. The file will be saved to your computer as a fasta file. **DO NOT EXIT OUT OF GENBANK.**

6. Now that you have saved all of your sequences it is time to create your phylogenetic tree! **OPEN MABL IN A NEW TAB OR NEW WINDOW.** http://www.phylogeny.fr/version2_cgi/phylogeny.cgi

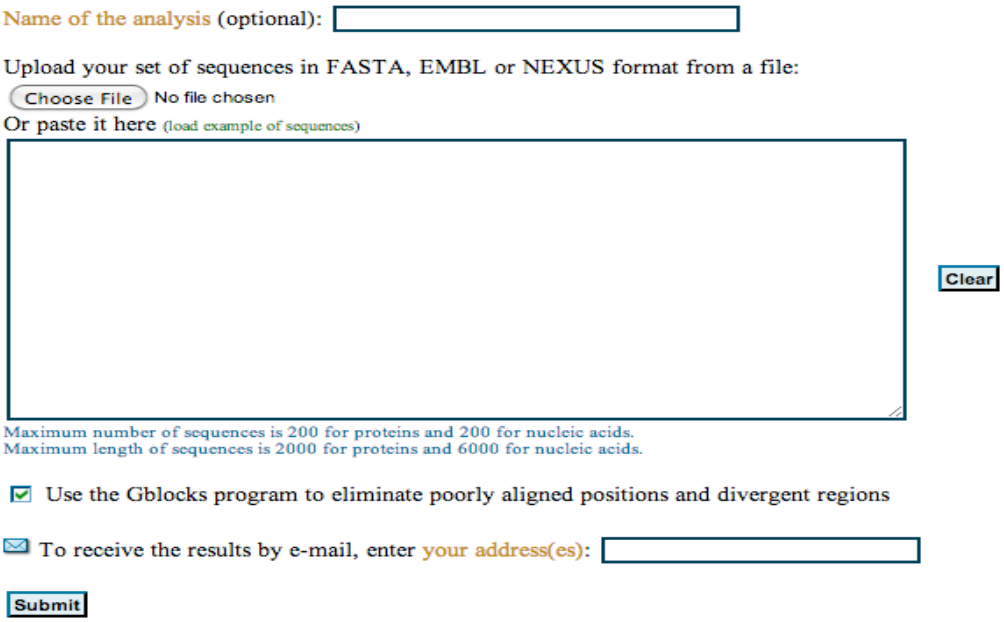
7. Make sure you are on **One Click Mode**. If you do not see the words One Click Mode at the top of the page, please select One Click from the Phylogeny Analysis tab and click on One Click. This will allow you to upload your file of sequences and generate your phylogenetic tree.



8. Name your analysis “Frullania Phylogenetic Tree.”

9. Upload your **FASTA** file from your downloads folder and select **Submit** (your file should be saved as **sequence(1).fasta)**. A FASTA file is a sequence record. A **FASTA** format consists of a single-line description (sequence name), followed by line(s) of sequence data. The first character of the description line is a greater-than (">") symbol.

10. Your sequence will align and produce a phylogenetic tree that you can download and analyze.



For analysis purposes, you are going to construct a second phylogenetic tree that includes the unknowns A, B, and C. You will study the relationship between your unknowns and the *Frullania* species.

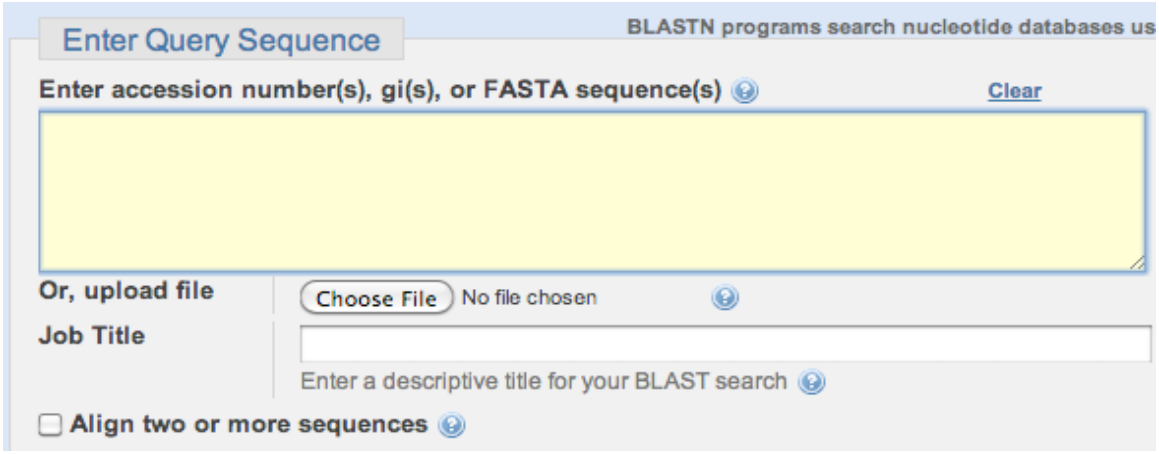
11. **On MABL,** above the tree rendering results there should be a list of steps**.** Select **Data & Settings.** Obtain the file with the unknowns. Copy/paste the unknown sequences into the search box at the end of the previously uploaded sequences.

The program will generate a new tree for you to download and analyze. Name your file "Unknown Phylogenetic Tree" and save it to your desktop where it can be easily accessed. Now that you have created two phylogenetic trees you will compare both and try to identify the position at which your unknown species have been placed.

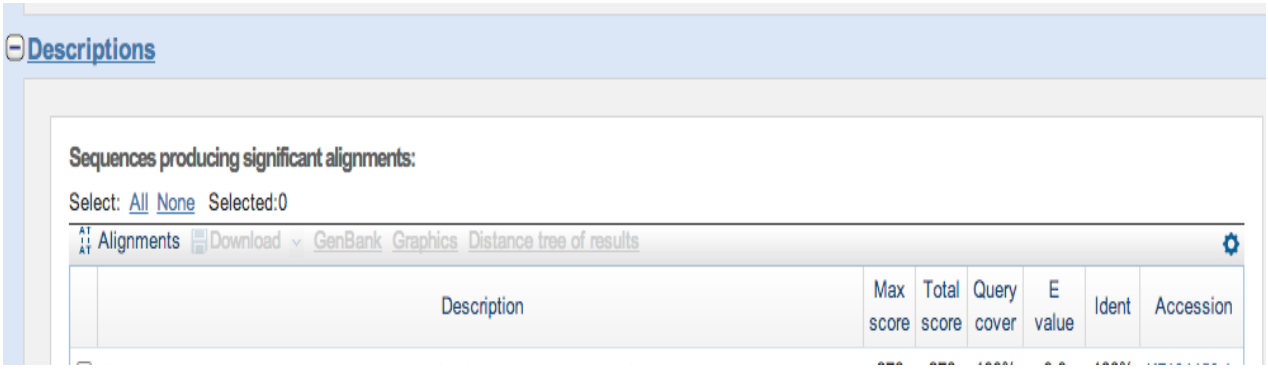
Next, you will perform a blast search on your unknown species.

12. **Go back to GenBank**. On the right side of the webpage click on **Run Blast**. A BLAST search will give you access to displays of the alignment results, links to related information for matched sequences, and a detailed description of your results.

13. Delete any numbers in the search box, copy/paste the first unknown sequence, and select "BLAST."



The BLAST search will give you the most significant alignments. Scroll down to the descriptions section and examine the columns to the right of the described species.



Each column lists…

- The description/title of matched database sequence
- The highest alignment score (Max score) from that database sequence
- The total alignment scores (Total score) from all alignment segments
- The percentage of query covered by alignment to the database sequence
- The best (lowest) Expect value (E value) of all alignments from that database sequence
- The highest percent identity (Max ident) of all query -subject alignments, and
- The Accession of the matched database sequence

Now that you understand how to create a phylogeny and why phylogenies are important, add/remove sequences to create your own evolutionary trees for comparison. Search through various accession information by entering species name into the Nucleotide search box on GenBank. Experiment with your trees by using different sequences and out-groups and write a short narrative about your tree(s).

Discussion

1. What is depicted in a phylogenetic tree? In your own words, describe what a phylogeny represents. What are sister groups? What is an out-group? Where would they and the last common ancestor be positioned in a phylogeny?

2. This activity describes a research project. What are researchers studying and why is it important? What are Bryophytes and why should researchers study them?

3. How are we accessing information about the genus *Frullania*? What information is provided when you search through an archive?

4. What is a FASTA file and what information does it contain?

5. Compare both phylogenetic trees. What has changed? Describe what you see. Where did the unknowns fall? Did the unknowns fall into a particular clade? Try to identify the species based on its position on the tree, and explain your reasoning.

6. Add and/or subtract branches from your tree. What has changed? What does this say about the evolution of liverworts and *Frullania*?

7. After running the blast search, did the names come in where you thought? Do all unknowns match? What is a BLAST search and how is/isn't it useful? Explain why GenBank may have produced BLAST results in which two species were identified as matching your unknown sequence.

8. What is depicted in a phylogenetic tree? In your own words, describe what a phylogeny represents. What are sister groups? What is an out-group? Where would the last common ancestor be positioned in a phylogeny?

9. Compare both phylogenetic trees. What has changed? Describe what you see. Where did the unknowns fall? Did the unknowns fall into a particular clade? Try to identify the species based on its position on the tree, and explain your reasoning.

10. Add and/or subtract branches from your tree. What has changed? What does this say about the evolution of liverworts and *Frullania*?

11. After running the blast search, did the names come in where you thought? Do all unknowns match? What is a BLAST search and how is/isn't it useful? Explain why GenBank may have produced BLAST results in which two species were identified as matching your unknown sequence.

Table 1 Species and accession numbers of *Frullania* isolates.

Genus	Species	Accession Number
A. *Frullania*	*rostrata*	JQ283999.1
B. *Frullania*	*rostrata*	JQ283997.1
C. *Frullania*	*rostrata*	JQ284001.1
D. *Frullania*	*rostrata*	JQ284001.1
E. *Frullania*	*tamarisci*	JQ284000.1
F. *Frullania*	*tamarisci*	HM167589.1
G. *Frullania*	*tamarisci*	HM167588.1
H. *Frullania*	*tamarisci*	HM167587.1
I. *Frullania*	*eborasensis*	HM167586.1
J. *Frullania*	*eborasensis*	HQ330414.1
K. *Frullania*	*eborasensis*	HQ330413.1
L. *Frullania*	*eborasensis*	HQ330412.1
M. *Frullania*	*dilatata*	HQ330411.1
N. *Frullania*	*dilatata*	HQ330403.1
O. *Frullania*	*dilatata*	HQ330402.1
P. *Frullania*	*dilatata*	HQ330399.1
Q. **Outgroup taxon**: Nipponiejeunea	*plifera*	HQ330401.1

References

Understanding Evolution. 2014. University of California Museum of Paleontology. 22 August 2008 <http://evolution.berkeley.edu/>.

Baum, D. (2008) Reading a phylogenetic tree: The meaning of monophyletic groups. *Nature Education* 1(1):190

Content developed by Justyna Drag, Lauren Hasan, Dr. Thomas Campbell, and Dr. Matthew Grief with additional resources provided by Dr. Matthew von Konrat.

PHYLOGENY UNKNOWNS

>UNKNOWN A voucher RA0886 recombination activating protein 2 (RAG2) gene, partial cds
CCAGGCTTTTCTCTGCTGAATTTTGATGGGCATGTTTTCTTTTTTGGGCAGAAAGGATGGCCAAAGAGAT
CCTGTCCCACTGGTGTTTTCCTCCTTGATATAAAACAGAATGAGCTCAAAATGAAACCTGCAATCTTCTC
TAAAGATTCCTGTTATCTTCCCCCTCTCCGCTACCCTGCTATTTGCACTCTCAGAGGCAATGCAGAGTCC
GATAAGTACCAGTATATCATTCACGGTGGGAAAACACCTAACAATGATCTTTCTGATAAGATTTACATTA
TGAGTCTTGTAAGCAAAAATAGCAAGAGAACCACATTTCAATGCATTGAGAAAGATCTGGCCGGTGATGT
TCCTGAAGCAAGATATGGACATACAATTAATGTAGTTCATAGCCGGGGAAAAAGCATGAGTGTTATTTTT
GGAGGAAGATCATATGTTCCTCTTGGACAAAGAACCACTGAAAAATGGAATAGCGTAGTTGACTGTTTGC
CATCTGTGTTTCTGGTTGATTTCGAGTTTGGATGCTGTACGTCATACATACTTCCAGAGCTTCAAGATGG
ACTTTCTTTCCATGTTTCTGTTGCCAAAAATGATACGATCTACATTTTGGGTGGTCATTCACTTCAAAAT
AATATCAGGGCCCCCAGCTTATACAAGCTAAAAGTTGATCTCCCACTGGGCAGCCCAGCTGTGACTTGCA
CTATATTGCCAGGTGGGATATCTGTGTCAAGTGCTGTTGTGACTCAGATCGGTGATACAGAATTTGTCCT
TGTTGGGGGCTATCAGTCTGACAACCAGAAAAGGTTGGTTTGTAACACCATAGTTTTAGAAGATAACAAG
ATAGAGATTGTTGAAAGGGAGACCCCACGCTGGACACCAGATATTAAACACTGCAGGATATGGTTTGGCT
GTGATATGGGCAAAGGGTCTGTATTGCTGGGCATTCCAGGGGACAACAAACAGATGATCTCAGATGCAAA
CTACTTCTACATTTTGAGATGCAGAGGAGCAGAAGAGGATGAGGAGGAAGAACTGACAGCCCAAACTTGC
AGTCAGACATCTACTGAAGACCCAGGAGACTCCACTCAGTTTGAAGATTCAGAAGAGTTTTGTTTTAGTG
CTGAAGCCAGTAGCTTTGATGTTGATGATACTGACACTTACAATGAAGATGATGAAGAAGATGAATCAGA
AACGGGCTACTGGATCACCTGCTCTGCTAGCTGCAATGTTAACATTAACACCTGGGTCCCTTTCTATTCA
ACAGAACTCAACAAGCCTGCCATGATCCTGTGTTCCAGCAGGGATGGCCACTGGGTCCACGCACAATGTA
TGGATCTGTCAGAGACCATGCTCCTGCATCTCTCGGAAGCAAATGTCAAGTACTTCTGCAACGAGCATGT
TGACCTTAATAAAGGGCTCCAA

>UNKNOWN B isolate B50f internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene, complete sequence; and internal transcribed spacer 2, partial sequence
CGAATCTCTCGTCCTCATCCTCGGGCACGTCTCGAGGAGAGGGGCGAGGAACCACTTCAAAACTATCAAA
ATGCACCTACGAAAAGGAACGGCAAGCTTTAGAAACCCGAGGGGATCGCACTGTAGAGGTGCGGTCGTCG
ACGATGCGATTCCGGGACTCGCAGGCTCCCTCCTCTCAGGGAGAGGGTGCTTGGTAGTGCCACCGCGGGT
GTGTCGCGGGCGTGTTTCCATGGTCGTCGGGGTGTCCCGCTCGCACGACTCCCAAACAACGTCTCCCCCC
ATCTCCAAGAGGCGGGGGAGACTTGCTGGGAGTAGGACGTTTCCTCCCCTCCTCTCGGGAGCACCCTTCA
CGCCGGGTCGTCCCCTCCTCTATCCCGGAGGAGGAGGCTCTCGGTTTGCGGGATCTCGGGAGCATCCTCG
GCTTGTCCTTCCGTGGGTGCAGAATCAGCCAAATTTCCAACGGTGAACCGAAGACAGACGAAATTTATCA
AACAAATGGACTCTCAGCAACGGATATCTTGGCTCTTGCAACGATGAAGAACGCAGCGAAATGCGATACC
TAGTGTGAATTGCAGAATTCCGCGAATCATCGAGTTTTTGAACGCAAGTTGCGCCCGAGGCTTTTACCGA
GGGCATGTCTGTCTGAGCGTCCATGGGGTCCCGCTCATCCGCCTCGCCGTGTTGGCGAGGATTCATAGGA
TGGGCGCTGGGATGGCCTATGCGCGGACCATCCATTGGGCTCGCCCAAGGGAAAGTGTCCCGCTCGGCTG
AAATCTATGTCCGTCCTGGAAATCGGAAGGGACGTTGCGAGAAGGTGCTCCGCTTTGCGAAAGTGAGCGG
ATGATATCATGTATGTCAACAACAGCACGCGACGCTCCCTTCCCTCTCGAAAGGACTCGGAAAAAACGGG
ACTCCCGTCATCGCACGGGAGTTTTTCACGTCTTGGACC

>UNKNOWN C isolate B63c internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene, complete sequence; and internal transcribed spacer

2, partial sequence

CGAATCTCTCTGTCCTCGATCCTTGGGCACGTCTCAAGGAAAGGGGCGAGGAACCACTTCAAAACTATCA
AAAATGCACCTACGAGAGGAACGGCAAGCTTTAAAAACCCGAGGGGATCGCACTGTAGAGGTGCGGTCGT
CGACGATGCGATTCCGGGACTCGCAGGTTCCCTCTGCTTAGGGAGAGGGTGCTTGGCAGTGCCACCGCGG
GTGCGTCGCGGGCGTGTTTCCATGGTCGTCGGGGTGTCCCGCTCGCACGACTCCCAGACAACGTCTCCCC
CCTCTCCCAGAGGCGGGGGAGGCCTGTCGGGAGTAGGACGTTTCCTCCCCTCCTCTCGGGAGCACCCTTC
ACTTCGGGTCGCCCCCTCCTCTGTCCCGGAGGAGGAGGCTCTCGGTTTGCGGGATCCTGGGAGCATCCCC
GGCTTGTCTTTCCGTGGGTGCAGAATCAGCCAAATGTCCAACGATGAACTGAAGACAGACGAAATTTATC
AAACAAATGGACTCTCAGCAACGGATATCTTGGCTCTTGCAACGATGAAGAACGCAGCGAAATGCGATAC
CTAGTGTGAATTGCAGAATTCCGTGAATCATCGAGTTTTTGAACGCAAGTTGCGCCCGAGGCTCTTGCCG
AGGGCATGTCTGTCTGAGCGTCCATGGGGTCCCGCTCATCCGCCTCGTCGTGTTGGCGAGGATTCATAGG
ATGGGCGCTGGGATGGCCTATGCGCGGACCATCCATTGGGCTCGCCCAAGGGAAAGTGTCCCGCTCGGCT
GAAATCTATGTCCGTCCTGGAAATCGGAAGGGACGTTGCGAGAAGGTGCTCCGCTTCGCGAAAGTGGGCG
GATGATATCATGGATGTCAATAGCACGCGACGCTCCCTTCCCTCTCGAAAGGACTCGGAAAAAACGGGAC
TCCCGTCATCGCACGGGAGTTTTTCACGTCTTGGACC