Molecular Phylogeny
Using DNA Sequence to Unravel Evolutionary Relationships

OVERVIEW

Each person will choose two isolated colonies that have been obtained from soil spread plates. We will use the Polymerase Chain Reaction (PCR) to amplify the 16S rRNA genes from these bacterial isolates. Gel electrophoresis will be used to determine if PCR reactions were successful and then we will purify and prepare the PCR products for DNA sequencing. DNA sequencing will be carried out at the Cornell Biological Resource Center. On the final day of the exercise we will use online tools to identify our isolates by conducting a phylogenetic analysis.

Many commercial and university facilities now accept samples through the mail for DNA sequencing at reasonable costs ($2-5$ per DNA sequence run, about 600-900 base pairs). Access to a thermocycler (cost > $1,500) and basic molecular biology equipment are required to complete the 'wet lab' portion of this exercise but other reagents and supplies are relatively easy to obtain.

It should be noted that the tools required for analyzing DNA sequence data and conducting a phylogenetic analysis are freely available over the web and that 16S rRNA gene sequences can be downloaded from Public databases (ie: Genbank). Thus, this exercise can be carried out starting on 'Day 3' by giving students 'unknown' DNA sequences that have been downloaded from Genbank and stripped of their identifying information prior to the start of the exercise. The only equipment required to carry out this option would be computers equipped with web browsers and internet access.

EQUIPMENT NEEDED

- Thermocycler
- Microwave, or hot plate with a magnetic stirrer for dissolving agarose (at least 1 per class)
- Gel rig (chamber top, bottom, tray and comb) (1 per 6 students)
- power supply (1 per four gel rigs)
- photographic apparatus (1 per class)
- 20 µl micropipettors (1 per pair)
- **FOR THE TEACHER ONLY:** One 200-µl micropipettor to use in aliquotting some of the reagents. **PLEASE NOTE:** This pipettor looks very much like the 20-µl pipettors. Keep this one separate!

MATERIALS

- Bacterial isolates
- 6 'PCR bead tubes' per group of 2
- 1 tube of positive control DNA per group of 2
- 1 tube of negative control water per group of 2
- 1-2 tube of primer mix (containing 0.3 uM of primers Bact8F and Bact 1492R) per student
- 1-2 tubes of 10 ul 1% Tween 20 (a non-ionic detergent) per student
- 5X loading dye (60 µl per pair - provided in DNA Profiling Reagent Kit)
- agarose (1g per 3 pairs)
- 1X TBE (gel running buffer) (1 liter per 2 pairs)
- Carolina Blu® DNA Stain (250 ml per pair) or GelStar® DNA stain (2 µl per pair) (if you are using GelStar® stain, you will also need a UV light (which comes with the kit) or a DarkReader®, (available separately from CIBT).
- type 667 Polaroid® film (available from Carolina Biological Supply Company 21-3679, also may be available at your local camera shop; 1 exposure per student required) Alternatively, you can use a digital camera.
- pipette tips (1 box small for P20 and P200 pipetmans per pair)
- staining trays (1 per 2 pairs)
- latex gloves

EXERCISE OUTLINE

Day 1 PCR amplification of 16S ribosomal RNA genes from bacterial isolates

Day 2 Gel electrophoresis of PCR products and preperation for DNA sequencing

Day 3 Conduct phylogenetic analysis of 16S rRNA genes

DAY 1: PCR AMPLIFICATION OF 16S RIBOSOMAL RNA GENES

The following procedure can be used with either whole cells from a colony/culture or with purified DNA. We will be amplifying the 16S ribosomal RNA genes (16S rRNA) from two isolates. Each person will choose two soil isolates provided by the instructor. Every PCR experiment should also include a ***negative template control*** (to which no colonies or DNA is added) and a ***positive template control*** (DNA which is known to amplify with your primer set). In our exercise we will perform one positive control and one negative control for each group of two people. The negative control reaction will tell you if there is DNA contaminating any of your PCR reagents. You should always include a negative control when performing PCR. The positive control reaction is a check to make sure all components of the reaction were present and that the thermal cycling program was appropriate. Because of the sensitivity of PCR it is best to wear gloves at all times in order to avoid contaminating your sample. Be sure to **ALWAYS USE A NEW PIPETTE TIP WHEN GOING INTO ANY VIAL** to prevent cross contamination of reagents and samples.

Materials needed by each group:

Gloves
100 ul pipettor & tips
6 'PCR bead tubes'
1 tube of positive control DNA
1 tube of negative control water
1 tube of primer mix (containing 0.3 uM of primers Bact8F and Bact 1492R)
4 soil isolates
4 tubes of 10 ul 1% Tween 20 (a non-ionic detergent)

Procedure:

1) Use a sterile inoculating loop or pipette to pick one large colony from one of your isolates and transfer this colony material into one of the tubes of Tween 20 (be sure that material sticks to your tip and that sufficient material is transferred – the quantity of material should be a little bigger then the head of a pin). Repeat with your other isolates being sure to label the tubes well. Each person should do 2 isolates. Mix the tubes well by pipetting up and down and then incubate 10 minutes at 95°C. This will help to lyse the cells (rupture their cell walls and membranes).

2) Label your 'PCR bead tubes'. Add 24 ul of the primer mix to each of your tubes. Each person should have two isolates and each group of two should share one positive control and one negative control tube.

3) Add 1 ul from each tube of your lysed cells from step 1 to a 'PCR bead tube'.

4) Add 1ul of positive control DNA to your positive control 'PCR bead tube'.

5) Add 1ul of negative control water to your negative control 'PCR bead tube'.

6) Make sure that the thermocycler has been programmed with the correct PCR profile by the instructor (see below) and begin your PCR reaction. The reactions will run overnight and we will analyze the PCR products in the morning.

The PCR thermal cycle profile:

# cycles	temp.	time.	Step.
1	95°C	5 min.	Initial Denaturation
25	95°C	30 sec.	Denaturation of template
	55°C	30 sec.	Annealing of primers
	72°C	1 min.	Extension
1	4°C	∞	Hold

DAY 2: GEL ELECTROPHORESIS OF PCR PRODUCTS AND PREPERATION FOR DNA SEQUENCING

AGAROSE GEL ELECTROPHORESIS

Agarose is a complex polysaccharide matrix derived from seaweed that permits rapid separation of large macromolecules such as DNA. DNA is pulled through the agarose matrix on the basis of charge, size, and state of supercoiling. Smaller molecules travel faster through the agarose matrix than larger molecules. The negatively charged phosphate backbone of DNA migrates towards the positively charged cathode of the electrophoresis chamber. The buffer in the electrophoresis chamber contains salts that provide conductance. The percentage of the gel also affects the migration of DNA; higher concentrations separate small molecules, lower concentrations separate large molecules. Each gel should be run with DNA size markers on each side of the experimental samples. The DNA size markers consist of double stranded DNA fragments of known size that form a "ladder" and provide a point of reference.

Preparation of Agarose Gel

1. The group that is to prepare the 1X TBE buffer will need to add 200 ml of distilled water to the bottle of TBE powder (USB #70454). **Mix thoroughly.** This is now a 10X concentration of the TBE buffer. Pour the contents of the bottle into a 2-liter beaker or Erlenmeyer flask. Carefully add 1800 ml of distilled water and mix until fully dissolved. This is enough buffer for four gels. If only 1 liter of 1X TBE is to be prepared, then use 100 ml of the 10X concentrate and 900 ml of distilled water. Share the extra buffer with another group of students.

2. Every 3 groups (6 students) will need to prepare and share 1 gel.

    a. Weigh out 1 grams of agarose and add it to 100 ml of 1X TBE buffer in a 250-ml Erlenmeyer flask. You can prepare more than one gel at a time by scaling up the recipe (2 grams in 200 ml, etc.).

    b. Use a microwave to dissolve the agarose. Microwave on high for 1 minute, then swirl the flasks. If the agarose is not completely dissolved, microwave for a few seconds longer until it is (the chunks of agarose will become translucent before they are dissolved. Hold your flask up to the light to see that all of the agarose is completely dissolved and no translucent flecks persist). (Alternatively, if a microwave is not available, place a magnetic stir bar into the flask and heat the flask on a stirring hot plate until the agarose particles are fully dissolved. The solution will have to boil for several minutes for this to occur. If a magnetic hot plate isn't available, heat the materials in a 250 ml or 500 ml beaker and **stir frequently**.)

c. Once the agarose is fully dissolved remove the flask from the hotplate, insert a thermometer, and allow the solution to cool to 70°C.

3 a. While the agarose is cooling, the next step it to prepare the casting tray. At the moment, CIBT has two types of gel apparatus in our lending librairy, ones that have not gasket on the casting trays and those that do. Examine your casting tray.

- If the casting tray has no gasket, place tape across the end of the tray so that it blocks off the open ends of the tray. Place the taped tray on a flat surface (not in the running apparatus).

- If the casting tray has a gasket, place it in the gel running apparatus so that the open ends of the casting tray are facing the sides of the gel running apparatus. The gaskets should make a seal with the sides of the gel box.

b. Pour the cooled solution into the tray to a height of 0.8 to 1 cm (approximate). It may be necessary to tip the tray gently to spread the agarose across the entire surface.

d. After the gel is poured, place the comb so that it fits in the slots that are closest to the end of the gel-casting tray. (There are two sets of slots in the gel casting tray, one close to the end of the tray and one in the middle of the tray. **Do not** use the slots in the middle of the tray.) The wells should be positioned closest to the black end (negative) of the running chamber.

e. Leave the gel undisturbed for 20 minutes until the agarose becomes firm and opaque.

4. a. When the gel has cooled, remove the gel casting gates from the ends of the tray.

b. Add 800 ml of 1x TBE buffer to the gel chamber. The gel should be fully submerged by about 2 - 3 mm of buffer.

c. Remove the comb by gently **pulling it straight upward**. You will load your samples into the slots created by the comb.

Loading of the Gel

1. Before you start be sure that the wells are on the side of the gel box that will connect to the black (negative) electrode. The DNA is negatively charged so it will move towards the red (positive) electrode when the current is applied. Remember DNA 'Runs to Red'.

2. Each pair of students will load 6 of the 20 wells in the gel. In addition to your samples, your teacher will also give you a sample of DNA size standard. One DNA size standard should be run on each gel.

3. Before you start loading your gel, discuss the order you would like to load your samples in with the rest of your group. It is important that you do this before you start loading your gel, since if you take too long when loading the gel your samples will diffuse out of their slots.

4. Use a P20 micropipette to place a spot of 1 μl 5X loading dye onto a

piece of parafilm. Use the P20 micropipette to add 4 ul of one of your PCR reactions to the 1ul spot of loading dye and mix by pipetting up and down several times.

5. Load your sample into the a well of the gel. The loading dye will make your samples dense so they will sink to the bottom of the slots in the agarose gel.

6. Repeat steps 4 and 5 for all of your samples.

7. Place the cover on the apparatus and connect the leads to the power supply. Electrophorese at 185-199 V for ~20-30 minutes or until the blue dye (from the loading dye) is near the bottom of the gel.

Staining and Photographing the Gel (if using Carolina Blu™ DNA stain)

1. Gently transfer the gel to a staining tray and cover with Carolina Blu™ Stain (250 ml). Leave the gel in the stain for 30 minutes. If bands are not visible, stain longer. Often several hours of staining produce the best results (overnight staining produces a dark background, but bands will still be visible).

2. Pour the staining solution back into the bottle, (put a check mark on the bottle, for stain is reusable a limited number of times) and add ~300 ml of distilled water to the gel in the tray. The DNA bands should be visible as the background clears. Change the water two or three times.

3. Scientific results need to be documented in permanent form. A photograph of the gel could be used as evidence in the courtroom (the gel itself would break down with age). Set the camera for *f*22, 1/125 second. Photograph your gel using Polaroid® type 667 film.

   If the photograph is dark change the *f* stop to *f*16. If the photograph is light, change the *f* stop to *f*32.

PREPARATION OF PCR PRODUCTS FOR DNA SEQUENCING

Following PCR we need to remove the PCR reaction components from the DNA that we created before we can sequence the DNA. We will use a commercial kit to carry out this purification. After DNA purification we will mix a portion of our DNA sample with our sequencing primer Univ519F and then submit the samples for sequencing on a 3730 Applied Biosystems DNA analyzer.

1) Read and follow the instructions that come with the DNA purification kit. These instructions are provided in the appendix of this protocol.

2) Use the remainder of your PCR product (~21 ul) from your bacterial isolates with the DNA purification kit. Do not purify the positive control, negative control, or any bacterial isolates that did not have a band when the samples were run on an agarose gel.

3) At the last step of the DNA purification protocol, when you elute your DNA, should be changed so that you add 17 ul of elution buffer rather then the amount suggested in the protocol. Be sure to carefully add the 17ul to the exact center of the resin bed in the column.

4) Add 1 ul of the sequencing primer Univ519F to your purified DNA.

5) You are now ready to give your sample to the instructor so that it can be sent for sequencing.

DAY 3 PHYLOGENETIC ANALYSIS OF 16S rRNA GENES

PHYLOGENIC ANALYSIS OF RAW 16S RRNA SEQUENCES

Raw sequence data (from a sequencing machine) will come as both a chromatogram (which shows the raw data) and as a .txt file which has the DNA base pairs only. You will use the .txt file for your analyses, but some small edits need to be made before you are ready to analyze the sequence. If you were to use sequence files downloaded from Genbank these editing steps would be unneeded (because the researcher that deposited the sequences has already done them).

We will perform the phylogenetic analysis using online tools at the Ribosomal Database Project (RDP). The RDP archives more than 250,000 16S rRNA sequences from every known bacterium on the planet (though there are still a lot of organisms that have yet to be discovered). Before you perform the phylogenetic analysis it would be helpful to review the RDP tutorial that is provided in the *Additional Information* which accompanies this exercise.

Editing the raw sequence data

1) Open the program Chromas (a free program that allows you to view chromatograms and is freely available on the internet). From the file menu in Chromas open the first chromatogram. A chromatogram from a good DNA sequence will have evenly spaced single peaks with low background and very few N's (an N means that the sequencing machine could not determine the identity of the nucleotide at this position). It is normal for there to be bad data at the beginning and at the end of raw DNA sequence. Determine where the good sequence data starts and ends, note number that indicates the sequence position for each of these points (depending on the sequencing method used a typical sequencing reaction will provide 400-800 bases of usable data).

2) Open the sequence text file that corresponds to the chromatogram you just viewed using Microsoft Word. Use the 'Word Count ' function from the 'Tools' menu in Microsoft Word to find the nucleotide positions that correspond to the beginning and end of the good sequence data. Delete the bad data from the beginning and the end of the sequence. Provide a name for the sequence by adding a '>'symbol immediately followed by a name to the beginning of your sequence (as below).

   >sequencename
   ACTGTGCAAGTGGGAACTGTGCAAGTGGGACTGTGCAAGTGGGA
   ACTGTGCAAGTGGGTGTGCAAGTGGGA……

3) Put all of your sequences in a single file by leaving a single blank line between each sequence (see below). Save this new file as a text (.txt) file.

   >sequencename1
   ACTGTGCAAGTGGGAACTGTGCAAGTGGGACTGTGCAAGTGGGA
   ACTGTGCAAGTGGGTGTGCAAGTGGGA……

   >sequencename2
   ACTGTGCAAGTGGGAACTGTGCAAGTGGGACTGTGCAAGTGGGA
   ACTGTGCAAGTGGGTGTGCAAGTGGGA……

Sequence Match

1) Use a web browser to go to the Ribosomal Database Project Online Analysis Tools (http://rdp8.cme.msu.edu/html/analyses.html). Select 'Sequence Match' from the menu (by clicking on the little arrow in the 'Run' column). Paste your sequence file into the space provided and then submit your file. Record the 5 most similar sequences in the database for each of your sequences along with the S_{ab} values (the first number the occurs after the sequence name, it ranges between 0 and 1 this tells you the % similarity value between your sequence and sequences already in the database). Note that values below 0.5 are generally meaningless.

Phylogenetic analysis

2) Return to the 'Online Analysis' page. Select 'Phylip Interface' from the menu. Record the number for you tree session from the box labeled 'Your Tree Session'.

3) Now click the button labeled 'Edit Sequence data'. In the 'Upload and Align Sequences' box click the browse button to choose the file containing your DNA sequence data. Next click the 'Include no. neighbors' tab and select the number 10 (this will put the 10 most similar sequences found by sequence match into your tree). Last go to the 'RDP Sequences' box and click the button to choose any additional sequences to use in your analysis. For this exercise find and select the sequence for '*Mt Fervid1'* (Methanothermus fervidus, 1.1.2.1) by finding it in the tree and then clicking 'add'. When you have selected *Mt Fervid1* you will click the button 'Click here when finished…' at the top of the page to return to the Edit Sequence Data page. Review the results in the alignment window. The green areas represent your sequence and the white areas represent where sequence information is missing. At least 1/3 of the green bars should overlap across all sequences you are analyzing for you to be able to make a good tree.

4) Next click the 'Distance Matrix' button to generate a similarity matrix, use the default options and calculate your matrix by clicking the button 'Calculate Matrix'. Finally click the 'Phylogenetic Tree' button. On the 'Phylogenetic Tree' page click the 'Outgroup' tab in the 'Options' menu at the lower left of your window. Select '*Mt Fervid1'* as your outgroup (an outgroup is a distant ancestor and it is useful to use an outgroup to understand the relationships among the sequences you are analyzing. *Mt Fervid1'* is an Archaea and since all of our isolates are likely to be Bacteria it will serve as an acceptable outgroup. Click 'Calculate Tree' to calculate your tree.

5) View your tree. You can refine your analysis by examining the tree to find the sequences most closely related to your sequence of interest. Return to step 3 and repeat your analysis but this time choose add additional sequences to your tree using the 'RDP Sequences' button in the 'Edit Sequences' page. From the RDP tree choose sequences that are similar to those you identified in the first tree you calculated (for example if your sequence seems to cluster with a number of Pseudomonas species then to to the Gamma Proteobacteria and select a lot of Pseudomonas sepecies to add to your tree). It is good to select a range of sequences from groups that are closely related to your sequence in order to build your tree. Be aware that the RDP contains a lot of incomplete sequences and that the green bars representing the sequences you choose need to overlap by at least 1/3 of the total size of the bar in the alignment window (this corresponds to about 500 bp of overlap) in order for you to get a meaningful tree. In general try to avoid incomplete sequences when selecting sequences to add to your tree. Calculate a second tree using these new sequences and compare it to your first tree.

6) After you have made this second tree you can repeat the tree building process using the same set of sequences and varying the options you choose for building your similarity matrix. If your tree is robust then it is likely to look similar regardless of the options you choose (small changes in tree topology are not unusual, but you should be suspicious of any large changes). Make sure you record the options that you choose for each tree that you constructed.