Unknown #3: Sequencing

Review the lab handout " **Identification of Unknown #3 by PCR and sequence analysis of 16S ribosomal DNA (rDNA)**" to see how the experiment is performed.

Summary

1. Extract DNA from unknown bacteria by boiling.
2. Set up a PCR reaction to selectively amplify a 779bp (base pair) fragment of the 16S ribosomal RNA gene.
3. Run agarose gel electrophoresis to verify that a DNA fragment corresponding to ~800bp was amplified.
4. Obtain a sequence from that amplified DNA fragment and use it to identify the unknown bacteria.

The ribosomal RNA gene is commonly used for sequencing purposes because it is found in living cells.

The Polymerase Chain Reaction (PCR) requires four ingredients:

1. Template DNA to be amplified
2. DNA polymerase – makes DNA, is thermostabile, does not denature with heat
3. dNTP's, the nucleotides A, T, G, C
4. primers – create region of double-strandedness so DNA pol can hop on and start replicating
   - primers are specific for the region to be amplified (replication/amplification only occurs where the primers bind)

The process:

The reaction mixture is heated up to 95°C to separate the two strands of DNA.
The reaction mixture is cooled down to 58°C to allow primers to find the complimentary DNA sequence and bind to it.
The reaction mixture is heated back up to 72°C to allow the thermophilic DNA polymerase to function and replicate new DNA. 72°C is hot enough to allow the DNA pol to function but not hot enough to separate (denature) the primer from the DNA.

This cycle is repeated 30-50 times to dramatically amplify the sequence of interest. The sequence of interest is also known as the amplicon.

The amplified DNA (amplicon) is run on an agarose gel along with a DNA marker/ladder. The DNA marker/ladder contains a series of fragments of known size (bp). The marker/ladder allows for the estimation of fragment size by comparing experimental fragments with known fragments.

The amplicon is then sequenced using the Sanger-Dideoxy Chain Termination method. There are other techniques, but the Sanger sequencing technique is the most common sequencing technique used. See video: https://www.youtube.com/watch?v=FvHRio1yyhQ

The Sanger sequencing technique requires <u>five</u> ingredients:

The first four ingredients are the same as for PCR.

1. Template DNA to be amplified
2. DNA polymerase – makes DNA, is thermostabile, does not denature with heat
3. dNTP's, the nucleotides A, T, G, C
4. primers – create region of double-strandedness so DNA pol can "hop" on and start replicating
   - primers are specific for the region to be amplified (replication/amplification only occurs where the primers bind)

5. ddNTP's – dideoxy nucleotides. Dideoxy nucleotides lack the hydroxyl group (OH) needed for nucleotides to be added to make DNA. Therefore, when DNA pol randomly adds a complimentary ddNTP, no more nucleotides can be added after that ddNTP. DNA elongation terminates.

The overall process:

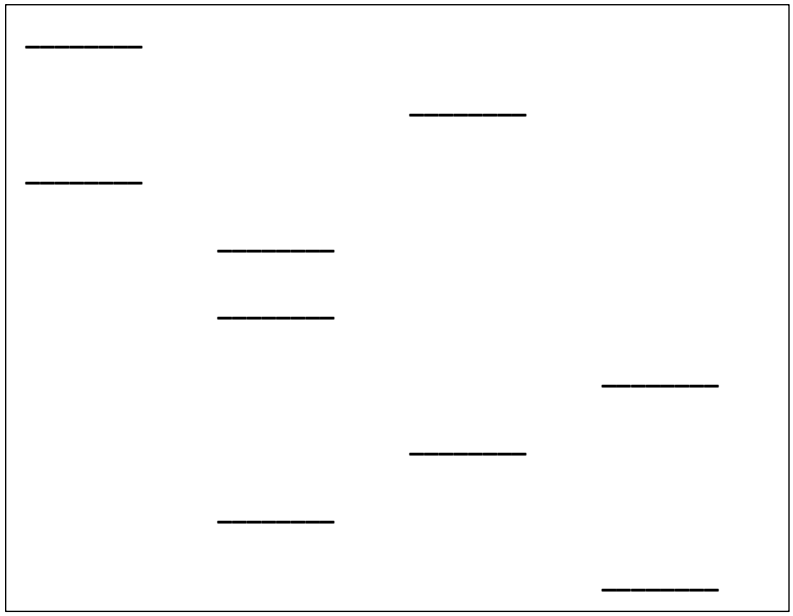
1. Four reaction mixtures are created. Each reaction mixture contains all of the first four ingredients plus a separate ddNTP.

For example: The first reaction mixture contains the four ingredients + ddATP.
The second reaction mixture contains the four ingredients + ddTTP.
The third reaction mixture contains the four ingredients + ddGTP.
The fourth reaction mixture contains the four ingredients + ddCTP.

2. In each reaction mixture, every possible size fragment will be generated. Each fragment terminates with a ddNTP.
3. Each reaction mixture is loaded into a polyacrylamide gel. Polyacrylamide is far more sensitive than agarose. Polyacrylamide gels can separate fragments that differ in size by **only one base pair!**
4. A gel separating hundreds or more of fragments is produced. Each fragment ends with a ddNTP that terminated that fragment.
5. The gel is read from bottom to top (see video). Each fragment corresponds to the nucleotide in that exact position in the sequence.

For example: **ddATP** **ddTTP** **ddGTP** **ddCTP**

Polyacrylamide Gel



The sequence in this gel is:
CTGCTTAGA

The fragment that migrated the fastest is the smallest.

The fragment that migrated the slowest is the largest.

From bottom to top, the sequences are:

-**C**
-C**T**
-CT**G**
-CTG**C**
-CTGC**T**
-CTGCT**T**
-CTGCTT**A**
-CTGCTTA**G**
-CTGCTTAG**A**

The last nucleotide in each sequence is a ddNTP that terminated elongation. The ddNTP tells us the nucleotide that goes in that position in the DNA sequence.

Finally:
Once the DNA sequence has been generated, the sequence can be used to identify the unknown bacteria using either of the two databases shown in the lab handout.