Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Section #: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

 11 • Sequence and Function Based Metagenomics

**INTRODUCTION**

***Summary***

Thus far, you have collected your sample, filtered your microbes, extracted your DNA, and performed PCR on it. Now it is time to analyze the data from your microbes. To do this, you must sequence the DNA, determine where the genes are located on the DNA, and ascertain the functional potential of the genes. Furthermore, it will also be important to know what the products of the genes do. You’ll be exploring the answers to these types of questions today as you learn about sequence-based and function-based metagenomics and their applications. During this lab you’ll also begin experiments to test your hypothesis about how a specific antibiotic used for human medicine or livestock could affect antibiotic resistance of microbes in the Mississippi River.

***Sequenced-based and Function-based Metagenomics***

There are two basic types of metagenomics studies: Sequence-based metagenomics and function based metagenomics. **Sequence-based metagenomics** involves sequencing and analysis of DNA from environmental samples. Sequence-based metagenomics studies can be used to assemble genomes, identify genes, find complete metabolic pathways, and compare organisms of different communities. Genome assembly requires lots of computer power but it can lead to a better understanding of how certain genes help organisms survive in a particular environment. For example, you might find a certain species collected from an iron mine has a number of different genes involved in iron metabolism. Sequence-based metagenomics can also be used to establish the degree of diversity and the number of different bacterial species existing in a particular sample. Analyzing microbial diversity is less costly and less computer intensive than assembling genomes and it can provide valuable information about the ecology of microbes in a sample. In contrast, **function-based metagenomics** involves screening for a particular function or activity. It has already been used to identify novel antibiotics, and proteins involved in antibiotic resistance, vitamin production, and pollutant degradation1. Many proteins and enzymes with useful functions exist within microbes. Recall that the Taq enzyme (a thermostable DNA polymerase) you used to perform PCR was originally isolated from the *Thermus aquaticus* bacteria in Yellowstone National Park. When using a fosmid library for functional metagenomics, you might, for example, screen a sample for the presence of genes that encode antibiotic resistance, or you might look for genes which allow bacteria to degrade StyrofoamTM. Such analysis allows scientists to identify a tremendous amount of genetic diversity without knowing anything about the sequence of the gene, the structure of the protein, or the microbe of origin. Each of these methods has advantages and disadvantages. Sequence-based screens can give you information about the microbial ecology and genomes of a particular environmental sample. However, if you find interesting genes in this type of screen you don’t know whether they are functional or what they do. A functional screen detects genes which produce functional gene products, but not much is revealed about what species the genetic material came from.

***Sequencing the Metagenome***

It is sometimes desirable to sequence metagenomes from different environments to identify organisms, genes, or metabolic pathways that are present. For all but a few highly prevalent species, it is impossible to sequence an entire genome2. Due to the length of bacterial genomes, ~0.5-10 Megabase pairs (Mbp), it is impossible to sequence the genome from start to finish in one reaction. Instead small pieces called **reads** must first be sequenced. (The terms ‘sequences’, ‘reads’, or ‘sequence reads’ are often used interchangeably). If a genome was a large book, then the individual sequence reads would be like words or sentences. Unfortunately, these smaller pieces of the genome are sequenced in random order so we have to use a computer to assemble all the pieces before we can get the full story. Adding to the complexity is the fact that environmental samples contain many different genomes. Reassembling all these genomes is a task comparable to piecing together a library of books from individual words and sentences.

Sequences are 20-700 bp in size depending on the sequencing method2. Sequencing technology is advancing very quickly and new methods are constantly being developed. In fact, sequencing technologies are advancing so fast that our rate of acquiring new sequences is rapidly outpacing our ability to store them3. Sequencing methods are broadly classified as first-generation, or second-generation technologies.

**First-generation sequencing** technologies rely on the **Sanger method** of sequencing. Sanger sequencing is very similar to a DNA replication reaction with the DNA you are sequencing being used as the template. DNA polymerase requires a user-provided primer to start copying the template. DNA replication will stop at random locations due to the presence of **chain terminators**. Chain terminators are deoxynucleotide triphosphates (dNTPs) which DNA polymerase is unable to use to extend the DNA chain. Historically, **di-deoxynucleotides** were used as chain terminators because they lack the 3’ hydroxyl group which is required for chain elongation. Four separate reactions were prepared, one reaction with di-deoxy versions of each base (e.g. ddATP, ddTTP, ddGTP, ddCTP). Together, all four reactions will produce all possible fragments that differ in length by only a single nucleotide. Because we know the last base in each fragment is the di-deoxy nucleotide present in a given reaction, we can determine the sequence of the DNA. In the modern version of Sanger sequencing called **dye terminator sequencing**, fluorescent chain terminator dNTPs are used. Because each base has a different fluorescent tag, DNA sequencing can be done in a single reaction instead of four separate reactions. DNA fragments are separated in a thin capillary tube and a computer reads the fluorescent signals and interprets them as bases in a linear DNA sequence. This method is accurate and can read relatively long sequences (~700 bp), two traits that made it useful for sequencing the human genome. Unfortunately it is rather expensive and time consuming so most new genome sequencing is being done with more advanced technologies.

There are several **second-generation sequencing** (also called next-generation) methods but they all rely on the use of polymerase colonies or “polonies”. **Polonies** are clusters of PCR products derived from a single molecule of nucleic acid. When sequencing metagenomics samples there are a large number of different templates available. Polonies can be thought of as a big pile of PCR products from one specific template. After the polonies have been formed, enzymatic reactions are done in parallel on all the polonies to sequence the DNA. Second-generation sequencing techniques have been used successfully as a cheaper alternative to traditional Sanger sequencing and have even been used to successfully sequence the entire panda genome4. The two most common second-generation sequencing machines are the **Roche 454** and the **Illumina GAII**.

***Assembly of the Individual Sequences***

Regardless of the method used to sequence the DNA, a large number of sequences are generated that need to be assembled into a complete genome. Computer algorithms are used to assemble overlapping shorter sequence reads into longer contiguous sequences called **contigs**. Contigs can be combined into large **scaffolds** which are then stitched together to build complete genomes. The more contiguous DNA we have to analyze, the more information about the organism we can gain. Unfortunately because metagenomics samples contain a huge number of different species, genomes of all but the most common species are incomplete. Although incomplete genomes are the norm, there is still a lot of information that can be gleaned from these sources including information about variation within and between genomes and the genes present in a given species.

Because assembly requires overlapping regions on the reads, it is generally easier and more accurate when longer sequences are used. In order to be certain that a given region is correct, it is good to sequence it many times over, especially when using methods that give very short read lengths (50-100 bases). **Coverage** is defined as the mean number of times each nucleotide in a given genome is sequenced. Some regions will be sequenced more times and other regions will be sequenced fewer times. For example, 5X coverage means that each nucleotide in a given genome is sequenced a mean number of five times although the actual number of times each specific nucleotide is sequenced can vary. If we could sequence an entire genome in a single read then 1X coverage would be enough, but because overlapping regions are required, more coverage is typically needed.

***Gene Identification and Description: “Gene Calling”***

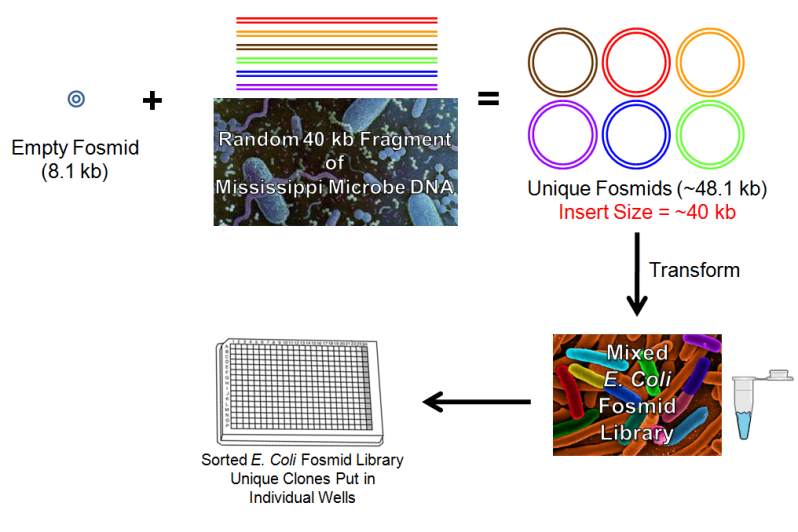
Does the sequence “AGGATGCCTAACTGTCCAT” mean anything to you? It is not enough to know the sequence of all the bases in a given genome; you also want to know how the genes are distributed in the genome and what they do. That’s where **gene calling** comes in. Calling a gene actually means finding and annotating a gene within a particular database. Genes can be identified by computer programs which search sequences for **open reading frames** or **ORFs** which extend from an ATG start codon to a TAG, TGA, or TAA stop codon. Various methods are used including comparing your metagenome data to a list of known genes and trying to find new ORFs by looking for promoters as well as start and stop codons. Because most open reading frames are at least 60 amino acids long, some programs use this criterion when selecting ORFs. If you are interested in a particular gene, you can search for similar genes in your metagenome data to see if related genes are contained within your metagenomics sample. This technique is great for finding new gene variants but it cannot be used to identify novel genes which have no existing homologs.

***Creating a Fosmid Library to Perform a Function-based Metagenomics Screen/Selection***

Recall from the previous lab, that in order to perform a function-based screen or selection, scientists must first create a **DNA library**. A DNA library consists of random fragments of DNA (genes from organisms in a particular metagenome), which are stored in circular DNA vectors to increase their stability. These circular vectors are then put into a microbe [usually *Escherichia coli* (*E*. *coli*)] so they can be replicated during the microbe’s life cycle. The entire library is too large to put into a single vector in a single bacterium; instead, the complete library consists of many individual bacteria each containing their own unique fragments of environmental DNA. This technique allows scientists to preserve and copy the DNA they have painstakingly obtained from the environment in the form of *E. coli* clones. The *E. coli* clones, each containing different fragments of DNA, may be all mixed together into a single tube, or each individual *E. coli* clone with its unique DNA fragment may be separated into different wells of a 384-well plate. The advantage of the later method is that it makes it really easy to find a particular clone and test it more than once.

For metagenomics projects, researchers chop up DNA into manageable pieces and insert these random chunks of environmental DNA into a circular DNA vectors called **fosmids**. Recall that a fosmid is a large circular piece of DNA that can contain many genes. The Minnesota Mississippi Metagenomics Project (M3P) uses fosmids which contain inserts that are about 39 kb in size. This is large enough to contain about 20-40 bacterial genes per fosmid. Usually only one or a few fosmid copies are present in each microbe and these fosmids are copied and distributed to daughter cells every time *E. coli* cells divide. To analyze protein function, fosmids must be placed in a host that is easy to culture. Usually a strain of *E. coli* is used, but other species can also be used. It is useful to make a library in more than one species because not all genes will be expressed in all bacteria and some proteins may be toxic to certain microbes but not others. After the environmental DNA fragment is cloned into a fosmid, and the fosmid has been taken up by a suitable host cell; the next step is to isolate individual fosmid clones. This is accomplished by plating your *E. coli* at a dilution such that single colonies, which each contain a single fosmid insert, can be seen growing on a large agar plate. A robot called a **Q-bot** can then be programmed to select individual colonies, which are subsequently used to inoculate one of the wells of a 384-well plate. Replicators, which have 384 prongs, can be dipped into a 384 well liquid culture plate and then touched to a solid agar plate designed to assay or test microbes for a particular function (e.g. antibiotic resistance). Antibiotics are applied to the solid agar plate, and microbes that survive in the presence of the antibiotic are considered to be antibiotic resistant. These types of life or death assays are called **selections** because they select for microbes which can survive and grow**.** Another type of assay called a **screen** will highlight the microbes with desired properties without killing them. Colorimetric assays are a type of screen in which microbes possessing a desired function are stained a particular color.

**Creation of an *E. coli* Fosmid Library**

****

**Figure 1. Creation of an *E. coli* Fosmid Library.** Random pieces of environmental DNA are put into fosmids. The fosmids are then cloned into *E. coli* to make a mixed *E. coli* library. In many cases, the mixed library is then streaked for isolation and unique clones are placed in each well of a 384-well plate for storage and screening purposes.

<http://www.bio-world.com/productinfo/4_74_587/4572/EZ-Open-Microcentrifuge-Tube-mL-Amber-PK.html>

<http://www.currentprotocols.com/protocol/ch100115>

<http://www.industryleadersmagazine.com/lethal-e-coli-infecting-1600-across-europe-deemed-mutant-super-bug/>

<http://topnews.ae/content/22519-scientists-analyze-why-do-bacteria-fight-consistently-inside-human-body>

The M3P project has produced an *E. coli* fosmid library which contains samples taken from 11 sites along the Minnesota portion of the Mississippi river. About twenty-five 384-well plates were produced from each site. If you do the math (11 X 25 X 384), there are about 105,600 individual fosmids clones; therefore, a screen of all of the clones could take a long time to complete. The *E. coli* fosmid libraries you’ll be working with today have already been prepared for you. You will be working with mixed libraries which will allow you to screen a large number of clones in a relatively short time. The selection that you will be performing is similar to preliminary tests performed to determine the type and levels of antibiotic resistance present in each sample. Many clones must be screened in order to find genes with a particular function, so it is quite possible that you may find no clones which are able to survive in the presence of your antibiotics. If that turns out to be the case, it is totally okay and it does not necessarily mean that you did anything wrong.

For more videos about DNA sequencing and fosmid library production visit the Dolan DNA learning center at <http://www.dnalc.org/>.

***Antibiotics and Antibiotic Resistant Bacteria***

Antibiotics typically either kill microbes or drastically slow their growth. Today we use many kinds of antibiotics as our first and last line of defense against many diseases. In 1928, Alexander Fleming discovered the first antibiotic: penicillin. Penicillin is produced by a mold. When penicillin was serendipitously discovered on a petri dish, it prevented bacteria colonies from growing nearby. Antibiotics work by effectively poisoning the molecular machinery of the cell. Typically they bind to proteins to block translation or DNA replication to prevent microbial growth, or they disrupt the cell wall to elicit cell lysis.

Like penicillin, most antibiotics that exist today were derived from substances produced by microbes. These microbes use antibiotics to restrict the growth of other types of microbes. Microbes that produce an antibiotic to which they themselves are also resistant can have an advantage over their neighbors. They can use more of the resources and nutrients in their area because they make their surroundings uninhabitable to their competitors. Other microbes have responded to these assaults by evolving mechanisms of antibiotic resistance. Microbes have evolved resistance to antibiotics in almost every way you can imagine; they have evolved cell walls to prevent the antibiotic from getting in, pumps to pump it out, proteins which degrade the antibiotic, and modified versions of target proteins which are no longer susceptible to the action of the antibiotic. In short, some microbes have been having an antibiotic arms race against one another for eons, long before humans started to use antibiotics for medicine.

When antibiotics were first used for medicinal purposes they worked wonders. Penicillin for example was found to be effective against Staphylococcus and syphilis, which have historically been serious human diseases. Many new antibiotics have been developed to help treat different kinds of bacterial infections. Antibiotics used in medicine and research labs include gentamicin, tetracycline, streptomycin, cephalexin, ciprofloxacin, levofloxacin, tigecycline, erythromycin, ampicillin, kanamycin, chloramphenicol, vancomycin and many others. Because antibiotics have been advertised to treat any infection, people often go to their doctors and ask for antibiotics whenever they are sick. What many don’t understand is that most antibiotics won’t work against viral infections like those that cause colds, flus, sore throats (except those caused by strep), and most coughs. In other cases where antibiotics are the preferred treatment, people often make the mistake of stopping the medication once they feel better. By not taking the full dose, microbes which were slightly resistant are allowed to survive and propagate, making future treatments more difficult. Even when antibiotics are successfully used, they can still have collateral damage by killing off ‘good’ bacteria which normally inhabit your gut. In addition to the use of microbes in human medicine, a large number of antibiotics are used in agriculture as a prophylactic and growth enhancer for cattle and other animals. Many types of cattle, swine, and poultry are fed antibiotics on a daily basis. Overuse of antibiotics encourages the evolution of new antibiotic resistance genes and the transfer of existing antibiotic resistance genes between different species of bacteria. This is a classic example of evolution in action. The use of antibiotics inadvertently selects for rare resistant microbes. In many cases antibiotics successfully wipe out 99.9% of bacteria, but it only takes one resistant mutant to create a new resistant strain. An unfortunate side-effect of antibiotic overuse and misuse is that multidrug resistant pathogens are becoming increasingly common and more difficult to treat.

As mentioned above, one goal of a functional metagenomics study might be to determine the prevalence of antibiotic resistance in a natural environment. By determining the presence of antibiotic resistance in the environment we can understand what naturally occurring resistance is out there and what new antibiotic resistance is evolving due to human antibiotic use. The more antibiotic resistance genes that are out there, the higher the risk that these genes will wind up in pathogens capable of infecting humans. Today, you will be screening part of an *E. coli* library containing Mississippi metagenomic DNA. You will perform a selection on never before tested clones to examine the level of resistance to your chosen antibiotics at your chosen sites along the Mississippi river. To do this experiment, you will be using a modified version of the Kirby-Bauer antibiotic susceptibility testing procedure which is routinely used to determine whether different strains of bacteria are susceptible to a particular antibiotic.

For a list of antibiotics and diseases they can treat see:

<http://www.merckmanuals.com/home/print/sec17/ch192/ch192a.html>

***Gel Electrophoresis***

Gel electrophoresis is a common technique for visualizing and separating DNA fragments based on their size. Today you’ll be using this technique to visualize products from the PCR reaction you setup last lab period. The gel you’ll be using today is a gelatinous slab of agarose (a substance derived from seaweed). At the molecular level this gel contains many intertwined polysaccharide molecules, which form a network of small tunnels going in all directions. DNA is loaded into the wells near the top portion of the gel. When an electric current is applied, the negatively charged DNA will migrate towards the positively charged cathode. DNA fragments of different lengths will migrate at different speeds through the small tunnels in the gel. Small DNA fragments can move through the tunnels more quickly than larger fragments so they will end up near the far end of the gel, while the slower moving larger fragments will end up near the wells at the top of the gel. A DNA ladder is usually loaded on one or both sides of your gel. The ladder (or molecular weight marker) contains DNA fragments of known sizes. The size of your DNA fragments can be compared to the ladder to determine their size.

**OBJECTIVES**

1. Compare and contrast sequence-based & function-based metagenomics and list the strengths and weaknesses of each.

2. Describe how a functional metagenomics screen is performed and explain how this approach can be used to uncover novel proteins for use in research, medicine, or industry (e.g. new cellulases for biofuels).

3. Explain how the 16S rRNA gene PCR product is used in the analysis of metagenomes, (i.e. what information can be gained from the analysis of this gene?).

4. Define an ‘*E. coli* fosmid library’ and use that definition to explain why many library clones must be assayed to identify proteins of interest.

5. Load and run an agarose gel and describe the principle behind gel electrophoresis.

6. Plate *E. coli* clones from the fosmid library on selective media with antibiotic discs and be able to interpret the results.

**MATERIALS**

**Functional Metagenomics: Antibiotic Selection**

Mixed *E. coli* fosmid library from each of the ten river sites

LB chloramphenicol plates

Antibiotic discs (See table below)

Permanent markers for labeling plates

Sterile Swabs

Sterile forceps

37°C incubator

Ethanol and oil lamps to sterilize the forceps

Large autoclave waste beaker for used swabs.

|  |  |
| --- | --- |
| **Available Antibiotics** | **Class** |
| Tetracycline | Tetracycline |
| Oxytetracycline | Tetracycline |
| Ampicillin | Penicillin |
| \*Penicillin (NEGATIVE CONTROL) | Penicillin |
| Kanamycin | Aminoglycoside |
| Neomycin | Aminoglycoside |
| Ciprofloxacin | Fluoroquinolone |
| Sulfathiazole | Sulfonamide |
| Sulfamethoxazole/Trimethoprim (SXT) | Folate pathway inhibitor |
| Vancomycin | Glycopeptide |

\**E. coli* are Gram-negative microbes that are naturally resistant to penicillin

**Sequence-based Metagenomics: Electrophoresis and Imaging Gels**

|  |  |
| --- | --- |
| PCR Products from previous lab period  1.5% agarose gels  Gel boxes  Electrophoresis power supplies  1-10 µl micropipettors and tips  Molecular weight markers (Ladder)  10X loading dye  0.5 X TBE buffer | Staining and destaining trays  EtBr stain (1 µg/ml)  Rubber spatulas  EtBr waste container (for gels and paper etc.)  UV transilluminator box  Digital gel documentation system  Pipette tip discard container  Vinyl gloves |

**EXPERIMENTS**

**I. Visualizing Your DNA Using Gel Electrophoresis (Work in groups of two)**

**Caution:**

1. Do not use the power supply if the electrical jacks are wet!

2. Do not operate the gel box system without a safety cover; it may result in a shock!

3. Ethidium bromide is a mutagen and probable carcinogen. The dilute solution you’ll be using poses a minimal risk if handled properly. However, always wear gloves when working with or handling the gel and during cleanup. Wash your hands thoroughly before you leave the lab.

**A. *Loading and Running Your Samples***

1. Using a micropipette, add 2.5 µl of 10X loading dye to your sample tube containing 22 µl of PCR product. Mix by tapping your tube on the bottom of the lab bench.
2. The instructor will also load 7.5 µl of molecular weight marker (DNA ladder) to the well of the first lane on the left of your gel so that you can confirm the size of your PCR product.
3. Use the micropipette to load 10 µl of your sample into the well of the next open lane of the gel. Write down which lane your sample is in, and record this information on the provided Gel Data Sheet.
4. Once everyone has loaded their sample(s), close the top of the electrophoresis box. Place your hand over the top of the gel box to stabilize it when you connect the power cords. Connect anode-to-anode (red- to- red) and cathode-to-cathode (black-to-black). Run at 150 volts (make sure it’s not set to milliamps) for 40-45 minutes. Monitor the progress of the migrating blue band and make sure you turn off the gel box before this runs off the gel. Inform your instructor if the blue band is not gradually moving as it may mean you have a broken power supply.
5. After the gels have run (i.e. the DNA has migrated) for the allotted time; first turn off the power supply, then slide the lid to the side and remove your gel from the electrophoresis box. WEAR GLOVES AND BE VERY CAREFUL WITH THE GEL. They are slippery and drop easily. A dropped or a torn gel is useless.

**B. *Staining and Viewing Your Gel***

**Caution:**

1. Remember to wear gloves as you are working with ethidium bromide.

2. Ultraviolet light can damage the retinas of your eyes. Never look directly into the UV light. Make sure the camera hood is seated properly before turning on the UV transilluminator.

1. Place your gel in the staining tray. Immerse your gel in staining solution containing ethidium bromide (1 µg/ml) for 5 minutes. Ethidium bromide will intercalate between the bases in the DNA and will bind to it so it can be visualized with a UV light. WEAR GLOVES when working with ethidium bromide because it is a mutagen and a carcinogen.
2. After the staining is complete, use the spatula to CAREFULLY transfer your gel to a tray containing distilled water. If you spill any ethidium bromide solution, wipe it up immediately using bleach and paper towels. Discard the towels in the designated waste container.
3. Let your gel soak in the distilled water to destain it for 5 minutes. This will remove background ethidium bromide from your gel.
4. To get an image of the DNA on your gel, it is necessary to illuminate your gel with ultraviolet (UV) radiation. Put a plastic number near the bottom of your gel to help identify it. Your lab instructor will place your gel in a tray that sits on a pull-out UV transilluminator of an imaging system. Your instructor will then help you take a picture of your gel using the digital gel documentation system. You will receive a digital copy of your gel. The gel photos for your class will also be projected from the computer for class discussion.

**Band Intensity:**

The intensity or brightness of the DNA bands on your gel is affected by a variety of factors. First, more ethidium bromide can intercalate between the bases of longer strands of DNA resulting in a greater intensity in larger bands. Second, the intensity of your band depends on microbial abundance in your sample. If a certain microbial taxon is not present, you will get no corresponding band for that group on your gel. A faint band could potentially indicate a low abundance of your microbe in your sample location. Third, the intensity of your band depends on how much template is available for PCR. Even if you have plenty of microbes of a given taxa present in your initial sample, you may not get a PCR band on your gel if those microbes were filtered out by the 5.0 µm filter or if the DNA extraction procedures that were used were not sufficient to break open those cells to expose their DNA. For example, cyanobacteria and archaea cells tend to be difficult to break open using the method that you used. In summary, band intensity is affected by microbe abundance, template abundance, and the size of your PCR product.

**Questions:**

1.Which direction does the DNA move? What about the molecular nature of DNA makes it move in this direction?

2.Do larger DNA fragments or smaller DNA fragments move faster through an agarose gel? Why?

3.Did your PCR reaction work? How can you tell?

4. What was the size of your PCR product? Is this what you predicted?

**II. Function-based Metagenomics (Work in groups of four)**

**Caution:**

1. While antibiotics are generally considered safe, some can cause cancer or lead to birth defects in individuals exposed to high concentrations. While you’ll be working with safe concentrations it is still a good idea to ALWAYS WEAR GLOVES WHEN WORKING WITH ANTIBIOTICS.

2. ENSURE THAT NO ONE IN YOUR GROUP IS ALLERGIC TO THE ANTIBIOTICS YOU SELECT. IF SOMEONE IS ALLERGIC TO PENICILLIN, USE A DIFFERENT ANTIBIOTIC DISC INSTEAD.

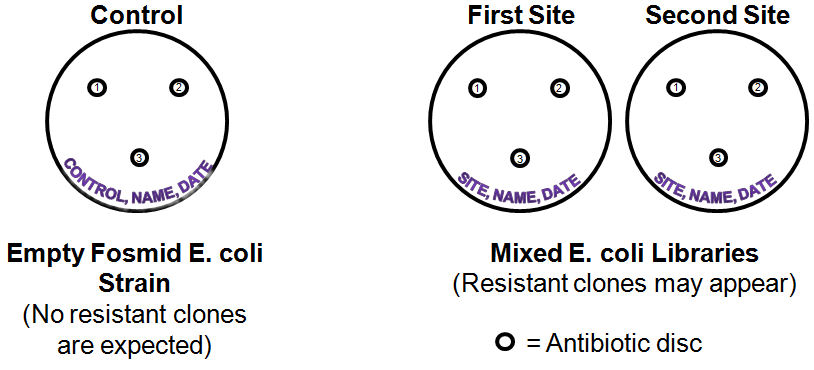
***Selection for Antibiotic Resistance at Different Mississippi River Sites***

Today you will be working with mixed fosmid libraries in *E*. *coli* bacterial cells. individual cell (i.e. a clone) contains a unique DNA fragment from the Mississippi metagenome. Because each fosmid has a random piece of DNA, most clones will not be resistant to your chosen antibiotic. You will be performing a **selection** to identify the few clones which contain genes within their fosmid that confer antibiotic resistance.

Each group will receive a total of three 100 mm LB-agar plates. Because the fosmids contain a chloramphenicol resistance gene, all plates will contain 7 µg/ml of chloramphenicol to maintain the fosmid in all clones. You will plate the mixed *E. coli* library from the two river sites you selected on two of these plates. A strain of *E. coli*, which contains a fosmid with no insert, will be plated on one additional plate as a control. Because this control strain of *E. coli* lacks any insert DNA, it should not be resistant to the antibiotics you are testing (except penicillin). Your control plate should have clearing zones but no resistant colonies within them. It is important to realize that the control strain contains many individual copies of identical cells while the libraries contain many cells which each have their own unique piece of environmentally derived DNA. The exciting part about this experiment is that no one knows what the results will be. This is real science!

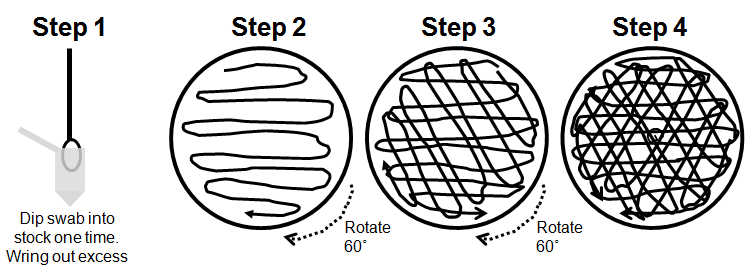
For this experiment, you **will work in groups of four.** At the end of the next lab period each group will present their results to the class. Be prepared to share with the class (1) your hypothesis, (2) why you selected this hypothesis, (3) the results you observed, and (4) what you might do differently in the future if you were to redo this experiment.

If you find an abundance of antibiotic resistant clones, or if your hypothesis was supported by your results please inform your TA so that your results can be further verified. It is also possible that you won’t find any resistant clones. If that is the case, do not worry; negative results do not necessarily mean you did anything wrong.

**Figure 2. Experimental Setup.** Label and prepare the plates as indicated.

1. **Label the plates with**:
2. The Mississippi river site you selected to sample from.
3. The name or initials of the members of your group.
4. The date.

1. **Inoculate the library and control strains onto the plate.** You will have to share stock tubes with your classmates so please be courteous and return tubes as soon as you’re done using them.
   1. After the libraries have thawed, select the site that contains your library of interest, invert the tube several times to mix.
   2. Carefully insert a sterile swab into the 1.5 ml tube containing the library or control strain being careful not to spill any *E. coli*.
   3. Withdraw the swab, place it in the cap of the 1.5 ml tube and spin it around with slight pressure to wring out excess liquid.
   4. Close the stock tube and return to its storage location.
   5. Streak out your *E. coli* across your plate in a zig-zag pattern (Fig. 3).
   6. When the entire length of the plate has been swabbed, rotate it by 60 degrees and then repeat the process. Rotate the plate again and repeat the process a third time to ensure the *E. coli* are evenly spread over the entire plate.
   7. You’ll inoculate a library from each of your two selected sites onto their own respective plate, and the control strain on the control plate. (Fig. 2).

**Figure 3. Plate preparation protocol.** Inoculate the entire plate surface with a swab as indicated.

**3. Add antibiotic discs to the plate**.

1. Locate disc dispensers for the antibiotics you intend to use for your experiment.
2. Select Penicillin (a good control antibiotic as *E. coli* is naturally resistant to it) and two other antibiotics you want to test.
3. Using a sterile forceps, remove a disc from the dispenser.
4. Gently set the disc down on your plate BEING CAREFUL NOT TO LET THE FORCEPS TOUCH THE PLATE. **Note**: If your forceps becomes accidentally contaminated with *E. coli*, you may re-sterilize it by dipping it in ethanol and then flaming it.
5. Place the discs evenly over the plate and in the same relative positions on all plates as diagrammed above (Fig. 2). Do not place the discs too close together or too close to the edge of the plate.

4. **Store the** plates in the incubator disc-side up at 37˚C for **three days**. The plates will then be removed from the incubator, wrapped with parafilm, and stored at 4˚C until the next lab period.

1. **Interpret the results.**
   1. Next lab period you will examine your plates.
   2. Around each antibiotic disc there should be a clearing zone where few if any *E. coli* cells can grow.
   3. The size of the clearing zone can be affected by a number of factors including the depth of the agar, the density of the microbes, the concentration of the antibiotic, the effectiveness of the antibiotic, and the solubility of the antibiotic.
   4. The edge of the clearing zone represents the minimal inhibitory concentration, (i.e. the amount of antibiotic that is just enough to kill your particular microbe).
   5. Resistant *E. coli* will appear as colonies within the clearing zone.
   6. Colonies which are larger and/or closer to the center of the clearing zone are more resistant to your chosen antibiotic than are colonies located near the edge of the clearing zone.
   7. Be sure to note the number of colonies, their distance from the center of the clearing zone, and their relative sizes and compare these attributes between sample sites to support or refute your hypothesis.

**Figure 4. Examples of Clearing Zones Containing Resistant Colonies.**

**Questions:**

1. Describe one advantage and one limitation of functional metagenomics.
2. Define a fosmid and describe why they are kept in *E. coli*.
3. What are the controls in this functional metagenomics experiment and what is their purpose?
4. What results do you expect if your clones are resistant to your antibiotic? What results would you see if they are not resistant?
5. What factors might affect the size of the clearing zones?
6. Why do all of the plates contain the antibiotic chloramphenicol?

**References:**

1. Sciences, N.A. of & NRC *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*. *Design* 171 (National Academies Press: Washington, DC, 2007).at <http:/www.nap.edu/catalog.php?record\_id=11902>

2. Wooley, J.C., Godzik, A. & Friedberg, I. A primer on metagenomics. *PLoS Comput Biol* **6**, e1000667 (2010).

3. Baker, M. Next-generation sequencing: adjusting to data overload. *Nat Meth* **7**, 495-499 (2010).

4. Worley, K.C. & Gibbs, R.A. Genetics: Decoding a national treasure. *Nature* **463**, 303-4 (2010).

NOTE TO TA AND LAB COORDINATOR:

1. Make sure the appropriate mixed *E. coli* fosmid libraries are thawed and kept on ice before class.

2. Avoid repeated freeze thaw cycles of the E. coli libraries, keep on ice in lab and place in the refrigerator between lab sections.

3. Only thaw the necessary amount of fosmid library. The rest may be stored and reused.

4. Note the three day incubation time. Ensure that you have enough incubator space for this period of time. Two days will work in an emergency but three is better.

5. Warn people that if they are allergic to penicillin or some other antibiotic that they shouldn’t touch it!