BIOL 101 H Lab

DNA FINGERPRINTING LABORATORY EXERCISE

Introduction

A. DNA and Genetic Identity

Every person who ever existed is a unique individual. We can easily recognize individual differences in traits such as facial appearance, size, skin and hair color, personality, and performance. More subtle characteristics such as fingerprints and patterns of vascularization in the retina of the eye can be used to precisely distinguish individuals from one another and almost all of these physical differences are genetically determined. Put simply, people are physically different from one another because they contain different genes or different combinations of genes. Physical traits are encoded by the sequence of DNA base pairs contained in a person's chromosomes. The four nucleotide bases are represented by the letters A, C, G, and T, which stand for their chemical names, adenosine, cytosine, guanine, and thymine. The A in one strand always pairs with the T in the opposite strand, whereas C always pairs with G. The strands are said to be complementary to each other, that is if you know the sequence of one strand you can determine the exact sequence of the other strand. The sequence of base pairs that make up a gene encodes for proteins. These are the molecules that determine the biochemical properties and physical form of an organism. Most physical traits are determined by multiple genes which act together during embryonic development to give form to an individual. Each gene exists in two copies in every person, since each of our 23 chromosomes is paired. The physical location of a particular gene on a chromosome is called a genetic locus (the plural is loci). In a population, many different forms of a single gene can exist at any locus. Each unique form of an individual gene is known as an allele. When one considers that thousands of gene pairs act together during development and that multiple alleles can exist for nearly every gene, the amount of diversity we see within the human species is easy to understand.

There are approximately 3 billion DNA base pairs in the human genome. (The base pair count for a species represents the sum of genetic information for that species.) About 99.9% of this DNA sequence is identical in all humans. The 0.1% difference is responsible for our genetic identity as individuals. The differences between one person's DNA and another person's DNA are called polymorphisms (Greek: poly = many, morph = form). There are 3 different types of polymorphisms: In one form of polymorphism, the sequence of the DNA base pairs are rearranged in a way that changes the meaning of the DNA code. For example, if you change the order of the letters in the word CAT to ACT, the meaning of the word is different. A second way to change DNA is to insert additional base pairs within a coding sequence. This is like adding letters to a word and changing the meaning, as in adding 2 additional letters to the word ACT can change the meaning to ACTOR. The third way that polymorphisms occur in DNA is by subtraction (or deletion) of nucleotides. By deletion ACT could become AT, a different word with a different meaning.

B. DNA Profiling of Individuals

DNA is a rather stable molecule in chemical terms. DNA released from dead or dying cells or tissues can be digested by microorganisms, but if the sample is dried, the DNA will remain intact for many years (or even thousands of years in mummified remains). Thus, material containing DNA from victims and perpetrators of a crime is often present at the scene. Since the sequence of base pairs in a DNA sample collected at a crime scene can contain the genetic identity of the individuals present, forensic scientists have borrowed genetic techniques to develop "DNA fingerprints" of suspects.

While it is technically possible to determine the base sequence of DNA fragments collected at a crime scene, this would be difficult, expensive, and time consuming. Therefore, rapid methods that type or characterize DNA from different individuals have been developed. Two methods that detect differences in DNA without actually determining the sequence are known as Southern blotting and PCR, (Polymerase Chain Reaction).

Southern blotting (named after its inventor, Ed Southern) uses special enzymes known as restriction endonucleases that cut DNA at specific base pair sequences to produce fragments of discrete size. To distinguish DNA fragments derived from a specific gene from the thousands of other fragments produced by an enzyme, the fragments are separated according to size by gel electrophoresis (see below), transferred from the gel to a piece of special membrane, and "probed" with a labeled DNA fragment that is complimentary in sequence to the gene to be identified. The size of the DNA fragments detected by the labeled probe often differs between individuals because of polymorphisms. Southern blotting is often used in forensics because many different genes can be examined simultaneously. The probability of 2 individuals having identical patterns of DNA fragments just by chance decreases dramatically as more and more genes are compared. Southern blotting has 2 main drawbacks. First, the DNA must be intact or at least not badly degraded. The molecules need to be long enough to give a discernable pattern of fragments after cutting with the enzymes. Also, a relatively large amount of DNA is required for analysis. Southern blotting is usually only practical when fresh samples of tissue such as blood or semen can be collected. We will not be using Southern blotting in this exercise.

The polymerase chain reaction was developed by Kary P. Mullis, who was awarded a Nobel Prize for the discovery. The PCR reaction uses a DNA polymerase enzyme that is isolated from the bacteria *Thermus aquaticus* (Taq) that lives in hot springs in Yellowstone National Park. This organism normally lives at temperatures exceeding 70° C. In nature Taq DNA polymerase replicates the bacterial chromosome during cell division; however, in a test tube reaction, Taq polymerase can be used as a thermo-stable reagent to allow multiple rounds of replication of relatively short, well-defined DNA segments from any organism. In the PCR reaction, a segment of a complex genome is "targeted" for amplification by Taq polymerase by defining the region with 2 short synthetic DNA molecules complementary to sequences flanking the region of interest. These short DNA molecules are known as primers. PCR primers are made on a special machine called an oligonucleotide synthesizer and they can contain any desired sequence. Therefore, any gene or DNA region where the sequence is already known can be amplified by PCR. The regions of DNA we will be amplifying are called microsatellites. Microsatellites are non-coding regions of the DNA (they do not make proteins and have no function as far as we know) consisting of long repeats of specific base pairs. For example the three microsatellites we will be amplifying consist of ATA repeats, that is they are series of ATA ATA ATA repeated

various numbers of times. Yes, in this case all three of the microsatellites have the same repeat unit. However, this no the case with all microsatellites, which, can have repeats using any combination of the four base pairs. Because microsatellites are non-coding there is a high mutation rate because it doesn't matter if you have 20 repeats or 50 repeats. Given this, every individual will have a unique combination of microsatellites and a unique DNA fingerprint using these markers. These markers are also used to help diagnose genetic diseases (none of our primers are associated with any know disease) and determine parentage. Each of you has two copies of each chromosome, one from each parent and thus two copies of each microsatellite. When both of your copies of the microsatellite are the same size (have the same number of repeat units) we will see only one band on the gel and you are called homozygous for that locus. If the alleles you got from your parents are different sizes we will see two bands on the gel and you are called heterozygous for that locus. We are working with three human microsatellite loci, one from chromosome 11, one from chromosome 12, and one on the Y.

A PCR reaction contains the target DNA and all the necessary components for DNA duplication: large quantities of all 4 deoxynucleotide bases, large quantities of the two primer DNA fragments, salts, and Taq polymerase. The polymerase chain reaction is carried out in 3 steps:

1) The 2 strands of the target DNA double helix are separated by heating to 94° C for a few seconds.

2) The reaction is then cooled to about 58° C. This allows the 2 primers to bind or "anneal" to their complementary strands in the target DNA. This takes 20 - 30 sec.

3) The reaction is then heated up to 72° C, the temperature at which Taq polymerase works best. The Taq polymerase begins adding deoxynucleotide bases to the ends of each primer and eventually makes a complementary copy of the template DNA. This completes one PCR cycle. A diagram of the PCR cycle is shown in Figure 1.

Theoretically, PCR can amplify a single molecule of target DNA. Thus, at the end of the first cycle of PCR, we would now have 2 molecules of the targeted DNA segment. The 3 steps in the PCR process - strand separation, annealing of the primers to the template DNA, and synthesis of the new DNA strands - takes less than 2 minutes. But the cycle can be repeated at least 30 more times before the Taq polymerase loses activity from the repeated heating. Each newly synthesized DNA strand can act as a new template, so after 30 cycles, 1 billion copies of the original DNA segment can be produced! Taking into account the time it takes to change temperature of the reaction vial, 1 billion copies can be ready in about 3 hours.

The speed and extreme sensitivity of the PCR reaction has revolutionized the use of DNA evidence in criminal investigations. A few million copies of an amplified DNA fragment is sufficient material to be detected by staining with fluorescent dyes following separation of the reaction products by electrophoresis, or via laser excitation of fluorescent markers attached to the fragments. Thus, it is possible for forensic scientists to analyze the genetic patterns found in only a few molecules of DNA collected at a crime scene. A large number of DNA primers flanking different regions of the human genome that are known to be polymorphic are now available. Most of these polymorphisms involve deletions or insertions; thus it is relatively easy

to detect the various forms of the genes that exist in different individuals. As with Southern blotting, if enough different genetic loci are examined by PCR, the probability that any two people have exactly the same PCR products becomes vanishingly small. Genetic "fingerprints" yielded by PCR can be just as effective as traditional fingerprints in distinguishing identity. The simplicity of the reaction now makes PCR the method of choice for many applications that require DNA typing. These applications include criminal investigations, paternity cases, screening for genetic disorders, and the identification of human remains following disasters. To facilitate identification of military personnel, samples of DNA from all U.S. servicemen are now kept on file in case DNA fingerprint analysis is ever required in the future.

C. Limitations of DNA Evidence

DNA fingerprinting is not fool proof and does not always provide conclusive evidence in criminal cases. There are some technical and statistical issues that affect the use of DNA evidence. Human error in handling tissue samples collected at a crime scene and during the laboratory work is always a possibility. Samples can accidentally be switched or mislabeled during DNA preparation or electrophoretic analysis. Also, the PCR reaction is highly prone to contamination. Remember that only a few molecules of extraneous DNA introduced into a sample are sufficient to produce a detectable PCR product. The most common source of DNA contamination is DNA from another sample or from the individuals collecting the samples or doing the lab work. The usual result of contamination is the presence of DNA fragments that typically do not match any of the individuals involved in the crime. While thoughtful analysis of contaminated samples can often distinguish the presence of a suspects DNA pattern despite the presence of contaminating DNA, attorneys tend to encourage jurors to discount any evidence containing contaminated DNA (as in the criminal trial of O.J. Simpson). Because only small segments of DNA are amplified, the PCR test works with partially degraded DNA; however, some DNA samples can be degraded to the point where PCR cannot amplify a fragment.

Finally, the use of DNA evidence to convict a suspect requires statistical evidence regarding the "uniqueness" of a particular DNA pattern. The frequency of a particular DNA pattern in the human population is determined from information stored in databases maintained by the FBI and other federal agencies. From the frequency of a particular DNA pattern in the entire database, it is possible to calculate the probability that an individual other than the suspect would carry the same fingerprint. Because multiple genetic loci are usually analyzed, the numbers can seem astronomical. For example, if the frequency of a suspect's DNA pattern at the first locus tested is 1 out of every 40,000 individuals and the frequency of the pattern determined at a second locus is 1 out of 1,000 individuals, then the probability of anyone having the same DNA fingerprint as the suspect would be

$$\frac{1}{40,000 * 1,000} = \frac{1}{40,000,000}$$

As additional loci are analyzed, the theoretical "uniqueness" of a pattern becomes even greater; however, the frequency of specific DNA patterns is not the same for all races and ethnic groups. While the frequency of a particular pattern might be 1 out of 40,000 for the general human

population, it might be 1 out of 25 for individuals from a particular village in Sicily or West Africa. It is becoming clear that databases containing the frequency of DNA patterns for different human subpopulations will be necessary. Data collected from military personnel will provide this information in the future. It is clear, however, that the use of DNA data to convict an individual of a crime will always require a statistical rationale. It is far easier to acquit a suspect using DNA evidence. If the DNA pattern of an individual does not match the pattern found at a crime scene, he or she could not have been the source of the DNA.

Laboratory Exercise

Your Homicide Case

Forensic scientists often use DNA evidence in criminal cases to bring about indictments and to prosecute cases before juries. Consider the following case. A professor at a medium sized Midwestern university is found murdered in his office from multiple stab wounds. There were indications that the professor had struggled with the assailant. His hand was lacerated, suggesting that he had tried to seize the knife from the assailant. His fingernails were broken and skin cells were found under his fingernails indicating that the professor may have scratched the assailant during the struggle. Blood was found on the inside doorknob of the office which was located 20 feet from the body. This professor was a notoriously tough (but fair) grader. The 19 students in his class weren't doing as well as they hoped and became logical suspects as several of them wanted to go to continue on to some form of medical school (requiring good grades in all biology classes) or needed the class in order to graduate. All of the suspects were seen on campus throughout the day the professor was murdered.

Samples of DNA were isolated from the victim and suspects. DNA was also recovered from the skin tissue under the victim's fingernails and from the doorknob. The samples were analyzed by PCR using a set of primers that amplified a polymorphic segment of DNA located on chromosomes 11, 12 and Y. Your task is to analyze all the DNA samples using gel electrophoresis and to decide who is the guilty party based on the evidence you collect.

The first step you will be doing is extracting DNA from the suspects (for whom you will be standing in), the poor victim, and the samples taken at the crime scene.

To do this you must follow the following procedures EXACTLY!!!!!

Extracting the DNA

1. Insert the toothpick into your mouth and scrape off several cheek cells.

2. Place the toothpick in a 1.5 ml microfuge tube and add 180 µl of buffer ATL.

3. Pluck 15 hairs from you head. Pull them out by gripping the hair with the forceps where it inserts into your head. Check each hair to make sure the follicle is attached (the ball at the end).

4. Cut the follicle off of the hair shaft and place it in the 1.5 ml microfuge tube you put the cheek cells.

5. Add 20 µl of Proteinase K, mix using the vortex and incubate at 55^0 C until the tissue is completely lysed (dissolved). Vortex occasionally during the incubation. This should take approximately 40 minutes.

6. Vortex for 15 sec and then add 200µl of Buffer AL and mix on the vortex.

7. Incubate at 70° C for 10 minutes.

8. Add 200 µl of ethanol to the sample and vortex.

9. Pipette the mixture from step 6 into the DNeasy mini column in the collection tube.

10. Centrifuge at 8000 rpm for 1 minute and discard the collection tube.

11. Place the spin column into a new collection tube and add 500 µl Buffer AW1.

12. Centrifuge at 8000 rpm for 1 minute and discard the collection tube.

13. Place the spin column into a new collection tube and add 500 µl Buffer AW2.

14. Centrifuge at maximum rpm for 3 minutes and discard the collection tube.

15. Place the spin column in a clean 1.5 ml microfuge tube and pipette 200 µl of Buffer AE directly on to the membrane in the spin column,

BUT DON'T TOUCH THE MEMBRANE WITH THE PIPET TIP.

16. Let the tube sit for 1 minute at room temperature.

17. Centrifuge the tube for 1 minute at 8000 rpms.

18. Discard the spin column.

This is a two part lab. The second portion of the lab will be done next week where we will create and run the PCR reaction and analyze it on an automated DNA sequencer

Creating the PCR mix

We are going to be running standard PCR reactions of 20µl (0.00002 L, or 0.000005 gallons). The PCR mix will contain the following amounts for each reaction you will perform. As you add the specific reagent put a check mark in the column under its heading.

Reagent	Volume 1 rxn	Total for 6 rxn	MM Y-392	MM X-7108	MM 2-163	MM 10-1714	MM 2-117
Ultra Pure Water	12 µl						
10x PCR Buffer	2 µl						
dNTP's	1µl						
Forward Primer	0.6 µl						
Reverse Primer	0.6µl						
Taq polymerase	0.3µl						
Template	5 µl						
Total	20µl						

Table 1
Amount of reagents need for each primer. Place an X in the cell as the reagent is added.

You will be working in 5 groups of 4 each so each person in a group will be responsible for making one master mix.

1. Label four 0.5 ml tubes as the master mixes for each primer (MMY, MMX, MM163, MM1714, and MM 117) place these in a tube rack in an ice bucket.

ALL REAGENTS AND DNA MUST BE KEPT ON ICE AT ALL TIMES!!!!

2. Label a set of 0.2 ml PCR tubes as 1 10 and put these in the rack in the ice.
3. Add the required amount of each reagent from your calculations in table 1.

MAKE SURE THAT YOU ADD THE REAGENTS IN THE ORDER LISTED ABOVE!!!!

4. Vortex this mixture and place 15 µl of the appropriate master mix in the appropriate 0.2µl tubes.
5. Add 10µl of each individual's DNA to the appropriate tube.
6. Give me your samples and we will start the PCR machine running.

RESEARCH GENETICS' GENOME SERVICES PROTOCOL

The following recommended protocol was developed by the Research Genetics' genotyping center to use with GENEPAIRS primers.

Preparation for PCR:
We use 20 ng of DNA for each reaction, add 2.0µl of DNA (10ng/µl) to each reaction well and allow it to dry. Within 24 hours,
add 10 µl premix to each recently dried DNA sample.

Reagents	Volume/RXN (1.5mM MgCl2)
5x PCR buffer	2µl
dNTPs (1.25 mM @)	1.6µl
MgCl2 (25mM)	0.6µl
sterile di H2O	5.155µl
forward primer (as supplied)	0.3µl
reverse primer (as supplied)	0.3µl
Taq DNA polymerase (5U/µl)	0.045µl
Template TG (100ng/µl)	2.0µl
TOTAL VOLUME/RXN	10.0µl

Multiply these volumes by the number of DNA samples you will run with each marker. We add several extra reactions to our calculations to ensure there is enough PCR mix. We generally have good results with the 1.5 mM Mg++ premix. Almost all markers that don't work well

with the 1.5 mix will work well with 3.0 mM Mg++ premix.

PCR cycling conditions:

Initial Denature 96°C - 2 minutes (1 cycle)

30 cycles:
Denature 94°C - 45 seconds
Anneal 57°C - 45 seconds
Extension 72°C - 60 seconds

Final Extension 72°C - 7 minutes (1 cycle)

These condiitons may have to be optimized for your thermocycler, but this should give you a place to start.

Stock reagents:
5x PCR buffer
8 ml 1M (NH4) 2SO4
33.5 ml 1M Tris-HCl (pH8.8)
500 µl 10% Tween-20
qs to 100 ml with sterile di H2O
25 mM MgCl2
510 µl 4.9 M MgCl2 (Sigma)
qs to 100 ml with sterile di H2O
1.25 mM dNTPs
1 ml 100mM dATP
1 ml 100mM dCTP
1 ml 100mM dGTP
1 ml 100mM TTP

QS to 80 ml with sterile di H2O

*If amplification results are not optimal under these standard conditions, try:
(1) adjusting Mg2+ concentration to 3.0 mM
(2) lowering annealing temperature 2-3°C, when no product is observed OR
raising annealing temperature 2-3°C and/or run product in a gel lane by
itself, when spurious bands are visible
(3) adding double primer (5µl) and rerunning reaction with fresh premix

Store primers at -20°C for maximum shelf-life