

# Laboratory 1: Exploring Biodiversity I - Marine Bacteria

This laboratory is designed to introduce you the diversity of marine bacteria, including some that are still unknown to science. The lab will also demonstrate the value of molecular biology approaches in biodiversity and taxonomy studies. This 3-part laboratory involves studying and isolating a marine bacteria from a coastal Georgia marsh, extracting DNA and sequencing a taxonomically informative gene (the 16S rRNA gene) from one isolate, and using web-based databases to identify the organism (or its closest described relative, if it's new to science!).

Understanding the natural world depends to a large extent on knowing the species that inhabit it. Unfortunately this is no easy matter! The current number of described species worldwide is about 1.4 million, but the actual number is likely to be 30 million or more. Species extinction is occurring at alarming rates (the rate of extinction is currently estimated to be as high as one species per hour) and many organisms may become extinct before we know they are here. Some scientists estimate that several million species will be lost in the first few decades of the 21st century.

Although the challenge of describing biodiversity is great throughout the animal and plant worlds, it is probably greatest among the microbes. Bacteria and Archaea are too small to see with the naked eye, and even when viewed through a microscope, their morphology often isn't distinct enough to provide clues to identity. We think that less than 1% of the existing species of Bacteria and Archaea have currently been described, and it may be much less. Current knowledge of prokaryotic diversity is so poor, in fact, that a significant number of bacteria isolated randomly from natural environments will turn out to be new organisms. In this lab, we'll learn about some of the bacterial species that have already been described, and try to discover a new species in a coastal Georgia salt marsh.

## Discovering a New Bacterium

Modern molecular biology provides a number of tools that permit the exploration of bacterial diversity in natural ecosystems. Most of these new methods are based on sequences of the 16S rRNA gene, a gene encoding a molecule of RNA used in bacterial and archaeal ribosomes. The 16S rRNA gene is approximately 1500 bases in length (Fig. 1), and contains regions that are highly 'conserved' (i.e., have the same sequence in all bacteria and archaea) and highly 'variable' (i.e., have sequences that are unique at the genus or species level). Thus the conserved regions of the gene can be used to bind primers for PCR and sequencing, and the variable regions to determine the identity of the organism. In this lab, you will isolate a bacterial colony from salt marsh seawater or sediments, determine the sequence of its 16S rRNA gene, and use Internet-based 16S rRNA databases to learn whether it is has been isolated previously or whether you have discovered a new organism.

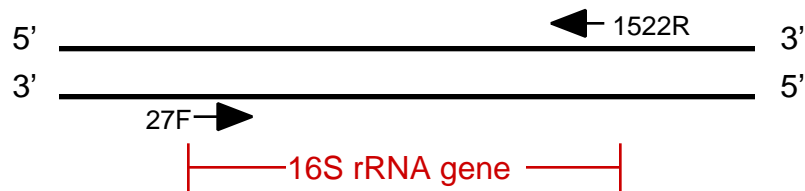


Figure 1. Schematic of the Bacterial and Archaeal 16S rRNA gene (approximately 1500 bp in length) and the PCR primers to conserved regions that are used in gene amplification.

## Isolating Salt Marsh Bacteria

In large groups (4 per lab) select a salt marsh microenvironment to serve as the source of your marine bacterial isolates. For example, depending on available material, you might choose marsh sediments, decaying plant leaves, surfaces of marsh invertebrates, or any other appropriate microhabitat.

### A. Set up a dilution series:

- 1) Set up a series of 10 test tubes containing sterile seawater (each tube contains 9 ml seawater). Label the tubes from " $10^{-1}$ " through " $10^{-10}$ ".
- 2) Inoculate the first tube in the series ( $10^{-1}$ ) with a small amount of material (approximately 1 ml, if available) from your selected microenvironment.
- 3) Vortex to mix the tube contents and then use a 1 ml sterile plastic pipette to transfer 1 ml of the contents of this tube into the  $10^{-2}$  tube.
- 4) Vortex and continue transferring down the tube series until all the tubes have been inoculated. Use a new sterile pipette for each transfer.

### B. Spread plate onto agar:

- 1) Obtain 10 agar plates (made with half strength YTSS<sup>†</sup> agar). Label the plates (on the bottom) to match the dilution series tubes.
- 2) Using a new sterile plastic pipet, remove a 0.5-ml sample from the lowest dilution tube ( $10^{-10}$ ) and deposit it on the surface of the appropriately labeled agar plate.
- 3) Quickly spread the liquid evenly around the plate using a plastic plate spreader. You should use the plate spreader for about 30 seconds to spread the liquid as evenly as possible. Use the plastic lid to shield the plate from air-borne bacteria while you spread.
- 4) Using the same sterile pipette, move up to the next dilution tube and repeat the spreading. Continue until all the dilution tubes are spread. Cut a strip of parafilm and wrap the edges of the spread plates to prevent drying.

- 5) Place plates in the 25°C incubator right side up. We will turn the plates over later to prevent problems with condensation on the agar surface, but will wait until the liquid has had time to fully soak into the agar.

*Before you leave the lab:*

- 6) Organize to work with a partner if desired. You may also work alone.
- 7) Record details of the isolation procedure and keep good records of the procedures you use in future analyses. You will need this information to write a lab report later in the semester.
- 8) Demonstrate to the TA your ability to use sterile technique when streaking agar plates. This is a *critical* skill for the next part of the lab, but it's easy enough to master with a little practice. Make sure you understand when you need to come back and why (otherwise you won't be able to continue with the lab next time).

†YTSS agar contains 4 g yeast extract, 2.5 g peptone, 20 g sea salts, and 18 g agar per liter.

## C. Obtain a pure culture:

- 1) Work alone or in pairs for the rest of this lab.
- 2) After 3-4 days, return to the laboratory and examine the plates from the dilution series. Use a sterile plastic loop to pick a single bacterial colony and streak onto a new agar plate (Fig. 2). Your life will be a lot easier if you choose a colony that is well separated from other colonies. Look through all the dilution plates from the series and choose one that has large, well separated colonies.
- 3) Select a second isolate and repeat the streaking process on a second agar plate. Place the dilution plates back in the incubator when you are finished for others in your lab group to use. Place your two plates upside down in the incubator also.
- 4) After 3 days, return to the laboratory and re-streak your selected colonies onto two fresh agar plates. Using sterile loops and following Fig. 2 as before, choose an isolated colony from each plate you streaked in the previous visit and streak it again onto a fresh plate. You will use these plates to obtain cells for PCR amplification in Week 2, so it's essential that you get pure colonies at this stage.

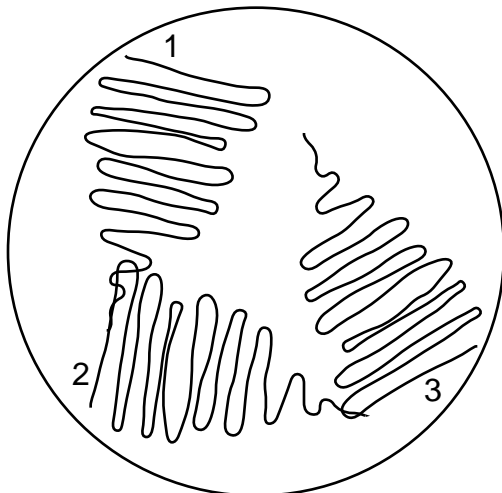


Figure 2. Method for streaking to single colonies. A plastic loop containing cells is touched to the plate at location 1 and the cells are then gradually diluted by streaking the loop back and forth across the plate. A second (sterile) loop is positioned at location 2, passed through the previous streak one time, and then diluted further by streaking back and forth. A third (sterile) loop is positioned at location 3 and the process is repeated.

- 5) When your bacteria have grown up on the second plates and if there is no evidence of contamination, wrap the edges of the plates with parafilm and store in the refrigerator. If there appears to be more than one organism growing on the plates, obtain two more fresh plates and re-streak a third time according to Fig. 2.

## Laboratory 2: Exploring Biodiversity II - Molecular Tools for Diversity Studies

In this second week of the 3-part biodiversity lab, you will amplify the 16S rRNA gene from your two bacterial isolates using PCR, and then submit the product for sequencing.

The polymerase chain reaction (PCR) can be used to amplify regions of DNA using 'primers' that bind to either end of the region of interest. In this lab we will use PCR to amplify the 16S rRNA gene of each bacterial isolate. A pair of short synthetic DNAs (i.e., primers) that can base pair with opposite ends of the 16S rRNA gene (and on opposite strands; see Fig. 1 in Lab 4) will be mixed with DNA samples from your organisms along with dNTPs and a DNA polymerase. Cycles of heating (to make the DNA single-stranded), primer binding, and extension of the primers by DNA polymerase produce thousands of copies of the 16S rRNA gene from your bacterium.

This procedure works well for identifying unknown organisms because the primers are designed to bind to highly conserved regions of the 16S rRNA gene (i.e., the regions that are pretty much the same in all organisms). That means you don't have to know the sequence of the particular 16S rRNA gene you are amplifying.

### A. Extract DNA and set up a PCR reaction:

*(Wear gloves!)*

- 1) For each of your two bacterial isolates, use a sterile plastic loop to scrape a small blob of cells from the agar plate.
- 2) Stir the loop into a microcentrifuge tube containing 500  $\mu$ l of sterile water. Close the tube and vortex very well (about 1 minute or until the cell clump is dispersed).
- 3) Place the tubes into a boiling water bath for 10 min. Use the special tube holder to prevent the tops of the tubes from opening during boiling.
- 4) Remove from the bath and centrifuge the tube contents for 10 min at 13,000 rpm in the microcentrifuge.
- 5) Transfer approximately half the supernatant to clean, labeled microcentrifuge tubes using an automatic pipettor. This will serve as the DNA template for your PCR reaction. If you need to stop at this point, you can store the DNA preparation (i.e., supernatant) in the freezer.
- 6) Set up the PCR reaction:
  - a) Obtain three microcentrifuge tubes containing a bead (of reaction mix and enzyme) in the bottom, and one tube containing sterile water.
  - b) Using a permanent marker, label the top and sides of the tubes containing the beads. Label two with the names of your bacterial isolates and the third as a negative control.

- c) Keep all tubes in the ice bucket as much as possible. To the tubes with a reaction bead in the bottom add, in this order:
  - 23  $\mu\text{l}$  sterile  $\text{dH}_2\text{O}$
  - 1  $\mu\text{l}$  DNA preparation or sterile  $\text{dH}_2\text{O}$  (two tubes get DNA from an isolate; the negative control tube gets sterile  $\text{dH}_2\text{O}$ )
  - 1  $\mu\text{l}$  forward primer (27F) (get this from your TA)
  - 1  $\mu\text{l}$  reverse primer (1522R) (get this from your TA)
- d) Gently vortex the contents of the tubes, and then spin briefly (30 sec) in a microcentrifuge. Keep the tubes in an ice bucket until it is time to load the thermal cycler.
- e) Load the three tubes into the thermal cycler when the other groups are ready. The thermal cycler will cycle through three temperatures:  $94^\circ\text{C}$  to melt the double stranded DNA into single strands,  $55^\circ\text{C}$  to allow the primers to bind to complementary regions on the bacterial 16S rRNA gene, and  $72^\circ$  to allow *Taq* polymerase to fill in nucleotides between primer binding sites. After 35 temperature cycles (which takes about 3 hours), the machine will cool to  $4^\circ\text{C}$ .
- f) When the reaction is complete, we will move the tube to the refrigerator for storage until you return to run out the PCR reaction on an agarose gel.

## B. Examine PCR product on Agarose Gel

*(Wear gloves; ethidium bromide is a mutagenic compound)*

- 1) An agarose gel has already been prepared, poured, and allowed to solidify. A plastic “comb” was placed in the gel before it solidified to make small wells that can hold liquid.
- 2) Retrieve your three PCR tubes from the refrigerator. Cut a small square of parafilm and place it on the bench top. Place three small dots containing 1  $\mu\text{l}$  of blue loading dye about 2 cm apart on the parafilm (one dot for each PCR tube).
- 4) Remove 7 microliters of PCR reaction mixture from each tube and add to one of the spots of loading dye. Mix gently with the pipette tip. Use a separate pipette tip for each sample. Put the PCR tubes back in the refrigerator (you will need them later).
- 5) Pull up the loading dye/PCR mixture from one dot with the pipette. With a steady hand, load this liquid into a well in the agarose gel. Repeat for the other dots, using a clean pipette tip and loading each into a different well.
- 6) In a fourth well, load 5  $\mu\text{l}$  of DNA marker (labeled MWIII). This will provide a ladder of DNA fragments of known length, so you will be able to check the size of your PCR product.
- 7) When all the lanes are loaded, the TA will add buffer to the gel box and start the current.
- 8) After 40 minutes, a line of blue bands should be visible about a third of the way down the gel (this is the loading dye). The TA will turn off the current. **Make sure you are wearing gloves!** Remove the gel from the gel box and place on the transilluminator (a light box containing a UV lamp). Close the plastic cover of the box and turn on the light. **Do not expose your eyes or skin to the UV lamp. Wear UV safety glasses and make sure the plastic cover is closed before looking at the transilluminator.**

- 9). Look at the bands in the gel. A compound added to the gel (ethidium bromide) interacts with the DNA to make it visible (as light pink fluorescent bands) under UV illumination. Ideally, you should see one band in each of your bacterial DNA lanes with an approximate size of 1500 bp (Fig. 1) and no band in your negative control.

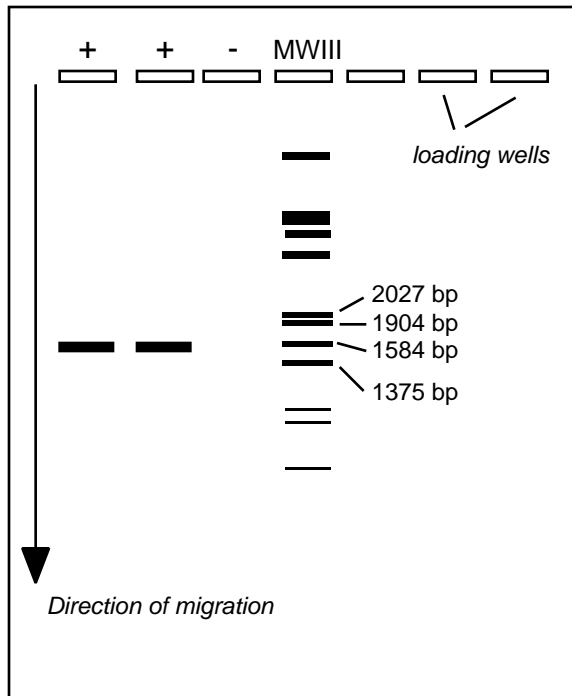


Figure 1. Agarose gel showing ~1500 bp PCR product from the bacterial 16S rRNA gene. The first two lanes (labeled '+') are PCR products from two bacterial isolates. The third lane (labeled '-') is the negative control. The fourth lane (labeled 'MWIII') is the molecular weight marker.

### C. Take your PCR product to MGIF for sequencing

When you have a PCR product of the appropriate size and concentration, take the product to the Molecular Genetics Instrumentation Facility in the Life Sciences Building for DNA sequencing. Choose your isolate with the best PCR product (i.e., strongest band) to continue working with.

- 1) Fill out the "DNA Sequence Request Form" (provided in the laboratory handout) with the required information.
- 2) Take the request form and the tube containing the rest of your PCR product to MGIF in Room A108A of the Life Sciences Building. To keep your PCR product cold during transport, put the tube in a plastic cup filled with ice.
- 3) In approximately 1 week, MGIF will e-mail the DNA sequence to you. Save a copy of the sequence in a text file on a floppy disk or save the e-mail so you can access the file from another computer (if you use Arches, that means leaving it in your Inbox or saving it to a file folder on the Arches server, [imap.arches.uga.edu](mailto:imap.arches.uga.edu)). You will need to have the sequence with you to do Part III of the Biodiversity lab.

## Laboratory 3: Exploring Biodiversity III - Internet Databases

This week (in part three of the Biodiversity lab) you will work with the 16S rRNA sequence of your bacterium using databases and programs available on the Internet. This exercise will tell you whether your isolate corresponds to a previously characterized species, or whether it is something new. If it is new, you can determine which organisms are its closest relatives, and perhaps predict some of the physiological properties of your unknown based on what is known about its relatives.

### BLAST search at GenBank

GenBank is the National Institutes of Health (NIH) genetic sequence database, a collection of all publicly available DNA sequences. Scientists who determine the sequences of genes (or sometimes of entire genomes) deposit the sequence information in GenBank. Aside from doing this as a service to the scientific community, many journals require submission of new sequence data to GenBank as a precondition to publishing a manuscript referring to the sequence. GenBank has approximately 13,543,000,000 bases in 12,814,000 sequence records as of August 2001, derived from viruses, bacteria, and complex eukaryotic organisms.

Included among these many sequences are a number of 16S rRNA gene sequences from Bacteria and Archaea. The current number of 16S rRNA sequences is approximately 12,000. Some of these are 'complete' (they cover the entire 1500+ bases of the 16S rRNA gene), while other are 'partial' (like yours will be) and cover only a portion of the gene (your sequence includes only ~400 bases near the beginning of the gene).

You will use the internet server at the National Center for Biotechnology Information (NCBI) for sequence searches. A sophisticated computer program called BLAST compares the query sequence you submit to all of the sequences previously deposited in the GenBank database. BLAST finds sequences in the database that are most similar to the submitted sequence, and then returns the names and a brief description of the matching sequences in the output from the program.

- 1) Open the text file that contains your sequence and trim off the beginning and end of the sequence to remove most of the "N"s. These Ns represent bases that could not be properly identified by the DNA sequencer, and are typically clustered at the ends of the sequence. By trimming the first and last ~25 bases, you can get rid of most of them. It's OK to have a few Ns in the rest of the sequence, but a large number of them indicates that your sequence is not very good.
- 2) Using a web browser, go to the following URL: <http://www.ncbi.nlm.nih.gov/BLAST/>.
- 3) On the web page that opens, choose the first link ("Standard nucleotide-nucleotide BLAST [blastn]").
- 4) Back in the browser window, the web page (shown in Fig. 1) provides a box into which you should paste your trimmed 16S rRNA sequence. You may first need to click inside the box in order to activate it. There are a number of other boxes and pull-down menus

that allow you to make choices about the database to be searched and to specify formats. We will use all the default settings, so you do not need to make any changes. Just click on "BLAST!" to start the query.

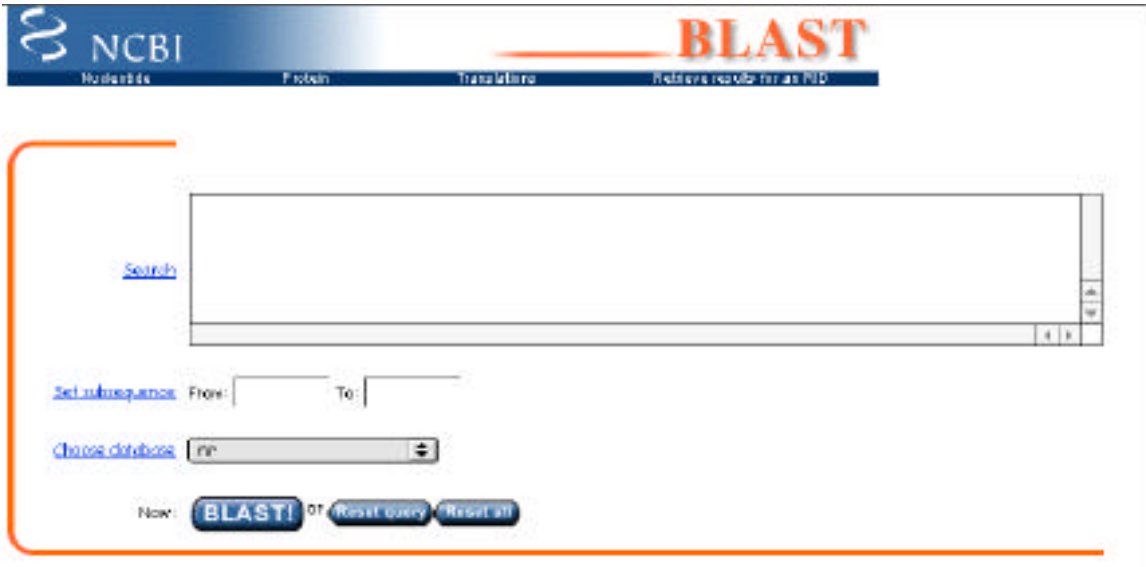


Figure 1. View of NCBI BLAST page. Paste your sequence into the top box labeled 'Search'. Start the query by clicking the 'BLAST!' button.

- 5) The next web page tells you that your query has been submitted and provides options for formatting the results. We'll take the default values again, so you do not need to change any settings. Click on the "Format!" button. The next page will show the results when they are ready. Typically, it takes less than 2 minutes for the query to be completed.

Once the results of your database search are returned you will see that they are presented in three different formats:

A. **The Graph.** The first format (under the heading "Distribution of Blast Hits on the Query Sequence") is a graphical representation of the results, showing the matches to your input sequence as a stack of colored bars with the best matches at the top. The color of the bar indicates the match score (i.e., degree of similarity) between your sequence and the database matches. Red indicates the best matches, black indicates the poorest. The length of the bar shows the length of the sequence that was noticeably similar to the query.

B. **The Sequence List.** Below the graphical output is a list of the top 100 sequences that matched your query, giving their reference number in the database, a very brief description, a score (the higher the score, the more similar the sequences), and an E value (the odds that the similarity between the query sequence and the match might have occurred purely by chance; the lower the score, the more similar the sequences).

C. **The Alignments.** Finally, the third data format provides a base-to-base alignment of the query sequence with each of its matches, starting from the "best match" and progressing down the list. This format repeats the alignment score and E value given above, but also has a measurement called "Identities". The identity give the number of exact matches between the your sequence and the sequence from GenBank and is presented as a fraction (the

number of positions in the sequences that matched divided by the number of positions that the two sequences had in common) and as a percent.

- 6) Scroll down to the alignments section and look at the identities of the best match (the first one listed). Did you get a 100% match? Look down the alignment list and record the names of the 5 best matches, including the number of positions compared and the percent match.
- 7) Follow the link to the description of the top 5 matches by clicking on the accession number (listed to the left of the organism name). This page shows you the GenBank entries for the matches, and provides information on who entered those sequences and where they came from. Record information about where each sequence was found, whether it came from an isolate or a clone (i.e., a sequence that was obtained without isolating the organism), whether or not it is marine, whether anything is known about its physiology, and anything else that looks interesting. Because each researcher enters his/her own sequences into GenBank, there is a lot of variation in the type and amount of detail given. In some cases, there may be a reference to a published paper that will give you more information.
- 8) Many of the entries in GenBank are not described species. Instead, they are listed as an 'unidentified bacterium' or a 'uncultured bacterium'. Scroll down through the list of matches to find the ten closest relatives that actually have species names (you may already have up to 5 from the step above). As above, follow the link to the actual GenBank entry and collect information on where and how these related bacteria were isolated as well as anything that is known about their physiology. Record the "identities" of the sequences (i.e. number of matching positions and percent match).

## Phylogenetic Tree Building at RDP

The Ribosomal Database Project (RDP) (run by the Center for Microbial Ecology at Michigan State University) hosts a web site dedicated to the analysis of bacterial 16S rRNA sequences. You will use the tools at the RDP site to build a phylogenetic tree showing the relationship of your bacterium to some better-known and better-studied bacteria. The tree will group 16S rRNA genes according to sequence similarity, providing information on how your bacterium fits into the existing taxonomic framework for prokaryotes.

- 1) Use a browser to access the RDP analysis page (<http://rdp.cme.msu.edu/html/analyses.html>)
- 2) Find the 'Phylip Interface' row and click on the 'run' triangle under the Small Subunit heading.
- 3) You are now on the 'Start' page for the Phylip analysis tools. You may want to record the session number shown on this page, since the server will save the number for several weeks, allowing you to return at another time to continue with your session.
- 4) Choose 'Edit Data Set' from the top row of buttons. This new page allows you to choose the 16S rRNA sequences that will be included in your phylogenetic tree. The first sequence to add is your own. In the "Upload and Align Sequences" box (Fig. 2), choose "Browse" and select the text file (make sure it is a text file, not a word processing file) that contains your sequence. Now change the third button to read

"Include 3 Neighbors" and then click on the "Upload and Align" button. This brings your sequence into RDP and formats it for tree building.

RDP Phylip Interface : Edit Data Set

Start Edit Data Set Distance Matrix Phylogenetic Tree Help

---

Upload and Align Sequences

Browse... Upload and Align Include no neighbors.

- You may upload files containing sequence data in FastA, GenBank, Phylip, or raw text format.
- The sequences will be aligned to the RDP data as they are uploaded. You may upload as many files (one at a time) as you like, but please note that a maximum of 100 sequences (including templates) will be used for creating a Distance Matrix.
- Long sequence names (over 10 characters) will be trimmed to 10 characters and made unique, so for your own convenience you should limit your names to 10 characters.

RDP Sequences

Select RDP Sequences

- Any template sequences from the RDP database that are included when your sequences are aligned will be displayed here.
- You may also use the Hierarchy Browser tool to include more RDP sequences in your Distance Matrix and Phylogenetic Tree.

Figure 2. View of the 'Edit Data Set' page at RDP.

- 5) Next you will add sequences of six organisms to the tree that come from a variety of bacterial groups. These six organisms are listed in Table 1. In the 'RDP sequences' box, click the 'Select RDP Sequences' button to enter into the lists of sequences that are available to add to your tree. The lists are arranged hierarchically according to phylogeny. Use Table 1 to find your way through the sequence lists. You can **search** using key words for the species or strains given in Table 1 (the fastest way) or you can **browse** the lists of organisms. As you find each of the organisms in the hierarchical lists, click on 'Add' to include the sequence in your tree. If you make a mistake, 'Del' deletes a sequence from your list. When you have found and added all six sequences, go to the top of the page and click on the button labeled 'Click here when finished selecting sequences' to return to the "Edit Data Set' page.
- 6) Next you will add two or three close relatives of your bacterium. If possible, add some of the top matches from your GenBank search. To do this, click again on the 'Select RDP Sequences' button and choose 'Search' to query RDP with the names of the close relatives you got from your BLAST search. If these sequences aren't in RDP, the three neighbors you included when you uploaded your sequence into RDP can be used instead (these should already be listed in the 'RDP Sequences' box). Now check the list of sequences in the 'RDP Sequences box' to make sure it's correct. If you added a sequence by mistake that you don't want in the tree, simply un-check the box to the left of the sequence name.
- 7) Now choose the 'Distance Matrix' button from the top of the page. This takes you to a tool that will compare the sequences. Although the algorithm the computer uses is complicated, it essentially takes each possible pair of sequences and compares each position in their 16S rRNA genes to see if it has the same base or not; the organisms with the most matches get the highest relatedness values. We'll take all the default

settings for this page, so just click on the 'Calculate Matrix' button near the top left of the page. This generates a table showing relatedness for all pairs of sequences, and these are the values that will be used to make the tree.

- 8) Next click on 'Phylogenetic Tree' from the top row of buttons. In the resulting window, make only one change to the default settings: check the 'show full names for RDP sequences' box. Now click on the 'Calculate Tree' button near the top left of the page. Your phylogenetic tree is generated in the main window.
- 9) Save the tree as a pdf file using the key combination 'control-F' and print a copy.
- 10) Based on the phylogenetic tree you generated and Table 1, make your best guess at assigning your organism to a Division and Subdivision within the Bacteria.



Table 1. Bacterial sequences to be used in constructing the phylogenetic tree.

Domain	Division	Subdivision	Group	Group/Subgroup	Organism
Bacteria	Cyanobacteria & Chloroplasts	Cyanobacteria	Prc.marinus		<i>Prochlorococcus marinus</i> str. SSW5
Bacteria	Gram Positive	Bacillus-Lactobacillus-Streptococcus	B.subtilis	B.subtilis	<i>Bacillus subtilis</i> str. 168
Bacteria	Proteobacteria	Alpha	Sphingomonas	Erb.longus	<i>Erythrobacter longus</i> str. Och101
Bacteria	Proteobacteria	Alpha	Rhizobium-Agrobacterium	Sta.stellulata	<i>Stappia stellulata</i> IAM 12621
Bacteria	Proteobacteria	Gamma	Pseudomonas and relatives	Ps.putida	<i>Pseudomonas putida</i> str. mt-2
Bacteria	Proteobacteria	Gamma	Enterics and relatives	Escherichia	<i>Escherichia coli</i> str. K-12

## Sequence Submission to SIMO and GenBank

This section provides instructions for submitting your 16S rRNA sequence to GenBank. By adding your new sequence to the existing databases, scientists that discover similar 16S rRNA sequences in the future will be able to find out about your bacterium, as well as the habitat and environmental conditions under which your bacterium was isolated. You will first enter your isolate into the Sapelo Island Microbial Observatory (SIMO) database. This database is run by scientists at UGA who are building a repository for isolates obtained from the Sapelo Island salt marshes. The National Science Foundation is funding the project (you can find out more by visiting the SIMO web site at <http://simo.marsci.uga.edu>). After you obtain a SIMO database number for your isolate, you will use this number to identify your sequence as a Sapelo Island bacterium when you submit your 16S rRNA sequence to GenBank.

- 1) Get a Sapelo Island Microbial Observatory isolate identification number for your bacterium. Your TA needs to help you do this.
- 2) After obtaining a SIMO isolate number (it will be in the form SIMO IS-Sxx-xxx) use a web browser to go to the GenBank entry page at <http://www.ncbi.nlm.nih.gov/BankIt>.
- 3) Scroll down the page until you see the section shown in Fig. 3. Enter the number of nucleotides in your trimmed sequence in the box and click the 'New' button.

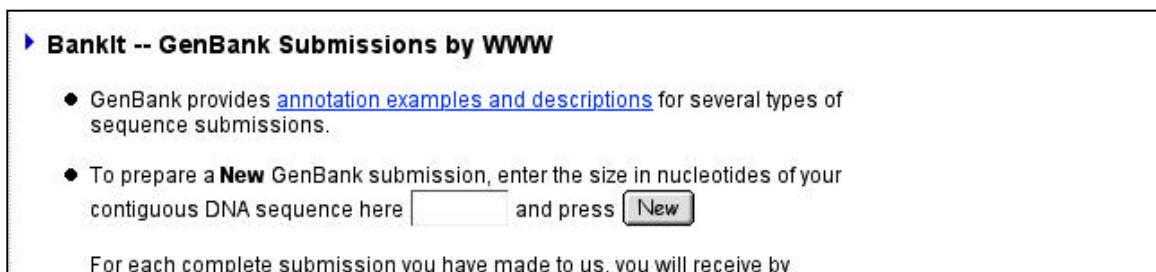


Figure 3. Portion of the GenBank submission page showing how to enter the BankIt page.

- 4) The web page that appears next is a BankIt entry form that assigns a unique number to your entry. Scroll down through the form, entering information exactly as shown in the appendix.
- 5) After you've finished filling in the BankIt form, click the "Validate and Continue" button at the bottom.
- 6) A series of pages will appear asking you to check for possible errors. Ignore the warning message about not providing a recognized organism name (your isolate doesn't have a one); just scroll to the bottom of the form and click 'Validate and Continue' again. When asked if you want to modify your submission regarding coding regions or other features, don't make any changes; select 'Submit to Genbank'. And when asked if you forgot to indicate a coding sequence interval in your submission, say you didn't (your gene doesn't code for a protein). Read these error pages carefully to make sure your submission isn't cancelled without you knowing it.

- 7) When you get to the final “Thank you for using Bankit” page (see what it looks like in the Appendix), your submission is complete. GenBank will contact you by e-mail to:
  - 1) give you a copy of your submission, 2) give you the GenBank submission number assigned to your sequence, and 3) show you your final submission after GenBank staff have checked/corrected the format.

# Exploring Biodiversity Laboratory Report

Prepare a written laboratory report summarizing your activities during the three-part Biodiversity lab. The lab report must be typewritten and should be about 3-5 pages in length. If you worked with a partner during the lab, do not write the lab report together...it must be entirely your own work. Include the following sections in your report:

*Introduction:* What was the reason for carrying out this study? You might choose to focus your introduction on topics such as global biodiversity issues, or the role of bacteria in salt marshes, or any other framework that you think is relevant to the lab. There is no single right reason...you just need to choose something reasonable and present it well. Include at least 3 outside references (books, scientific papers), 2 of which must be different from the papers provided to you in lab. (approximately 1 page)

*Bacterial Isolation and Sequencing:* Describe how you obtained your bacterial isolate, and provide information on the laboratory methods you used to obtain a sequence of the 16S rRNA gene.

*Sequence Analysis:* Describe what you learned about your organism and its relatives from the BLAST analysis and tree building exercise. Provide GenBank accession numbers and SIMO submission numbers for your sequence.