MOLECULAR SEQUENCE-BASED IDENTIFICATION

INTRODUCTION

The traditional approach to identifying bacterial strains is based largely on growth-dependent physiological and biochemical tests that have been developed since the beginning of the 20th Century, and are still widely used in clinical laboratories. We perform a number of these classic diagnostic test methods in a separate exercise. These traditional methods are being augmented, and in some cases supplanted, by molecular sequencing methods.

You and your lab partner will attempt to identify 2 different bacterial strains by submitting a 16S ribosomal DNA sequence to an online database for comparison with known sequences. The 2 strains include the LAB strains that you isolated, plus a "Known" strain that I provide (E. coli or S. epidermis). The project will be separated into 6 phases that you complete on different days:

I. Genomic DNA Isolation
   You isolate and purify high molecular weight (HMW) DNA from overnight cultures of an enteric unknown and from your fecal isolate strains.

II. Analysis of the Genomic DNA Preparation
   Spectrophotometric and Agarose Gel Electrophoresis Analysis

III. PCR Amplification
   You use the Polymerase Chain Reaction to amplify a variable region of the 16S rRNA gene.

IV. Cleanup PCR Product

V. Analysis PCR Products
   Analysis by absorption spectrophotometry and agarose gel electrophoresis before incurring the expense of sequencing.

VI. Analysis of DNA Sequence
   The product of the PCR amplification in this exercise will be "outsourced" to a dedicated facility for automated DNA sequencing. Results are returned to us by email.

Your returned sequences are used to search the NCBI DNA sequence database for similar sequences.
PCR AMPLIFICATION OF 16S rDNA

Familiarize yourself with the basics of the Polymerase Chain Reaction:

http://en.wikipedia.org/wiki/Polymerase_chain_reaction
http://www.youtube.com/watch?v=_YgXcJ4n-kQ&playnext_from=TL&videos=RT4-xCAd87I

Standard PCR amplification reactions have the following components:

• Thermostable DNA Polymerase (usually Taq polymerase)
• 4 deoxynucleotide triphosphates (dNTPs)
• Buffer at appropriate pH, ionic strength, and [Mg++]
• Oligonucleotide Primers (usually synthetic)
• Template DNA

All of the reagents except the template DNA are ordinarily purchased from biotech companies. The template DNA is your genomic DNA prep.

The primers are synthetic oligonucleotides made to order by a company specializing in custom DNA synthesis. The primer sequences we use ("27F" and "519R") hybridize, in opposite orientations (and to opposite DNA strands) of highly conserved regions in the 16S rRNA gene of virtually all organisms in the Domain Bacteria. The primer sequences are given below. Notice that there are three positions that are "degenerate". For example, 27F is actually 2 different primers, one with C, the other with A, at the position indicated by "M". This is necessary to recognize exactly, all Bacterial 16S rRNA genes.

27F: 5' AGA GTT TGA TC(M) TGG CTC AG 3' [=20 nt]

519R: 5' G(W)A TTA CC G CGG C(K)G CTG 3' [=18 nt]

M = C/A
W = A/T
K = T/G

The region of the 16S gene between the primer binding sites will be amplified and then sequenced. This region is sufficiently variable that any 2 given species will probably have sequences that differ by at least one nucleotide. On the other hand, the region is conserved enough that different strains of the same species are usually identical.
Procedure

The amplification reaction requires 2-3 hours (see below). Once the amplification is started, everything is automatic and we will turn our attention to elsewhere.

The reaction setup is:

<table>
<thead>
<tr>
<th>Component</th>
<th>Stock Conc.</th>
<th>Volume per Rx.</th>
<th>Final Conc.</th>
<th>Total Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR Master Mix*</td>
<td>2X</td>
<td>12.5 µl</td>
<td>1X</td>
<td></td>
</tr>
<tr>
<td>upstream primer</td>
<td>10µM</td>
<td>2.5 µl</td>
<td>1.0µM</td>
<td></td>
</tr>
<tr>
<td>downstream primer</td>
<td>10µM</td>
<td>2.5 µl</td>
<td>1.0µM</td>
<td></td>
</tr>
<tr>
<td>Nuclease-Free Water</td>
<td></td>
<td>2.5 µl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA template</td>
<td>&lt;50ng/µL</td>
<td>5 µl</td>
<td>&lt;10ng/µl</td>
<td>&lt;250ng</td>
</tr>
</tbody>
</table>

*The PCR Master Mix contains:

- Taq DNA Polymerase 50 U/ml in reaction buffer concentrate
- 400µM each: dATP, dGTP, dCTP, dTTP
- 3mM MgCl₂

We have added all of the reagents to the PCR tubes for you except for the template DNA. You only need to add 5 µl of your template DNAs.

Note: When pipetting small (<50 µl) volumes into a microfuge tube, you must use adhesion to achieve full and accurate delivery of the volume. If you do not understand what this means ask before pipetting anything.

!!! First, dilute your genomic DNA preps to a concentration of 50 ng/µl with di H₂O. Be sure your dilution calculation is correct before proceeding. (If your DNA prep. is <50 ng/µl to begin with, use it as is without further dilution.)

Mix the PCR rx. tubes by gently flicking them with your finger. Then briefly spin them in a centrifuge to collect all the liquid at the bottoms.

Label the tubes so that you can identify them again.

Place your tube in the thermocycler. Record the position of your tubes so that you can find them later.

DNA Amplification parameters programmed in the thermocycler:

- Denaturation 95°C / 1 min
- Annealing 55°C / 1 min.
- Polymerization 73°C / 2 min.

- # Cycles 35

At the end of the amplification we will transfer the tubes to storage at -20°C for you.
Assignment

1. On the 16S rRNA secondary structural diagram (next page) use colored pencils to indicate exactly the regions corresponding to the primer binding sites in the rDNA.

   HINT: The numbers "27" and "519" in the primer designations refer to the nucleotide position in the 16S rRNA sequence that corresponds to the 3' end of the primer.

2. How many nucleotide mismatches are there between the 2 primers and their binding sites on the E. coli 16S rDNA gene?

3. Determine the exact length of the PCR amplification product.

4. In theory, the number of copies of the target DNA sequence doubles in each cycle of the PCR reaction. Calculate the expected number of copies predicted at the end of the PCR reaction for each copy of the target sequence initially present. This may be referred to as the "Amplification Factor". (i.e. if the PCR reaction produces 1,000 amplified copies of copy of the target sequence originally present, the Amplification Factor is 10^3 X.)