Lab Topic 17

Molecular Phylogeny of Plants

Laboratory Objectives

After completing this lab topic, you should be able to:

1. Describe the evolutionary history (phylogeny) of land plants.
2. Construct a phylogenetic tree using morphological evidence.
3. Use the tools of bioinformatics to test hypotheses in the form of phylogenetic trees.
4. Analyze results, and write an illustrated report, using critical thinking skills and synthesizing data from a variety of sources.
5. Describe evolutionary relationships, including common ancestry and changes over time in lineages.
6. Discuss the importance of bioinformatics and molecular evidence in evolutionary analysis.

Introduction

In Lab Topics 15 and 16 you investigated the diversity of the plant kingdom, comparing representative plants. You studied vegetative and reproductive structures, observed fossils, and investigated moss, fern, pine, and flowering plant life cycles. Land plants have been classified into ten phyla based on their similarities and differences. Morphological and biochemical features shared by all land plants provide evidence that the closest living relatives to land plants are complex green algae called charophyceans. Advances in biotechnology in the 1970s and 1980s that permitted the analysis and sequencing of DNA provided additional genetic evidence for this relationship. In 1994, the Green Plant Phylogeny Research Coordination Group was formed at the University of California, Berkeley, to facilitate the efforts of various research groups working on the study of genetic relationships between plants and their algal ancestors. This began the "Deep Green" project, a large-scale collaborative study of plant evolution based on examination of nuclear and chloroplast gene sequences of algae and land plants. You can learn more about the Deep Green initiative at http://ucjeps.berkeley.edu/bryolab/GPphylo/proj_summ.php.

In this lab topic you will use your previous laboratory experience and knowledge to develop a hypothesis for the evolutionary relationships among land plants and their green algal relatives. You will test your hypothesis by

comparing molecular evidence from DNA sequences of a chloroplast gene using the tools of bioinformatics (the discipline connecting computer science and biology). You will construct a tree that represents plant evolutionary history (phylogeny). Before leaving the laboratory, you will prepare a final report incorporating your hypothesis and results.

Understanding Phylogenetics

Phylogeny is the study of evolutionary relationships among groups of living and extinct organisms on earth. Scientists analyze structural, reproductive, physiological, or molecular changes in specific characters. In particular, they utilize **homologous** characteristics, those features that are similar as a result of shared ancestry (common descent). Molecules, including DNA, can also reveal homology, since these are passed down from one generation to the next.

Phylogenetic systematics is the field of biology that examines morphological characteristics, biochemical pathways, and gene sequences to establish relationships among groups of organisms. To show evolutionary relationships, a **phylogenetic tree** (a visual representation of the lineages among organisms) is constructed. All phylogenetic trees are hypotheses that are to be tested, modified, and tested again. One type of phylogenetic tree, known as a **rooted tree**, contains a root, nodes, branches, and clades (Figure 17.1).

The **root** of the tree represents a common ancestor from which all the organisms in the group are derived. Within the tree are several **nodes**. A node represents a branching point for a taxonomic unit (such as group, phylum, genus, or species). **Branches**, the lines that extend from nodes, establish how closely related one group is to the other. The shorter the line, the more closely related the groups. Groups sharing a node share a common ancestor and make up a **clade**. The branching pattern of the tree and the length of the branches reflect the number of changes or differences that have accumulated among the species or groups. In Figure 17.1 wolves, skunks, and *Drosophila* (fruit flies) are all animals, and therefore their **shared ancestral traits** are eukaryotic, multicellular organisms that consume food. But wolves and skunks share some traits that are not shared by *Drosophila*. These shared traits (for example, hair and milk production) are known as **derived traits** and determine the specific clade—mammals—of the organisms that share them.
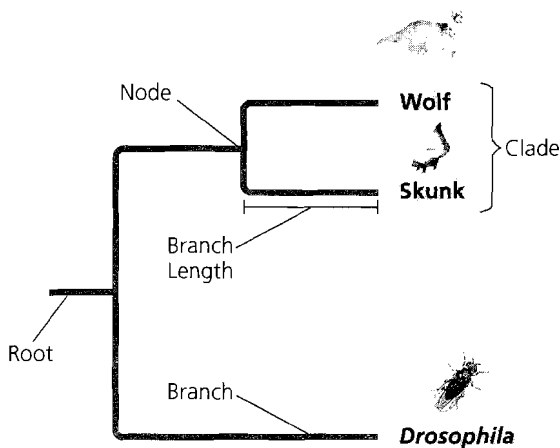
Figure 17.1.
Terminology associated with a phylogenetic tree reflecting evolutionary relationships.

Phylogenetic trees can be constructed using a variety of different methods to establish the nodes, branches, and topology of the tree. Refer to your textbook (see Campbell and Reece, 8th edition, Concept 26.3) for further information on methodologies.

Molecular Phylogenetics

Rapid technological advances in molecular biology have allowed scientists to obtain DNA sequences of genes and entire genomes of a large variety of organisms. **Molecular phylogenetics** is the field that examines evolutionary relationships among groups specifically based on changes occurring in DNA and protein structure. The number of differences in DNA sequences between different groups reflects the accumulation of mutations over the time since they shared a common ancestor. Phylogenetic trees can be constructed to illustrate these differences in DNA sequences for groups of organisms.

Four steps are important in the construction of phylogenetic trees based on molecular data (NCBI, 2004).

1. **Align** similar DNA sequences from different groups to detect similarities and differences in nucleotide bases.
2. **Establish sequence variation** by observing the level of homology or similarity of sequences among groups.
3. **Build a tree** by arranging groups based on the percentage of matching bases for sequences and other factors.
4. **Evaluate the tree**, including the analysis of the resulting tree and comparison with trees constructed with nonmolecular data.

Using DNA sequences in phylogenetics can generate very large data sets. Imagine comparing 10 different species for a gene sequence with 100 nucleotides and the possibility of 4 different nucleotides at each position. To cope with such huge amounts of data, scientists use the tools of bioinformatics to construct molecular phylogenies.

Using Bioinformatics

Bioinformatics is the field of science combining the disciplines of biology and computer science. Bioinformatics uses computer algorithms (a clear set of instructions for solving a problem) to conduct analyses of large biological data sets. Applications include aligning DNA and protein sequences, predicting protein structure, creating restriction enzyme maps for a DNA sequence, and predicting translation products. Many of these computer algorithms are now available for public access through the Internet and have become a staple analysis tool for molecular biologists. One important website is the **National Center for Biotechnology Information (NCBI)** established by the National Library of Medicine at the National Institutes of Health. NCBI manages GenBank®, an international database and collection of all publicly available DNA sequences, which now number over 61 million. NCBI also provides links to tools for molecular data analysis such as BLAST, a program used to find DNA sequences with similarity to an unknown sequence. Another excellent resource for bioinformatics applications is **Biology WorkBench**, a Web interface that allows rapid access to biological databases, such as GenBank®, and analysis tools, such as BLAST and CLUSTALW.

In the following laboratory exercise, you will use bioinformatics tools in Biology WorkBench to construct a phylogenetic tree or gene tree, using DNA sequences of land plants and green algae.

EXERCISE 17.1

Establishing Molecular Phylogenetic Relationships among Rubisco Large Subunit Genes

Materials

computers with Internet access, one per pair of students
2 computer desktop folders: "images" and "rbcL nucleotide seq" (available
    at www.masteringbio.com in the Study Area under *Investigating Biology*).
printer access
living examples of *Chara, Nitella, Polytrichum, Marchantia, Equisetum,*
    Polypodium, Psilotum, Pinus, Zamia, Arabidopsis, and *Lilium,* if available
20 sheets of large paper (11" × 17")

Introduction

The ribulose bisphosphate carboxylase (Rubisco) protein is essential to carbon fixation in photosynthesis and is found in green algae and all land plants. Therefore it is an ideal choice to establish phylogenetic relationships among green algae and land plants using nucleotide sequences. The gene sequence for the large subunit (rbcL) of the Rubisco protein has been isolated from a large variety of algae and plants. The DNA sequences of rbcL genes for *Chara, Nitella, Polytrichum, Marchantia, Equisetum, Polypodium, Psilotum, Pinus, Zamia, Arabidopsis,* and *Lilium* are available in GenBank.

Working in pairs, you will use your knowledge and laboratory experience from previous lab topics to develop a phylogenetic tree that depicts the evolutionary relationships for the organisms listed above. This will serve as your hypothesis, which you will test using molecular data for the rbcL nucleotide sequences. You will access the supercomputer, which will use the tools of bioinformatics to compare the sequences for rbcL genes for each of the above organisms, then perform an alignment of sequences, and construct a phylogenetic tree reflecting the relationships among these genes and groups.

Developing Hypothesis and Prediction

Examine the organisms selected for this exercise and determine which are charophytes, bryophytes, pterophytes, gymnosperms, or angiosperms. The plants may be on demonstration in the laboratory and images are provided in a folder on the desktop of your computer. Based on your knowledge of plant diversity and the *ancestral* and *derived* characteristics of the major phyla of plants, develop a hypothesis for which organisms will be more closely related to each other. Begin by arranging the organisms most closely related to each other into clades. Next, consider which clades might share a common ancestor. Refer to Table 16.6 to review characteristics of land plants studied in Lab Topic 15 Plant Diversity I: Nonvascular Plants (Bryophytes) and Lab Topic 16 Plant Diversity II: Seed Plants.

Hypothesis

Based on your hypothesized relationships, draw a phylogenetic tree showing appropriate branching and distances to reflect where each organism should be placed relative to the others. Make preliminary sketches and a final drawing on the large sheets of paper available in the laboratory. Use your lab manual (Lab Topics 15 and 16) and your textbook (Campbell and Reece, 8th edition, Chapters 26, 29, and 30) to guide you in constructing your phylogenetic tree. This tree is the visualization of your hypothesis, which incorporates diverse sources of information: morphology, vegetative, and reproductive structures, variations in life cycles and fossils. What is the **root** of your tree? Label **branches** and **clades**. Refer to the introduction at the beginning of this Lab Topic to review terminology and concepts. (*Note: this tree will be referred to as the "morphological" tree, even though other evidence was included in its development.*)

Prediction

You will be testing your hypothesis that the morphological tree accurately represents land plant phylogeny. You will use molecular data from the nucleotide sequences of rbcL. Do you think the molecular evidence will support (be consistent with) your hypothesis or falsify it? Write your prediction below.

Using Biology WorkBench

Imagine how many comparisons are possible for 11 species, comparing changes at each nucleotide position in rbcL DNA! To make this possible, you will use Biology WorkBench with access to the San Diego Supercomputer Center, which will compare and align the nucleotide sequences for all of your organisms. The supercomputer will then construct phylogenetic trees using a mathematical algorithm to select the one phylogeny from those possible that is best explained by the molecular data. Remember that the more similar two sequences are, the more closely related they are (the more recent their common shared ancestor). The more nucleotide changes that occur, the more time since two groups share a common ancestor, and therefore they are more distantly related. There are many computer programs that use different algorithms for analyzing these large data sets. In this exercise you will use the one named **ClustalW**. You and your lab partner will do the evaluation and analysis of results, comparing the molecular phylogenetic tree to your hypothesized phylogenetic tree, which was based on morphology and life cycle features.


While using Biology WorkBench—*do not use the "Back" button* on your Web browser. Use only the "Return" or "Abort" keys on the WorkBench menus. ***Using the "Back" button will overload and crash the supercomputer.***

Procedure

1. Open a new Word document. Next, open a Web browser and type http://workbench.sdsc.edu/ to enter the Biology WorkBench website.

Figure 17.2.
Biology WorkBench home
page. Register and enter WorkBench.
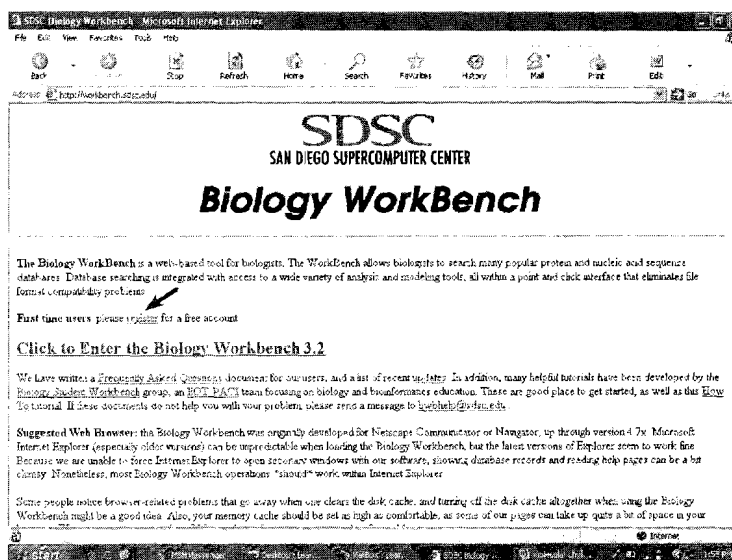
Note that during this exercise you will move back and forth between these two windows.

2. To acquire access to Biology WorkBench, you must register your team. Click on the link "register" under first-time users (Figure 17.2). Supply your name (choose one member of the team) and your email address. Select a user ID and a password that is easy to remember; *be sure to write it down*. Click on the "Register" button. Only one member of the team should register.

3. After you register, a new dialog box will appear on your screen asking for your user ID and password; enter the user ID and password you just selected. This will allow you to enter the WorkBench site. Alternatively, if you return to the main page, click on "Enter the Biology WorkBench 3.2" (Figure 17.2), which will prompt you for the user ID and password.

4. When you enter Biology WorkBench, the first page describes the web tools and provides an introduction. Scroll toward the end of the page and select "Session Tools" (Figure 17.3). This will lead you to a new page entitled "Default Session."
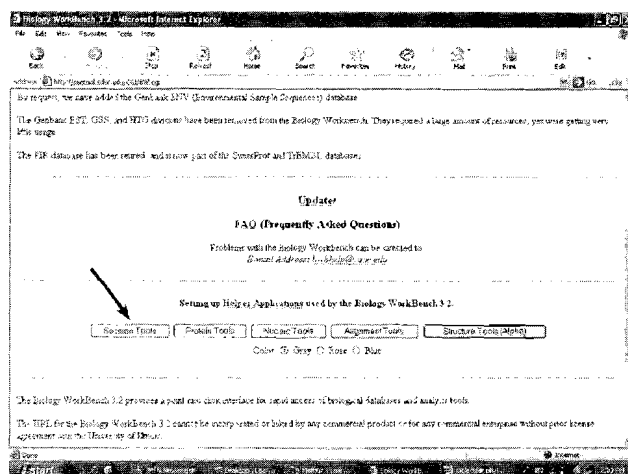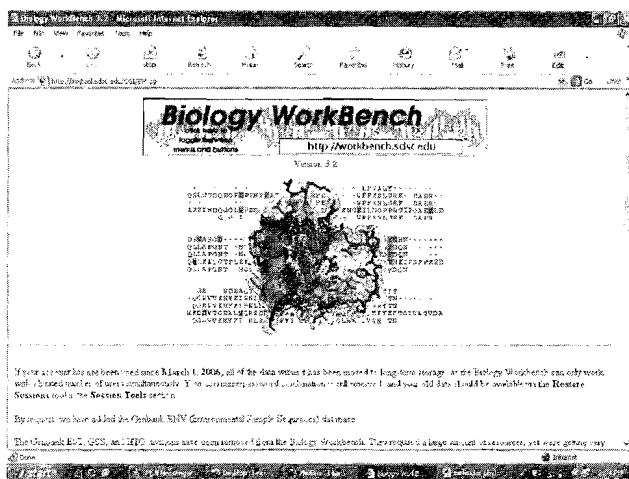

Figure 17.3.
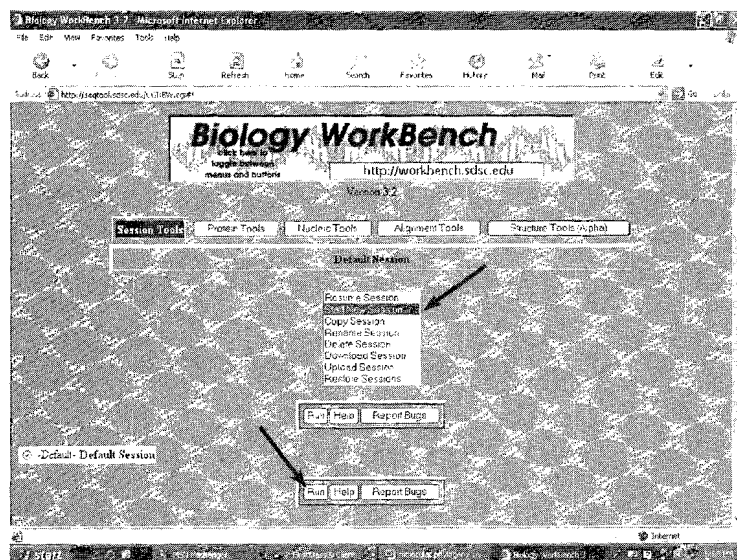Biology WorkBench. Selecting "Session Tools."

Figure 17.4.
Biology WorkBench. Starting a new session.

5. Using the scroll menu in the middle of the page, select "Start New Session" (Figure 17.4). Then click the button "Run" immediately below the scroll menu. A new page will open. Select a name for your session—such as molecular phylogeny lab—and type this in the box labeled "Session Description." Click the button "Start New Session" right below the session description (Figure 17.5).
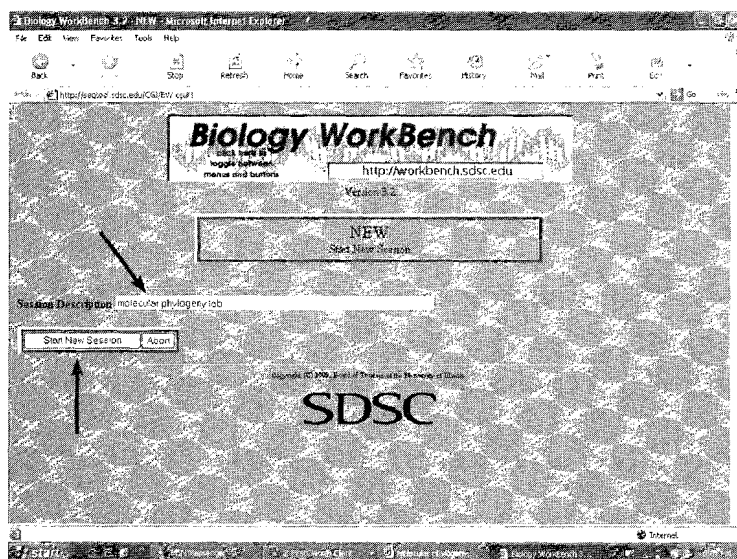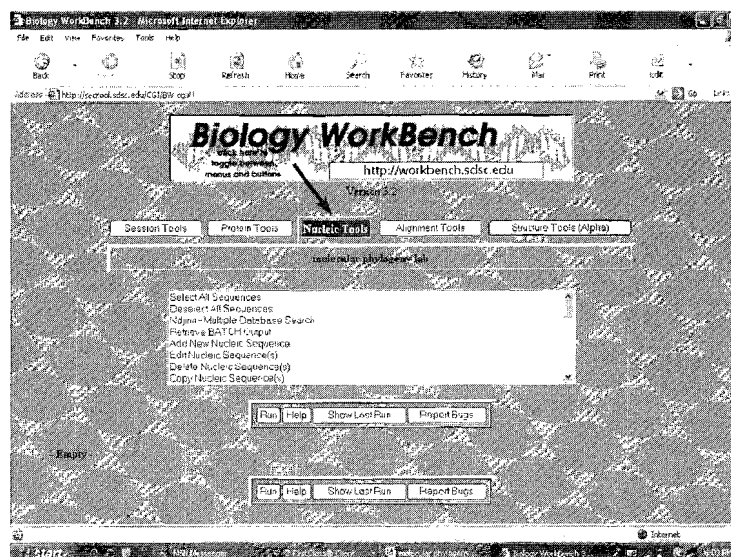

Figure 17.5.
**Biology WorkBench.** Entering a session description.

6. A new page will open with your new session listed (Figure 17.6). Next select "Nucleic Tools" from the menu at the top. After you select this option, a new page will appear with a scroll menu in the middle of the page.

Figure 17.6.
Biology WorkBench. Selecting "Nucleic Tools."

7. Minimize (but do not close) the Biology WorkBench window. (*If you accidentally close the WorkBench window, follow the steps to log in again.*) Locate the Word file labeled "rbcL nucleotide seq" on your computer Desktop. Open the document, then go to "Edit" and "Select All" to highlight entire text in the document. Right click on the mouse and select "Copy."

8. Minimize the Word document and return to the Biology WorkBench window. Select "Add New Nucleic Sequence" on the scroll menu in the middle of the page (Figure 17.7). Click on "Run."

Figure 17.7.
Biology WorkBench. Adding a new nucleic acid sequence.
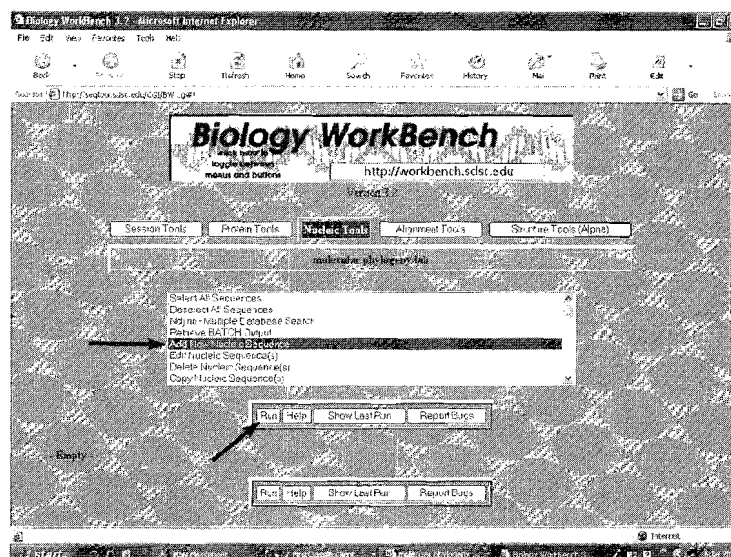
A new page requesting input of new sequences will open. Click on the large box labeled "Sequence." Right click on the mouse and choose "Paste" from the menu. You should now see the DNA sequences pasted in WorkBench (Figure 17.8).
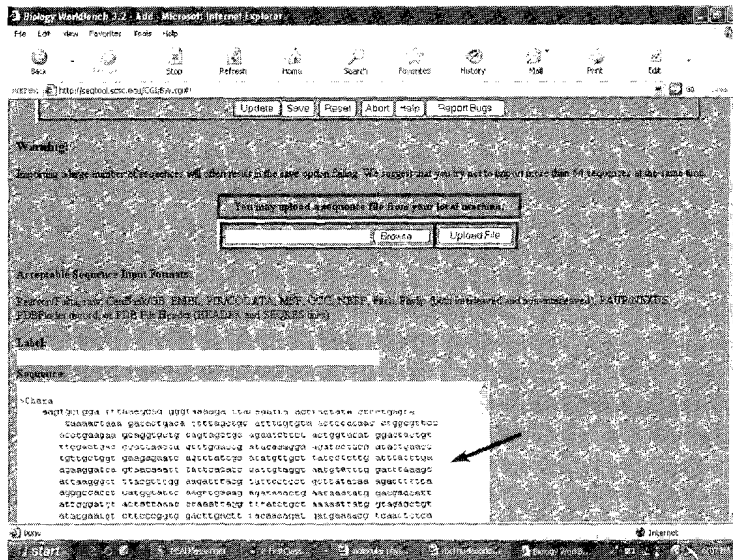

Figure 17.8.
Biology WorkBench. Pasting the nucleotide sequences.

9. After you have pasted the rbcL sequences into the Sequence box, click "Save" at the bottom of the Web page. A new page will open, containing the main scroll menu. The organism names for your uploaded sequences will appear in the middle of the page as "User Entered" (Figure 17.9). Select one of the uploaded sequences by checking the box on the left of the sequence. Select "View Nucleic Sequence" from the scroll menu (scroll down to find this). Then hit "Run." This allows you to check to be sure that the sequence has been uploaded for the organism listed. (You do not have to check all the organisms.) Next hit "Return" at the bottom of the page to return to the previous page.
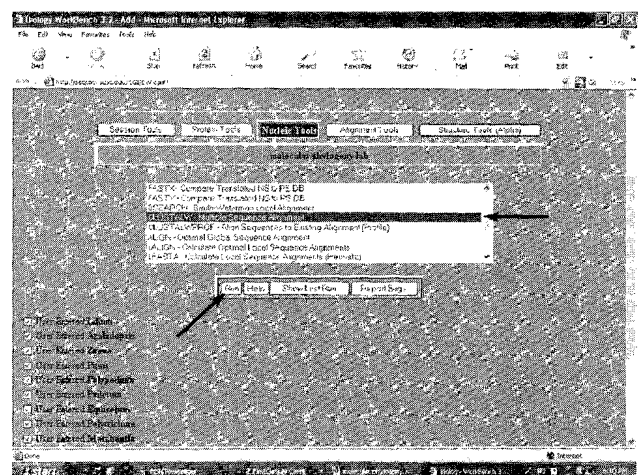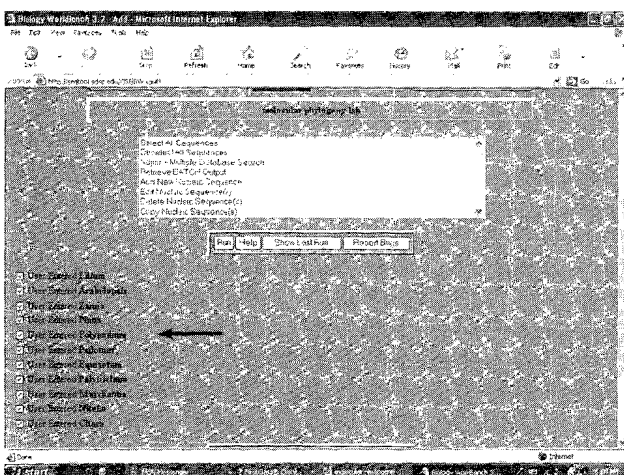

Figure 17.9.
Biology WorkBench. Entered sequences.

※ **Do not use the Back Button!**

---

10. Now select *all* of the uploaded sequences under the main menu by checking the appropriate boxes. From the scroll menu in the middle of the page, choose "CLUSTALW—Multiple Sequence Alignment" and click "Run" (Figure 17.9). The program will select the sequences and open a new page.

11. In the new ClustalW page, below the selected sequence files, find "Guided Tree Display" (Figure 17.10). Change this to "Rooted Tree." Next click on "Submit" at the bottom of the page. *Be patient* as ClustalW analyzes and processes the data. Do not click on links or arrow keys while the computer is retrieving data. A new page will appear with the results.

Figure 17.10.
Biology WorkBench. Options in ClustalW.
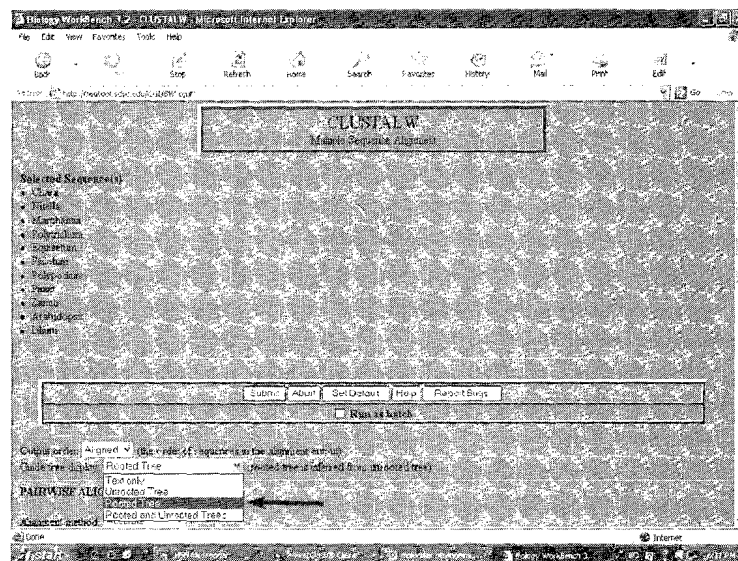
12. Scroll to the bottom of the ClustalW results page. Find the phylogenetic tree for the entered sequences. Place the cursor on this image, right click on the mouse, and choose "Copy." *Do not close the browser window!* Return to the open Word document on your desktop and paste the image into the document (alternatively you can paste the image in Paint). Save the document to the Desktop. Keep the Biology WorkBench/ClustalW browser window open to continue analyzing the results as instructed. *Closing this window will result in the loss of your alignment results.*

13. Print a copy of the molecular tree pasted in the Word document.

14. Return to the Biology WorkBench/ClustalW window. Scroll to the sequence alignments located above the phylogenetic tree. Review the sequence alignments and record any interesting observations about the sequence variability among the rbcL genes of the selected organisms. Continue your analysis in the following Results section.

Results

1. Record any information about the variation in rcbL sequences you observed in the ClustalW analysis.

2. Review the results. Can you determine the number of nucleotides for this gene? Note: The nucleotide sequence for some of the species may not include the entire rbcL sequence.

3. Observe the alignments and list the plants or algae that have the most similar alignment.

Discussion

Discuss your results in the report you prepare in Exercise 17.2. In this report you will analyze your molecular phylogenetic tree and compare your results with the morphological tree that you hypothesized for these eleven organisms.

EXERCISE 17.2
Analyzing Phylogenetic Trees and Reporting Results

Materials

computer with printer access
computer desktop folder with images
scissors and tape

Introduction

The supercomputer completed the enormous task of aligning the nucleotide sequences and developing the phylogenetic tree that best fits the data, looking at shared sequences (evidence of common ancestry) and sequence differences (accumulated after divergence from a common ancestor). Now you will compare your results to your hypothesized phylogenetic tree. If you predicted that the molecular data would be consistent (or not) with your hypothesis, what evidence do you have to support or falsify your hypothesis? Once you have completed your analysis, you and your partner will use your results and analysis to construct a report to submit at the end of the laboratory period.

Procedure

1. Along with your lab partner, examine the observations you made about the sequence alignment results provided by ClustalW. Examine the phylogenetic tree constructed based on rbcL genes of different organisms. Label the **root**, a **branch**, a **node**, and a **clade** within your phylogenetic tree. Paste the images of each organism from the folder on your computer desktop onto your phylogenetic tree. (*Note:* you may cut and paste using the computer or print the images and use scissors and tape!)

2. Provide an explanation of the evolutionary relationships among the organisms as shown by this tree. Compare your molecular phylogenetic tree to your hypothesized morphological tree. Describe any similarities or differences and your thoughts on why there may be differences.

3. Together with your partner, create and type a one-page report (double spaced, size 12 font), that includes your hypothesis and prediction and a description of your results and analysis. You should attach your morphological tree as part of your hypothesis. Be sure to identify the phylum for each of the organisms and briefly describe their relationships in the text of your report. Label your hypothesized morphological tree and the molecular tree as figures and refer to them in your report. Include both names of the team members on your report. Turn in your report along with the figures before leaving the laboratory.

Questions for Review

1. Define and describe the following terms: *phylogeny, phylogenetic tree, phylogenetic systematics, molecular phylogenetics, bioinformatics, sequence alignment, homology.*

2. Draw a simple phylogenetic tree for two sister clades with a common ancestor. Define the following terms and use them to label your diagram: *clade, node, root,* and *branch.*

3. What is rbcL, and why is it a particularly useful molecule for studying evolutionary relationships in plants and green algae?

Applying Your Knowledge

1. For the following phylogenetic tree, label the boxes on the branches with the derived trait that is shared by the members of each clade. For example, on the lower branch a box has already been labeled for the feature "cuticle," which is shared by all land plants.
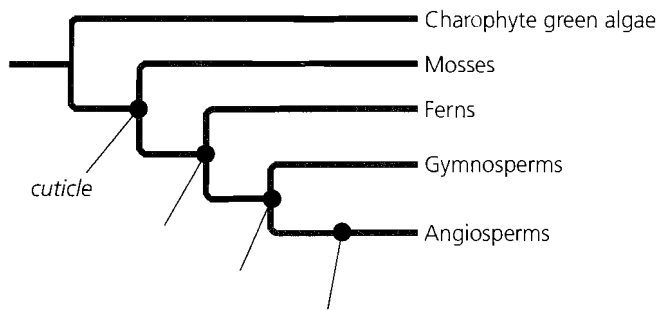
2. Can you suggest reasons why a phylogeny based on molecular evidence and a phylogeny based on morphology and other evidence might not be exactly the same?

3. Zoologists worldwide are sequencing a mitochondrial gene CO1 (cytochrome c oxidase subunit), which is found in all animals and appears to be distinctive for each species. The sequence of nucleotides can be used as a universal DNA bar code. By comparing the CO1 DNA sequence for an animal to a growing database of DNA sequences, scientists can accurately identify any animal and also discover species not previously known to science. How might DNA bar coding, which uses molecular biology and bioinformatics, be useful in enforcing international laws for banning the import of endangered species? How might these approaches stimulate the study of biodiversity in remote areas?

4. When scientists compare the phylogenetic trees based on molecular data with existing trees based on morphological characteristics, in over 90% of the cases the molecular data confirms the relationships previously recognized. However, there are some surprises. For example, water lilies and the water lotus were classified as close relatives, but recent molecular analyses have shown that the water lotus is more closely related to the sycamore tree than to water lilies. Scientists have to critically evaluate these conflicts. Do you think that the evidence from molecular analyses should be the deciding factor in resolving these conflicts with other kinds of evidence (morphology, reproduction, biogeography, fossils, etc.)? How do you think scientists should weigh the evidence?

Investigative Extensions

You may have time to modify your analysis during the laboratory period, or you may be inspired to pursue bioinformatics and molecular phylogeny as an independent investigation.

1. To modify the analysis of the species featured in this laboratory, consider doing the same comparisons (using all species), but select "Unrooted Tree" in the CLUSTALW screen. Then compare your results for rooted and unrooted trees. Go to the NCBI website and scroll to Figure 1 to read about rooted and unrooted trees.
   http://www.ncbi.nlm.nih.gov/About/primer/phylo.html

2. Do you think the results of your analysis might be different if you selected only some of the eleven species used in this lab topic? Return to Figure 17.9 in the Procedure section and select only those sequences that you would like to investigate in a simplified or at least different phylogenetic analysis.

Student Media Activities and Investigations

Activities—Ch. 29: Highlights of Plant Phylogeny
Investigations—Ch. 26: How Is Phylogeny Determined by Comparing Proteins?
www.masteringbio.com

References

This lab topic was coauthored with Dr. Nitya Jacob, Assistant Professor of Biology, Oxford College.

Cameron, K. M. "Bar Coding for Botany." *Natural History*, 2007, 114(2), pp. 52–57.

Campbell, N. A. and J. B. Reece. *Biology*, 8th ed. San Francisco, CA: Benjamin Cummings, 2008.

Futuyama, D. J. *Evolution*. Sunderland, MA: Sinauer Associates, Inc., 2005.

Mauseth, J. D. *Botany: An Introduction to Plant Biology*. Sudbury, MA: Jones and Bartlett Publishers, 2003.

National Center for Biotechnology Information. Revised 2004, April. "Just the Facts: A Basic Introduction to the Science Underlying NCBI Resources—Systematics and Molecular Phylogenetics." http://www.ncbi.nlm.nih.gov/About/primer/phylo.html. Accessed Aug. 1, 2007.

Websites

Bedrock: Bioinformatics Education Dissemination: Reaching Out, Connecting and Knitting-together. Integrating bioinformatics in undergraduate biology curriculum:
http://www.bioquest.org/bedrock/

Biology WorkBench home page:
http://workbench.sdsc.edu.

Green Plant Phylogeny Research Coordination Group home page:
http://ucjeps.berkeley.edu/bryolab/GPphylo/

National Center for Biotechnology Information home page:
http://www.ncbi.nlm.nih.gov/

Tree of Life Web Project site for exploring diversity and phylogeny:
http://www.tolweb.org/tree/home.pages/abouttol.html

University of California Berkley, understanding evolution website with excellent pages on phylogenetics:
http://evolution.berkeley.edu/evolibrary/article/phylogenetics_01