Name_____

Ah, Lou! There really are differences between us!

Maria C. Abilock
BABEC

Frank H. Stephenson, Ph.D.
Applied Biosystems

We gratefully acknowledge David Micklos of the DNA Learning Center at Cold Spring Harbor Laboratory for his generous help. Some materials for this exercise were adapted, by permission, from the *Genomic Biology: Advanced Instructional Technology for High School and College Biology Faculty* laboratory manual, Cold Spring Harbor Laboratory, copyright 1999.

Introduction

We are humans. We are bipedal and stand upright. We have hands, feet, fingers, and toes. You can look at the student next to you and easily recognize that person to be human too. What makes us look similar to each other while different from frogs, fish, or fuchsias is the molecule **deoxyribonucleic acid** (**DNA**).

The basic building block of DNA is the **nucleotide** comprising a deoxyribose sugar, a phosphate, and one of the four bases A (adenine), C (cytosine), G (guanine), or T (thymine). In the DNA molecule, nucleotides are linked together in a chain. DNA is a **double helix**; two chains of nucleotides are wound around each other to form a spiral structure. Interactions (hydrogen bonds) between the bases on the opposing strands hold the double helix together. The A's on one strand hydrogen bond with the T's on the other strand. The G's on one strand interact with the C's on the other. Therefore, A's and T's are said to be **complementary** as are G's and C's. Complementary bases, when hydrogen bound in the double helix, are called **base pairs** (**bp**). It is the order of the bases along the strands of the DNA molecule that makes each species unique.

Our bodies are caldrons for thousands of chemical reactions carried out to support the process of life. We ingest food for energy and for the raw materials needed to build the structures of the cell. We breathe oxygen; it assists in the moving of electrons from one molecule to another. We manufacture protein molecules called **enzymes** needed for the building or breakdown of still other molecules. We all look like humans because we all share the same cellular makeup.

The information for the construction of all the enzymes in the cell and all the proteins giving the cell its shape and function is stored within DNA's sequence of bases. One particular base sequence may carry the information for the assembly of hemoglobin, a protein that carries oxygen to your cells. Another sequence of bases may direct the manufacture of an actin molecule, a protein found in muscle. The region of bases on DNA that holds the information needed for the construction of a particular protein is called a **gene**. The average gene is approximately 10,000 base pairs long. There are approximately 35,000 genes in human DNA.

The human **genome** (the total sum of our genetic makeup) is made up of approximately 6 billion base pairs distributed on 46 chromosomes. All cells in your body, except red blood cells, sperm, and eggs, contain these 46 chromosomes (sperm and egg cells contain only 23 chromosomes). Only 3 to 10 percent of this enormous amount of DNA is used directly to code for the proteins required for supporting cellular metabolism, growth, and reproduction. The protein-encoding regions are scattered throughout the genome. Genes may be separated by several thousand bases. Furthermore, most genes in the human organism are themselves broken into smaller protein-encoding segments called **exons**, which, in many cases, have hundreds or thousands of base pairs intervening between them. These intervening regions are called **introns** and they make up between 90 to 97 percent of the entire genome. Since these non-coding areas such as introns have no defined role, they have been referred to as "Junk DNA". Whatever their function may entail in the genome, closer examination of these intervening DNA regions has revealed the presence of unique genetic elements that are found in a number of different locations. One of the first such repeating elements identified was *Alu*.

Alu repeats are approximately 300 base pairs in length. They got their name from the fact that most carry within them the base sequence AGCT; the recognition site for the *Alu* I **restriction endonuclease**, a type of enzyme that cuts DNA at a specific site. There are over 500,000 *Alu* repeats scattered throughout the human genome. On average, one can be found every 4,000 base pairs along a human DNA molecule. How they arose is still a matter

of speculation but evidence suggests that the first one may have appeared in the genome of higher primates about 60 million years ago. Approximately every 100 years since then, a new *Alu* repeat has inserted itself in an additional location in the human genome. *Alu* repeats are inherited in a stable manner; they come intact in the DNA your mother and father contributed to your own genome at the time you were conceived. Some *Alu* repeats are fixed in a population, meaning all humans have that particular *Alu* repeat. Others are said to be **dimorphic**; different individuals may or may not carry a particular *Alu* sequence at a particular chromosomal location.

The Polymerase Chain Reaction

Objectives:

1. You should be able to list and explain the importance of each component of PCR and compare it to cellular DNA replication.
2. You should be able to associate the temperature changes with the cycling steps of PCR.

The polymerase chain reaction (PCR) is a method used by scientists to rapidly copy, in a test tube, specific segments of DNA. By mimicking some of the DNA replication strategies employed by living cells, PCR has the capacity for churning out millions of copies of a particular DNA region. It has found use in forensic science, in the diagnosis of genetic disease, and in the cloning of rare genes. One of the reasons PCR has become such a popular technique is that it doesn't require much starting material. It can be used to amplify DNA recovered from a plucked hair, from a small spot of blood, or from the back of a licked postage stamp.

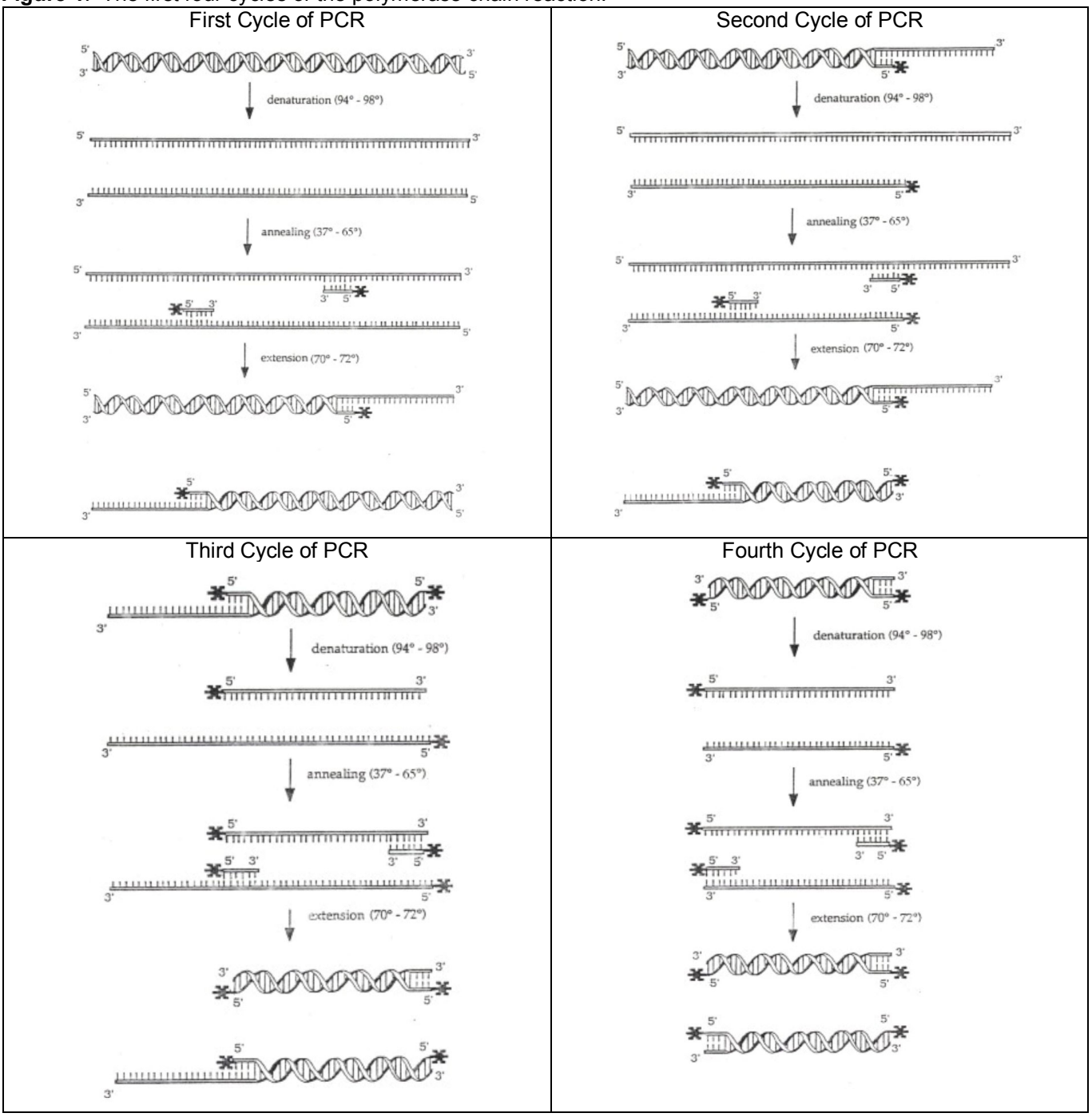
There are some essential reaction components and conditions needed to amplify DNA by PCR. First and foremost, it is necessary to have a sample of DNA containing the segment you wish to amplify. This DNA is called the **template** because it provides the pattern of base sequence to be duplicated during the PCR process. Along with template DNA, PCR requires two short single-stranded pieces of DNA called **primers**. These are usually about 20 bases in length and are complementary to opposite strands of the template at the ends of the target DNA segment being amplified. Primers attach (**anneal**) to their complementary sites on the template and are used as initiation sites for synthesis of new DNA strands. **Deoxynucleoside triphosphates** containing the bases A, C, G, and T are also added to the reaction. The enzyme **DNA polymerase** binds to one end of each annealed primer and strings the deoxynucleotides together to form new DNA chains complementary to the template. The DNA polymerase enzyme requires the metal ion magnesium ($\mathbf{Mg^{++}}$) for its activity. It is supplied to the reaction in the form of $MgCl_2$ salt. A **buffer** is used to maintain an optimal pH level for the DNA polymerase.

PCR is accomplished by cycling a reaction through several temperature steps. In the first step, the two strands of the template DNA molecule are separated, or **denatured**, by exposure to a high temperature (usually 94° to 96°C). Once in a single-stranded form, the bases of the template DNA are exposed and are free to interact with the primers. In the second step of PCR, called **annealing**, the reaction is brought down to a temperature usually between 37˚C to 65 ˚C. At this lower temperature, stable hydrogen bonds can form between the complementary bases of the primers and template. Although human genomic DNA is billions of base pairs in length, the primers require only seconds to locate and anneal to their complementary sites. In the third step of PCR, called **extension**, the reaction temperature is raised to an intermediate level (65˚C to 72˚C). During this step, the DNA polymerase starts adding nucleotides to the ends of the annealed primers. These three phases are repeated over and over again, doubling the number of DNA molecules with each cycle. After 25 to 40 cycles, millions of copies of DNA are produced. The PCR process taken through four cycles is illustrated on the following page (Figure 1).

In the following laboratory exercise, you will use PCR to amplify a dimorphic *Alu* repeat (designated "*Alu*" PV92), that, if you have it, will be found on your number 16 chromosome. You will use your own DNA as template for this experiment. DNA is easily obtained from the human body. A simple saltwater mouth rinse will release cheek cells, from which you will extract the DNA. After you amplify the *Alu* repeat region, you will determine whether or not you carry this particular *Alu* sequence on one or both or none of your number 16 chromosomes. This will be accomplished by separating the DNA in your PCR sample on an agarose gel via electrophoresis, a process that separates DNA by size. Finally, using a program developed by the DNA Learning Center at Cold Spring Harbor Laboratory, you will determine how rare this *Alu* sequence is in the human population and make some assessment as to when and where it arose.

Illustration of the Polymerase Chain Reaction

Figure 1. The first four cycles of the polymerase chain reaction.



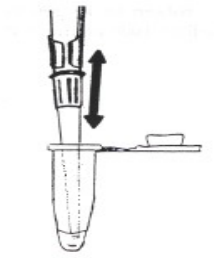
An excellent animated tutorial showing the steps of PCR is available at the DNA Learning Center web site.
http://www.dnalc.org/ddnalc/resources/pcr.html

Note: You will need Macromedia Flash plug-in to view this on-line and to download the animation files to you computer.

Laboratory Exercise

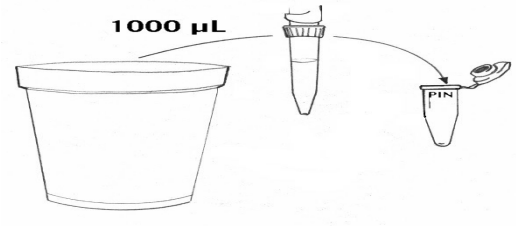
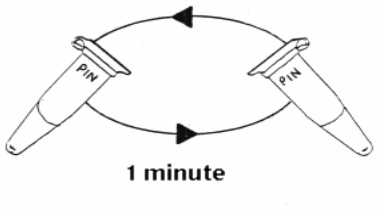
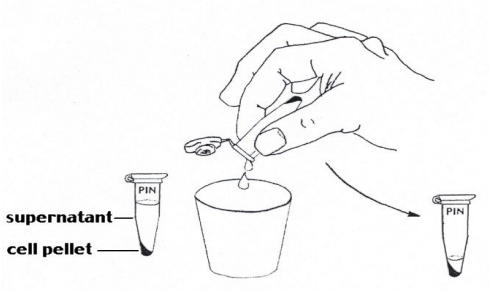
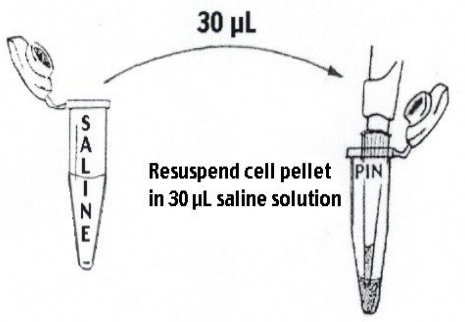
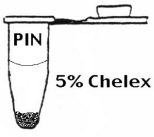
Objectives:

1. You should be able to successfully isolate your DNA from cheek cells.
2. You should be able to prepare a reaction for PCR amplification of an *Alu* insert.

IMPORTANT LABORATORY PROCEDURES	
a. Add reagents to the bottom of the reaction tube, not to its side. b. You should add each additional reagent directly into previously-added reagent and pipet the combined liquid up and down several times to ensure proper mixing. c. Pipet slowly to prevent contaminating the pipette barrel. 	a. Change pipette tips between each delivery. b. You should change the tip even if it is the same reagent being delivered between tubes. 

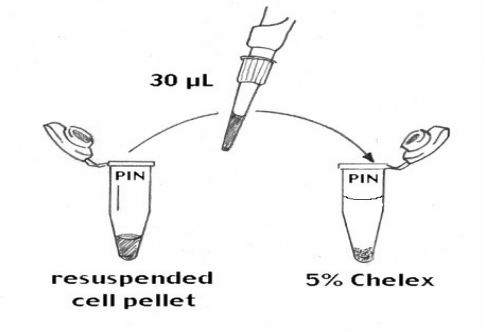
*** Place a check mark in the box of each step as it is completed.**

DNA Preparation Using a Saline Mouthwash	
1. Vigorously swirl 10 mL of saline solution in your mouth for 30 seconds. **Note:** The saline solution is a 0.9% NaCl solution, the salt concentration of your blood plasma.	 ☐
2. Expel saline into a cup and swirl to mix the cells.	 ☐
3. **TUBE #1:** Label a microfuge tube with your PIN.	 ☐

4. Transfer 1000 µL (= 1 mL) of the saline/cell suspension into the labeled microfuge tube [**TUBE #1]**.	 ☐
5. In a microcentrifuge, spin your saline cell suspension for 1 minute. Be sure to have another student's sample as a balance.	 ☐
6. Observe your cell pellet at the bottom of the tube. Pour off the supernatant into your cup, being careful NOT to lose your cell pellet. **Note**: It is okay if some supernatant is left in the tube.	 ☐
7. Add 30 µL of fresh saline to your **TUBE #1** and resuspend your cell pellet. Make sure the entire cell pellet is thoroughly mixed by vortexing, or "racking" your tube. **Note**: To "rack" your sample, be sure the top of the tube is closed, hold tube firmly at the top, and pull it across a microfuge rack 2-3 times.	 ☐
8. **TUBE #2:** Obtain a tube of Chelex from your instructor. Label with your PIN.	 ☐

9. Withdraw 30 µL of your cell suspension [**TUBE #1**] and add it to the tube containing Chelex [**TUBE #2**].

Note: Do not pipet up and down at this step or else you will clog the tip with Chelex beads.



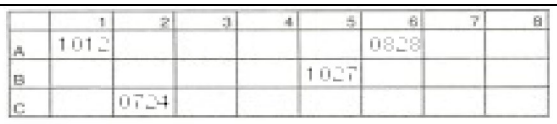
☐

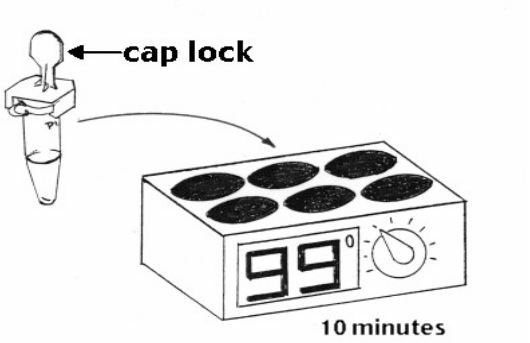
10. **PCR tube version**: If your Chelex (with your cell suspension) is in a tiny PCR tube, follow your teacher's instruction on placing it in a thermal cycler at 99ºC, for 10 minutes.

Note: Remember to record the location of your tube in the thermal cycler.

Heat block version: If your Chelex (with the cell suspension) is in a normal 1.5 mL microfuge tube, take your tube to a heat block station. Slide a cap-lock onto the tube lid and place it in the heat block for 10 minutes. Keep track of your tube in the heat block.

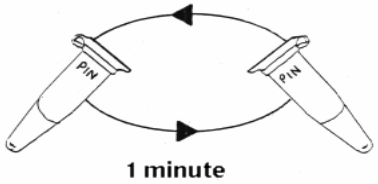
.





☐

11. After heating, shake your tube [**TUBE #2]** well, and then place it in a centrifuge to spin for 1 minute. Be sure to use another student's tube as a balance.



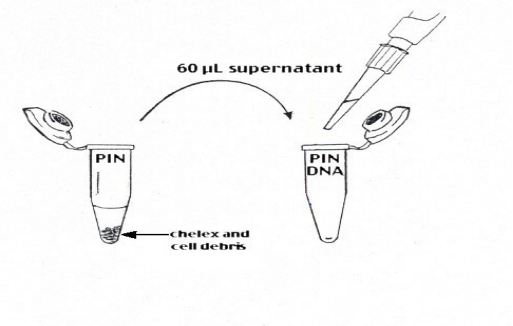
☐

12. **TUBE #3:** Obtain another clean microfuge tube and label with your PIN. Also write "DNA" on this tube.

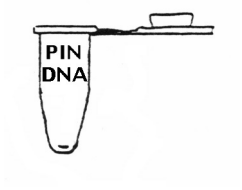
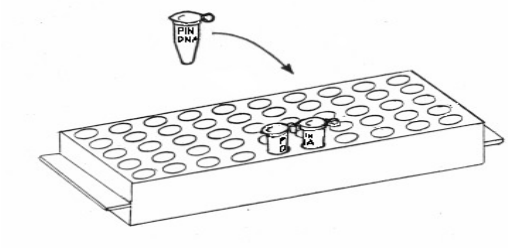
☐

13. Holding your tube at eye level, use a p200 to withdraw 60 µL of supernatant from the chelex tube [**TUBE #2**] to the new, labeled tube [**TUBE #3**]. Be sure NOT to transfer any Chelex beads.

Note: This is your stored "DNA" sample.



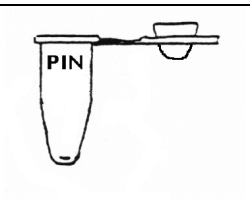
☐

14. Have someone check your tube [**TUBE #3**] to be sure that no Chelex beads were transferred to the "DNA" tube. Make sure there are NO Chelex beads present.	 ☐
15. Place your DNA tube [**TUBE #3**] in the class rack. Your teacher will refrigerate your isolated DNA until you are ready to prepare your PCR amplification.	 ☐

Polymerase Chain Reaction

1. Obtain a tiny PCR tube. Label it with your 4- digit PIN, just under the lip of the tube.

   Note: Keep your PCR tube on ice when setting up the reaction.

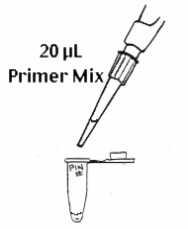


☐

2. Pipet **20 µL of Master Mix** into your PCR tube.

☐

3. Change your pipet tip and add **20 µL of Primer Mix** into your PCR tube.

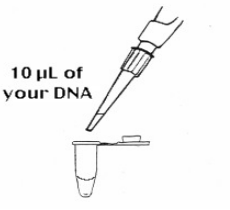


☐

4. With a new pipet tip, add **10 µL of your extracted DNA** [**TUBE #3]** into your PCR tube.

   Note: Slowly pipet up and down several times to mix all the reagents in your reaction tube.

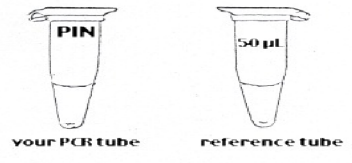
   What is the total volume in your tube? **___ µL.**



☐

5. Check the volume of your PCR tube by comparing it to a reference PCR with 50 µL in it. This should be near the thermal cycler, set by your teacher.

   Note: If the volume of your tube does not match, see your instructor to troubleshoot.



☐

6. Place your reaction into the thermal cycler and record the location of your tube on the grid provided by your teacher.

	1	2	3	4	5	6	7	8
A	1012					0828		
B					1027			
C		0724						

☐

7. Two students will be asked to set up the positive control reactions for the class.

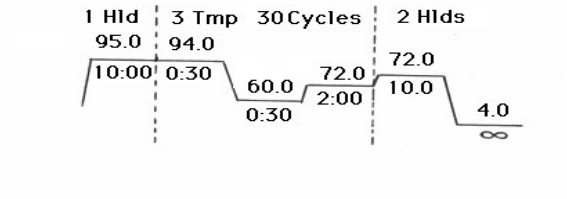
   Another two students will set up negative control reactions for the whole class.

Control	Master Mix	Primer mix	DNA
+	20 µL	20 µL	10 µL +C DNA
-	20 µL	20 µL	10 µL sterile water

☐

8. The cycling protocol for amplification of *Alu* PV92:

95°C—10 minutes
94°C—30 seconds
60°C—30 seconds } 30 cycles
72°C—2 minutes
72°C—10 minutes
4°C—hold, ∞ infinity

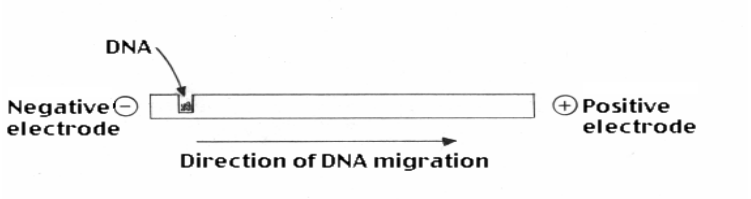


☐

Agarose Gel Electrophoresis

To determine whether or not you carry the *Alu* repeat, you will need to visualize the products of your amplification. This will be done using a process called **gel electrophoresis** in which electric current forces the migration of DNA fragments through a special gel material. Since DNA is negatively charged, it will migrate in an electric field towards the positive electrode (Figure 2). When electrophoresed through a gel, shorter fragments of DNA move at a faster rate than longer ones. The *Alu* repeat adds 300 base pairs of length to a DNA fragment and thus will slow its migration during electrophoresis.

Figure 2. Side view of an agarose gel showing DNA loaded into a well and the direction of DNA fragment migration during electrophoresis.



The gel material to be used for this experiment is called **agarose.** When agarose granules are placed in a buffer solution and heated to boiling temperatures, they dissolve and the solution becomes clear. A casting tray is set up with a comb to provide a mold for the gel. The agarose is allowed to cool slightly and is then poured into the casting tray. Within about 15 minutes, the agarose solidifies into an opaque gel having the look and feel of coconut Jell-O. The gel, in its casting tray, is placed in a buffer chamber connected to a power supply and buffer is poured into the chamber until the gel is completely submerged. The comb can then be pulled out to form the wells into which your PCR sample will be loaded.

Loading dye is a colored, viscous liquid containing dyes (making it easy to see) and sucrose, ficoll, or glycerol (making it dense). You will add loading dye to your amplification reaction and then pipet an aliquot of the mixture into one of the wells of your agarose gel. When all wells have been loaded with sample, you will switch on the power supply. The samples should be allowed to electrophorese until the blue loading dye is 1 to 2 cm from the bottom. The gel can then be stained with ethidium bromide and photographed.

Calculations for Preparing 2% Agarose gel

You will need a 2%, by mass, agarose gel for electrophoresis of your PCR products. If your agarose gel casting trays holds 50 mL, then how much agarose and buffer would you need?
Definition of % in biology is usually grams (mass) / 100mL (volume). Therefore, for
2 % agarose, it will be 2 g /100 mL buffer.

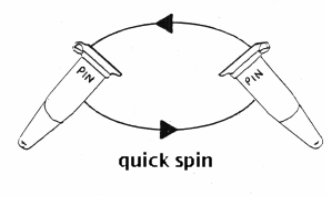
Step 1: Calculate mass of agarose needed for 50 mL total volume of agarose solution.

$$\frac{2\text{ g}}{100\text{ mL}} = \frac{X\text{ g}}{50\text{ mL}} \quad ; \quad X = 1\text{ gram}$$

Step 2: Calculate amount of buffer needed to bring the agarose solution to 50 mL. By standard definition, 1 gram of $H_2O \approx 1$ mL of H_2O. The amount of buffer for the 2% agarose solution will be 49 mL (50 mL – 1 mL (1 gram of agarose))

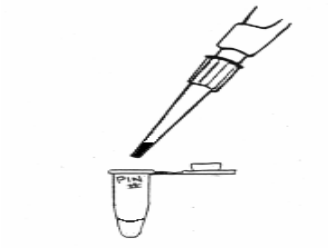
Electrophoresis of Amplified DNA

1. Retrieve your PCR tube and spin it briefly (~ 10 seconds) to bring the liquid to the bottom of the reaction tube. **Make sure the centrifuge is balanced before you begin spinning your sample!**



☐

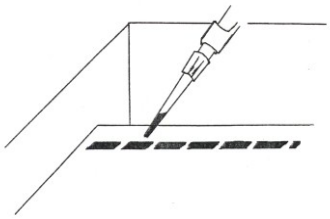
2. Add **5 µL** of loading dye to your PCR tube. Slowly pipet the mixture up and down until the contents in the tube are uniformly colored.



☐

3. Carefully load **15 or 20 µL** of your reaction into a well in your gel. Have students load 20 µL in 6-8 well gels and 12 µL in 10-12 well gels.

Note: Avoid poking the pipette tip through the bottom of the gel or spilling sample over the sides of the well. Use a new tip for each sample.

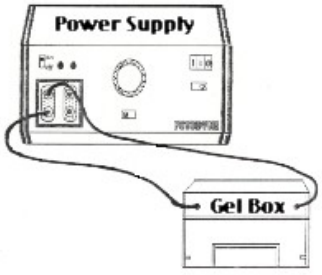


☐

4. One student (or the instructor) should load **5 µL** of 100 bp ladder (molecular weight marker) into one of the wells of each gel.

☐

5. When all samples are loaded, attach the electrodes from the gel box to the power supply. Have your teacher check your connections and then electrophorese your samples at **150 Volts** for **35 - 45 minutes**.



☐

6. After electrophoresis, the gels will be ready to stain and photograph.

☐

Staining and Photographing Agarose Gels

Your teacher will stain your agarose gel and take a photograph for you so that you may analyze your *Alu* results. Gel staining is done as follows.

Place the agarose gel in a staining tray. Pour enough ethidium bromide (0.5µg/mL) to cover the gel. Wait **20** minutes.

CAUTION: Ethidium bromide is a carcinogen. Always wear gloves and safety glasses when handling.

Pour the ethidium bromide solution back into its storage bottle. Pour enough water into the staining tray to cover the gel. Wait 5 minutes. Pour the water out of the staining tray into a hazardous waste container and place the stained gel on a UV light box.

CAUTION: Ultraviolet light can damage your eyes and skin. Always wear protective clothing and UV safety glasses when using a UV light box.

Place the camera over the gel and take a photograph.

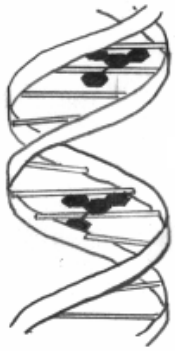


Figure 3. Ethidium bromide molecules stacked between DNA base pairs.

The PCR products run on your agarose gel are invisible to the naked eye. If you look at your gel in normal room light, you will not be able to see the amplified products of your reaction. In order to "see" them, we must stain the gel with a fluorescent dye called **ethidium bromide**. Molecules of ethidium bromide are flat and can nestle between adjacent base pairs of double stranded DNA (Figure 3). When this interaction occurs, they take on a more ordered and regular configuration causing them to fluoresce under ultraviolet light (UV). Exposing the gel to UV light after staining, allows you to see bright, pinkish-orange bands where there is DNA (Figure 4).

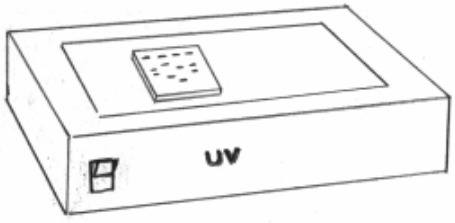


Figure 4. After staining an agarose gel with ethidium bromide, DNA bands are visible upon exposure to UV light.

Results

By examining the photograph of your agarose gel, you will determine whether or not you carry the *Alu* repeat on one, both, or neither of your number 16 chromosomes. PCR amplification of this *Alu* site will generate a 415 bp fragment if the repeat is not present. If the repeat is present, a 715 bp fragment will be made. Figure 5 shows the structure of an individual's two number 16 chromosomes in a case where one carries the *Alu* repeat and the other does not.

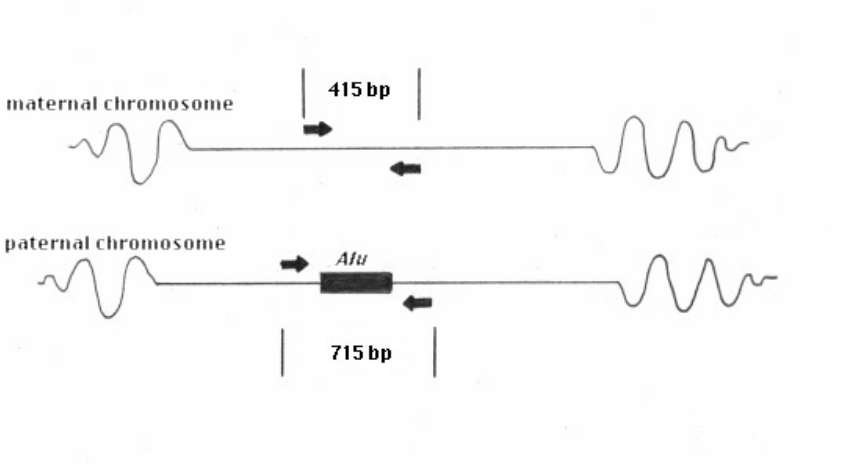


Figure 5. The chromosomes you inherit from your parents may or may not carry the *Alu* repeat on Chromosome 16.

When you examine the photograph of your gel, it should be readily apparent that there are differences between people at the level of their DNA. Even though you amplified only one site, a site that every one has in their DNA, you will notice that not all students have the same pattern of bands. Some students will have only one band, while others will have two.

We use the term **allele** to describe different forms of a gene or genetic site. For those who have the *Alu* repeat (they have at least one 715 bp band), we can say that they are positive for the insertion and denote that allele configuration with a "+" sign. If the *Alu* repeat is absent (a 415 bp band is generated in the PCR), we assign a "-" allele designation. If a student has a single band, whether it is a single 415 bp band or a single 715 bp band, then both their number 16 chromosomes must be the same in regards to the *Alu* insertion. They are said to be **homozygous** and can be designated with the symbols "-/-" or "+/+", respectively. If a student's DNA generates a 415 bp band and an 715 bp band during PCR, the student is said to be **heterozygous** at this site and the designation "+/-" is assigned. A person's particular combination of alleles is called their **genotype**. See the table below for a quick summary of the allele designations.

Possible bands	Allele designation	Genotype	*Alu* insert
1. One band at 415 bp	-/-	homozygous	No *Alu* insert
2. One band at 715 bp	+/+	homozygous	*Alu* insert present
3. One band at 415 bp and a second band at 715 bp.	+/-	heterozygous	*Alu* insert on one of the chromosome 16.

Figure 6 below, shows a representation of a possible experimental outcome on a gel, where all possible allele combinations have been generated.

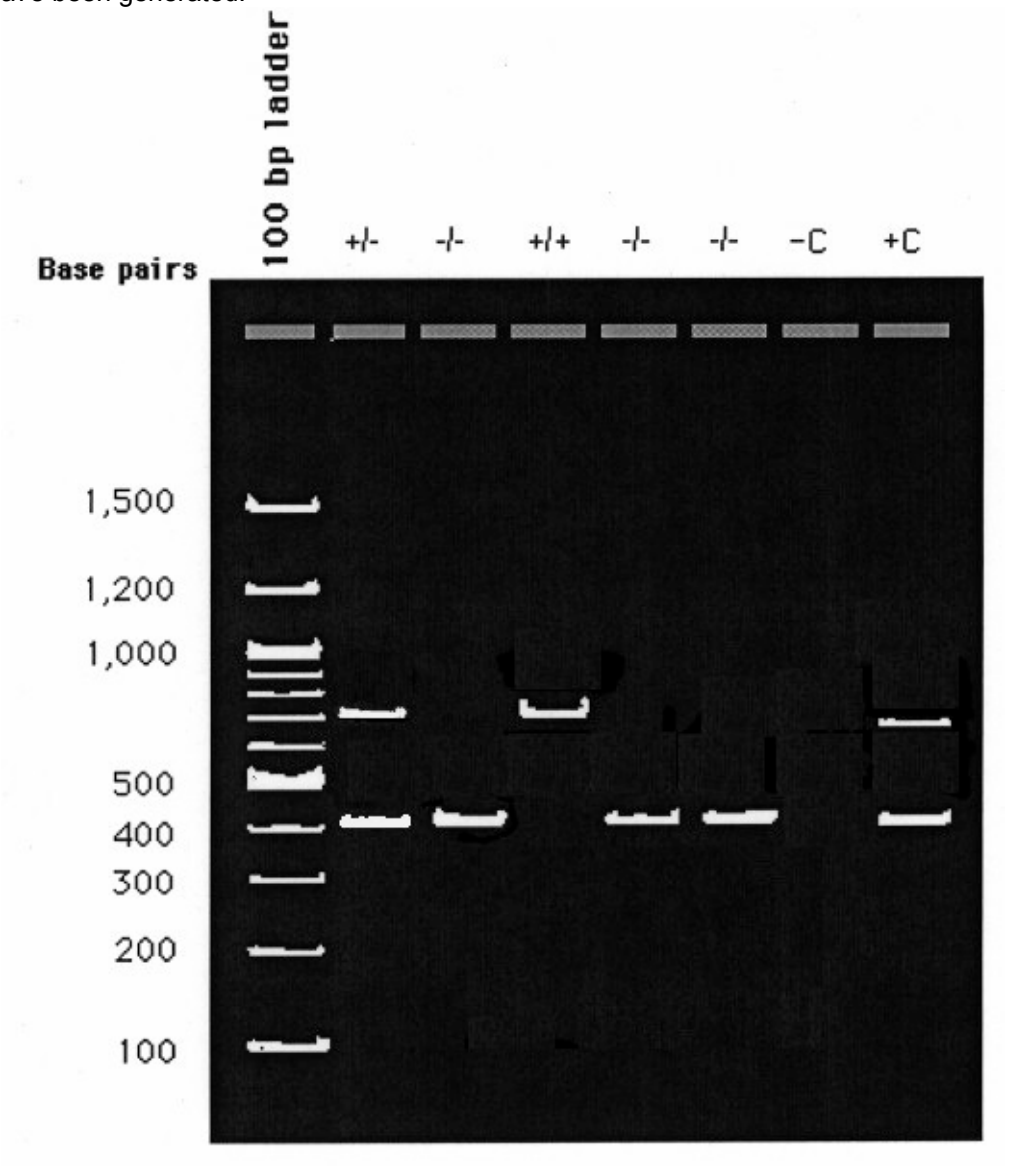


Figure 6. Agarose gel of homozygous and heterozygous individuals for the PV92 *Alu* insertion. A 100 base pair ladder is loaded in the first lane and is used as a size marker, where these bands differ by 100 bp in length. The 500 bp band and the 1,000 bp band are purposely spiked to be more intense than are the other bands of the ladder when stained with ethidium bromide. The next 5 lanes contain the results of homozygous and heterozygous individuals. A negative control (-C) does not contain any template DNA and should therefore contain no bands. The positive control (+C) is heterozygous for the *Alu* insertion; it contains both the 415 bp and 715 bp bands.

Biostatistics Activity 1: Calculating Allele and Genotype Frequencies

Objectives:

1. You should be able to calculate allele frequencies.
2. You should be able to calculate genotype frequencies.

Allele Frequencies

Within your class, how unique is your particular combination of *Alu* alleles? By calculating an allele frequency**,** you can begin to answer this question. An **allele frequency** is the percentage of a particular allele within a population of alleles. It is expressed as a decimal. You can calculate an allele frequency for the *Alu* PV92 insertion in your class by combining all your data. For example, imagine that there are 100 students in your class and the genotype distribution within the class is as follows:

Genotype	Number of Students having that Genotype
+/+	20
+/-	50
-/-	30

Since each person in your class has two number 16 chromosomes (they are diploid for chromosome 16), there must be twice as many total alleles as there are people:

$$\frac{2 \text{ alleles}}{\text{student}} \times 100 \text{ students} = 200 \text{ alleles}$$

To calculate allele frequencies for the class, therefore, 200 will be used as the denominator value. To calculate the "+" allele frequency, we must look at all those students who have a "+" in their genotype. There are 20 students who are "+/+"; they are homozygous for the insertion. Since these 20 students have two copies of the *Alu* insert on their chromosomes, they contribute 40 "+" alleles to the overall frequency:

$$\frac{2 \text{ "+" alleles}}{\text{homozygous "+/+" student}} \times 20 \text{ homozygous "+" students} = 40 \text{ "+" alleles}$$

There are 50 students heterozygous ("+/-") for the *Alu* insertion. Each heterozygous individual, therefore, contributes one "+" allele to the overall frequency, or 50 "+" alleles. Adding all "+" alleles together gives us:

$$\begin{array}{l} 40 \text{ "+" alleles from the homozygotes} \\ + \ 50 \text{ "+" alleles from the heterzygotes} \\ \hline 90 \text{ "+" alleles} \end{array}$$

The frequency of the "+" allele in this class, therefore, is:

$$\frac{90 \text{ "+" alleles}}{200 \text{ total alleles}} = 0.45$$

The frequency for the PV92 "-" allele is calculated in a similar manner. There are 30 students homozygous for the "-" allele. This group, then, contributes 60 "-" alleles to the frequency. There are 50 students heterozygous for the *Alu* insertion. They contribute 50 "-" alleles to the frequency. Adding all "-" alleles together gives us:

60 "-" alleles from the homozygotes
\+ 50 "-" alleles from the heterozygotes
110 "-" alleles

The frequency of the "-" allele in this class, therefore, is

$$\frac{\text{110 "-" alleles}}{\text{200 total alleles}} = 0.55$$

Notice that the sum of the frequencies for the "+" and "-" alleles is 1.0.

0.45 "+" allele frequency
\+ 0.55 "-" allele frequency
1.00

If the allele frequencies do not add up to 1.0, then you have made an error in the math.

Use the spaces below to calculate the "+" and "-" allele frequencies for your class.

Number of total alleles:

$$\frac{\text{2 alleles}}{\text{student}} \times \_\_\_\_\_ \text{ students} = \_\_\_\_\_ \text{ alleles}$$

Number of "+" and "-" alleles:

Genotype	Number of Students	Number of "+" Alleles	Number of "-" Alleles
+/+			**0**
+/-			
-/-		**0**	
Total			

Allele frequencies:

$$\text{"+" allele frequency} = \frac{\_\_\_\_\_ \text{ total "+" alleles}}{\_\_\_\_\_ \text{ total alleles}} = \_\_\_\_\_$$

$$\text{"-" allele frequency} = \frac{\_\_\_\_\_ \text{ total "-" alleles}}{\_\_\_\_\_ \text{ total alleles}} = \_\_\_\_\_$$

Do these allele frequencies add up to 1.00? _____

Genotype Frequencies

How does the distribution of *Alu* genotypes in your class compare with the distribution in other populations? For this analysis, you need to calculate a **genotype frequency**, the percentage of individuals within a population having a particular genotype. Remember that the term *allele* refers to one of several different forms of a particular genetic site whereas the term *genotype* refers to the specific alleles that an organism carries. You can calculate the frequency of each genotype in your class by counting how many students have a particular genotype and dividing that number by the total number of students. For example, in a class of 100 students, let's say that there are 20 students who have the "+/+" genotype. The genotype frequency for "+/+", then, is 20/100 = 0.2. Given the ethnic makeup of your class, might you expect something different? How can you estimate what the expected frequency should be?

If within an infinitely large population no mutations are acquired, no genotypes are lost or gained, mating is random, and all genotypes are equally viable, then that population is said to be in **Hardy-Weinberg equilibrium**. In such populations, the allele frequencies will remain constant generation after generation. Genotype frequencies within this population can then be calculated from allele frequencies by using the equation:

$$p^2 + 2pq + q^2 = 1.0$$

where p and q are the allele frequencies for two alternate forms of a genetic site. The genotype frequency of the homozygous condition is either p^2 or q^2 (depending on which allele you assign to p and which to q). The heterozygous genotype frequency is 2pq.

Let's use our fictitious class again (see page 16) to calculate expected genotype frequencies. We determined the following allele frequencies (we will assign p to the "+" allele and q to the "-" allele):

$$p = 0.45$$
$$q = 0.55$$

We expect, therefore, that the genotype frequency for "+/+" is equal to p^2 which is

$$p^2 = (0.45)^2 = 0.2025$$

The frequency for the "+/-" genotype is

$$2pq = 2(0.45)(0.55) = 0.495$$

The frequency for the "-/-" homozygous genotype is expected to be

$$q^2 = (0.55)^2 = 0.3025$$

To convert these decimal numbers into numbers of students, we multiply each by the total number of students. Since there are 100 students in this fictitious class, the number of students in the class expected to have the "+/+" genotype is

$$100 \times 0.2025 = 20.25 \text{ students who should be "+/+"}$$

The number of students who should be "+/-" is

$$100 \times 0.495 = 49.5$$

The number of students who should be "-/-" is

$$100 \times 0.3025 = 30.25$$

On page 16, you calculated the allele frequencies found in your class. Use these frequencies to determine the expected class genotype frequencies. (Let p represent the "+" allele and q the "-" allele.)

Expected "+/+" genotype frequency:

p^2 = _____ = _____

Expected "+/-" genotype frequency:

2pq = _____ = _____

Expected "-/-" genotype frequency:

q^2 = _____ = _____

Use the table below to calculate how many students in your class should have each genotype.

Genotype	Expected Genotype Frequency	Total Number of Students in Class	Expected Number of Students with Specific Genotype
+/+			
+/-			
-/-			

Now, calculate the **actual** genotype frequencies for this class (hint: use data on page 16).

Actual "+/+" genotype = _____
Actual "+/-" genotype = _____
Actual "-/-" genotype = _____

Name_____

Review Questions: Allele and Genotype Frequencies

1. A class is looking at a dimorphic *Alu* insert on chromosome number 3. How many total alleles are there in a class of 34 students for this *Alu* site?

2. The "-" allele frequency for the class is 0.3. What is the "+" allele frequency?

3. A class in Hardy-Weinberg equilibrium has a "+/+" genotype frequency of 0.64. What is the "+" allele frequency?

4. The "+/+" genotype frequency for a class is 0.49 and the "-/-" genotype frequency is 0.09. What is the "+/-" genotype frequency if the class is in Hardy-Weinberg equilibrium?

Biostatistics Activity 2: Using the Allele Server at the CSHL DNA Learning Center

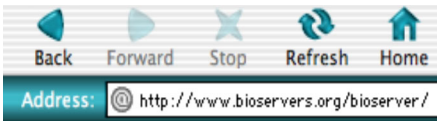
Objectives:

1. You should be able to view your class data in the Cold Spring Harbor Laboratory *Alu* PV92 database.
2. You should be able to plot the frequency of the *Alu* PV92 insert for different world populations on a map.
3. You should be able to formulate a hypothesis describing the origin and spread of the *Alu* PV92 insert across the globe.
4. You should be able to use Chi Square analysis to determine differences between your class data and data sets from other human populations.
5. You should be able to brainstorm reasons for genetic similarities and differences between populations.

The DNA Learning Center at Cold Spring Harbor Laboratory has developed a number of **bioinformatics** tools for student use. Bioinformatics tools are computer programs used to help scientists make sense of biological data and solve biological problems. You will be using the Allele Server for four different activities to help you learn more about *Alu* PV92 in human populations. In the first exercise, you will check your calculations for allele and genotype frequencies. Next, you will access data from a database to plot the "+" allele frequency from different world populations. In the third activity you will check to see if your class could be in Hardy-Weinberg equilibrium. The last activity includes computer analysis to find human populations similar to and different from your class.

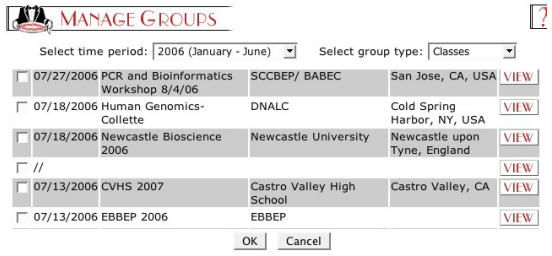
Part 1: Using the Allele Server to Check Your Allele and Genotype Frequencies

Your teacher has added your class results into a database at the Cold Spring Harbor Laboratory using the Allele Server program. In this activity you will also use the Allele Server to access your class data so that you can check your allele and genotype calculations.

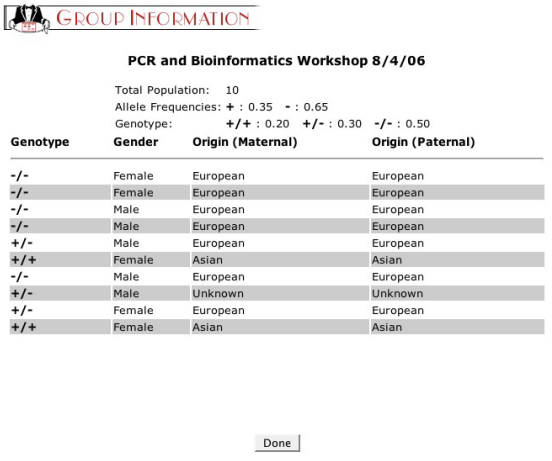
Looking at Allele and Genotype Frequencies	
1. Open up an Internet browser window. (This might be Internet Explorer, Safari, Firefox, or Netscape Navigator, etc.)	
2. In the internet address box, type in the following URL: http://www.bioservers.org/bioserver The Allele Server main page will be brought up.	
3. In the Bioservers page, type in your username and password in the **Allele Server** (in the middle of the field) and then click **LOGIN**. If you do not have a username and password, you must first click the **REGISTER** button and complete the registration form.	

4. Two windows will appear. The top pop-up window, “Using Allele Server”, provides a quick reference guide to navigate the Allele Server. You can either close the “Using Allele Server” popup or just click on the second, larger window, to bring it to the foreground. This window is the Allele Server workspace. Click on the **MANAGE GROUPS** button. This will bring up the MANAGE GROUPS window.

5. In the upper right hand corner of the MANAGE GROUPS window, is a self-scroll menu bar. This will automatically default and load “Classes” Otherwise, select “**Classes**” from this menu. Once you are in the “Classes” window, scroll down to find your class. Click the **VIEW** button on the right hand-side.



6. The GROUP INFORMATION window will show the data for your class.
   a. In the spaces provided, fill in the allele and genotype frequencies for your class as displayed in the GROUP INFORMATION window.
   “+” Allele Frequency: _____
   “-“ Allele Frequency: _____ “+/+” Genotype Frequency: _____ “+/-“ Genotype Frequency: _____ “-/-“ Genotype Frequency: _____
   b. Do these allele and genotype frequencies match the actual allele and genotype frequencies that you calculated (page 16 and 18 in manual)? _____ If you answered “No”, check your calculations again.
   Close the window by clicking the small close box at its top. **Do not** click the “Done” button. This will return you to the MANAGE GROUPS window.



Part 2: Using the Allele Server to Look at Different World Populations

Humans, monkeys, mice, canines, and corn have them running rampant through their genomes. They move undetectably from chromosome to chromosome. They are the so-called "jumping genes" and the *Alu* element is one of them. Although *Alu* is found only in primates, there are other related "jumping genes" that have found their way into the DNA of most eukaryotic organisms on Earth.

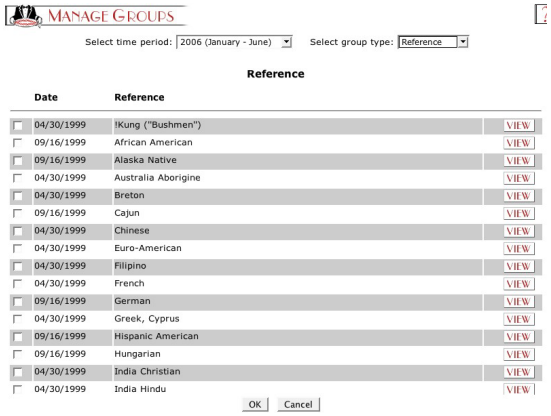
Alu is classified as a **retroposon** - a genetic element that uses the enzyme reverse transcriptase to copy itself from one chromosomal location to the next. For this reason, calling *Alu* a "jumping gene" can be misleading. *Alu* doesn't "jump" in the sense that it leaves one location to occupy another. When *Alu* moves, it leaves a copy of itself behind.

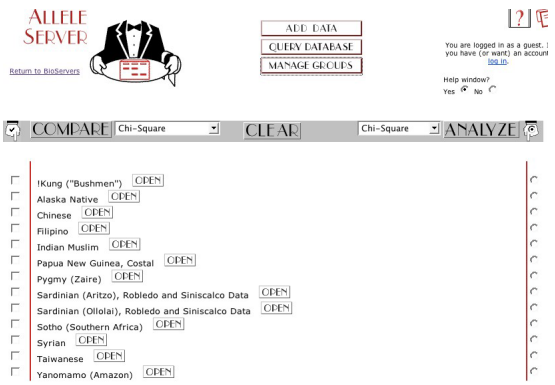
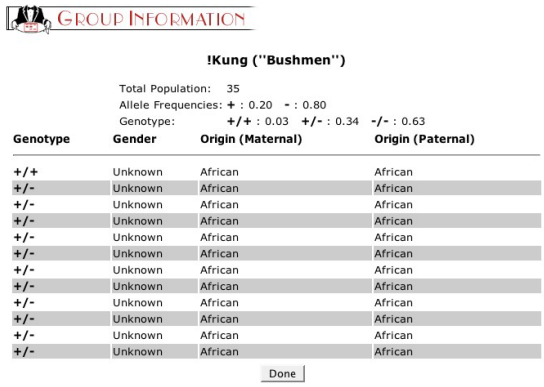
The *Alu* element first appeared tens of millions of years ago and since that time; it has been increasing within our genome at the rate of about one copy every 100 years. It is difficult to tell how *Alu* arose. It shows a striking similarity to a gene (called 7SL RNA) that performs a vital function in our metabolism. But *Alu*, it seems for now, has no specific function. It is self-serving and, like a parasite, takes advantage of us for its own replication without providing us any advantage to our own survival.

Most *Alu* elements are "fixed"; they are found at the same chromosomal site in every person on the planet. Fixed *Alu* elements must have arisen very early in our evolution, well before Homo sapiens appeared. When modern humans did arise some 200,000 years ago, the vast majority of our *Alu* insertions came to us already intact in our DNA. The *Alu* PV92 insertion, however, is not fixed. This insertion may or may not be present on one or both of a person's number 16 chromosomes. Since not everyone has the *Alu* PV92 element, it must have arisen after the initial human population began growing. It is a widely held belief that modern humans originated in Africa and then disseminated across the planet. Did the *Alu* PV92 insert arise in Africa or on some other continent during our spread across the globe? In the following exercise, you will plot the "+" allele frequencies for various populations on a world map and make some determination as to where this *Alu* arose and how it might have spread across continents.

Plotting the *Alu* PV92 Insert on the World Map

1. Return to the **MANAGE GROUPS** window. Wait for the MANAGE GROUPS window to load the classes before going ahead and selecting "**Reference**" from the popup menu, on the upper right corner of the page. Click on the boxes to the left of at least 10 population groups (from representative places), around the world. When done, click the **OK** button. This will place these groups in the Allele Server workspace – the original startup page.



2. Click on the **OPEN** button to the right of the population group of interest in this list. This will bring up the window displaying details about the population.	
3. Record the "+" allele frequency for that population on the world map provided. Close the window by clicking the close box at the top of the window and open the next population group. Record that "+" allele frequency on your map. Repeat this process for all the population groups listed on your workspace. When you are finished, clear the workspace by checking each box to the left of each population group and then pressing the CLEAR button.	

Name_____

RIGHT-CLICK to zoom in.
Map will resize with browser window.
1. African American
2. Alaskan Natives
3. Australian
4. Breton
5. Cajun
6. Chinese
7. Euro-American
8. Filipino
9. French
10. German
11. Greek, Cyprus
12. Hispanic American
13. Hungarian
14. India Christians
15. India Hindu
16. Indian Muslim
17. Italian, Northern
18. Java
19. !Kung ("Bushmen")
20. Malay
21. Maya
22. Moluccan
23. Mvskoke
24. Nguni
25. Nigerian
26. Pakistani, Batzer Data
27. Papua New Guinea
28. Papua New Guinea, Costal
29. Pushtoon (Afgani)
30. Pygmy (Central African Republic)
31. Pygmy (Zaire)
32. Sardinian (Artizo)
33. Sardinian (Marrubiu)
34. Sardinian (Ollolai)
35. Sardinian (San Teodoro)
36. Sotho
37. South India
38. Swiss
39. Syrian
40. Taiwanese
41. Turkish, Cyprus
42. United Arab Emirates
43. Yanomamo
http://www.dnalc.org/map.html

Name_____

Review Question: Plotting *Alu* PV92

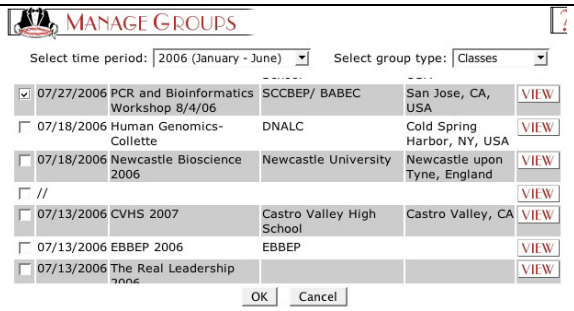
Look at the "+" allele frequencies for the various world populations that you entered on your world map. Do you notice any patterns? Formulate an explanation for where you believe the *Alu* PV92 insert originated and how it spread throughout different world populations. You may use the following map.

Part 3: Using Allele Server to Test if Your Class is in Hardy-Weinberg Equilibrium

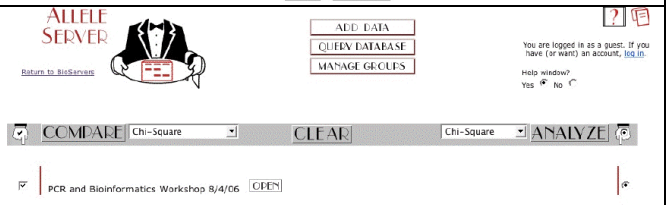
On page 18, you calculated the expected genotype frequencies for your class using the Hardy-Weinberg equation. Are the expected genotype frequencies you calculated similar to the actual class frequencies? If they are different, then it may mean that the population in your class is not in Hardy-Weinberg equilibrium. If we do observe differences, how can we account for them? How do we even know when there is actually a significant difference between the observed genotype frequencies and the expected genotype frequencies? You will use the Allele Server program to address these questions.

Chi Square Analysis of Your Class Data

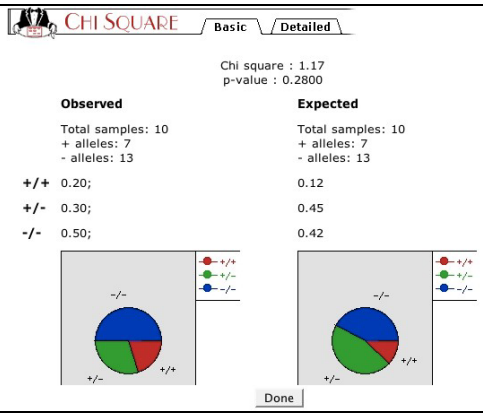
1. In the MANAGE GROUPS "Classes" window, locate you class and place a check mark in the box to its left. Click the OK button at the bottom of the window. This will bring you back to the ALLELE SERVER workspace window. Your class data will have been placed in the workspace.



2. Click on the box to the left of your class name in the workspace. In the scroll box to the left of the ANALYZE button, make sure it is on "Chi-Square". Below the ANALYZE button is a circular radio button (a small hand on the right side of the screen points to it). Click on it then click on the ANALYZE button.



3. The CHI SQUARE window that appears displays the observed genotype frequencies for your class and the genotype frequencies that would be expected if your class were in Hardy-Weinberg equilibrium. Do these look similar? Close the CHI SQUARE window when you are finished.



By following the above steps, you have directed Allele Server to use a test called Chi-square, a statistical test used for comparing observed frequencies with expected frequencies. The Allele Server analysis gives you a Chi-square value and a p-value. The larger the chi-square value, the greater is the difference between the observed and the expected values. When using the Chi-square analysis, we test the null hypothesis that there is no difference between samples (observed and expected) and we assume that if there is any difference, then it arose simply by chance and is not real. For this study, our null hypothesis is that your class is in Hardy-Weinberg equilibrium. Whether or not we can accept the null hypothesis is given by the p-value. If the calculated p-value is less than 0.05, the null hypothesis is disproved; the population is not in Hardy-Weinberg equilibrium. If the p-value is greater than 0.05, the population may be in Hardy-Weinberg equilibrium; we cannot prove that it is not in Hardy-Weinberg equilibrium.

As an example, let's say that Chi-square analysis of your data gives a p-value of 0.17. This means that there is a 17% probability that the difference between the observed and the expected values is due to chance. It also means that there is an 83% (100% - 17% = 83%) probability that the difference is not due to chance; the difference is real.

What is the Chi-square value for your class? _____

What is the p-value for your class data? _____

Name_____

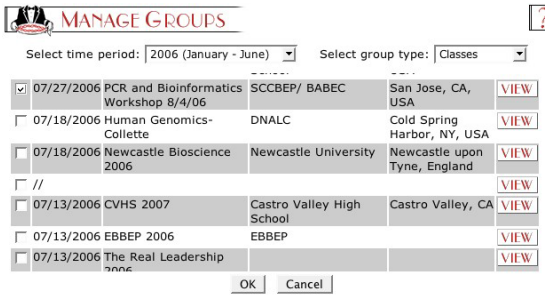
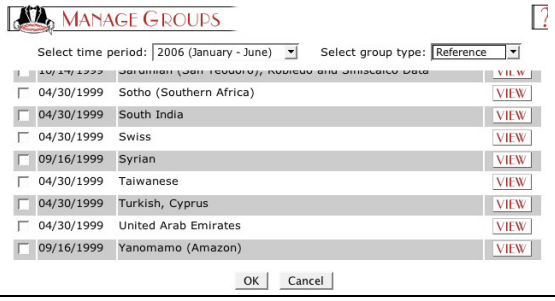
Review Questions: Hardy-Weinberg Equilibrium Analysis

1. What is the probability that the difference between the observed and the expected genotype frequencies calculated for your class is due to chance?

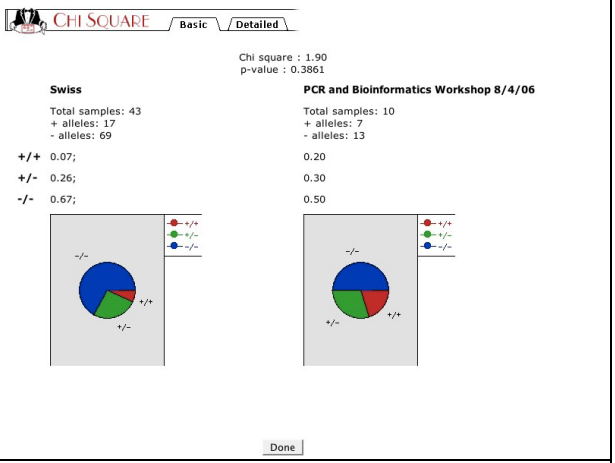
2. Based on the Chi-square and p-values, do you believe your class is in Hardy-Weinberg equilibrium? Why or why not?

Part 4: Using Chi-Square to Compare Two Population Groups

In this part of the exercise, you will use chi-square analysis to determine whether or not there is any difference in the genotype distribution between your class and another population group. If the p-value that is calculated for this comparison is less than 0.05, then the difference between your class and the other population group is probably real. If the p-value is greater than 0.05, then there is probably no difference between your class and the other population group.

Chi Square Analysis of Two Populations	
1. In the Allele Server workspace, click the MANAGE GROUPS button. This will take you to the MANAGE GROUPS window. Select "Classes" from the pop-up menu, select your class and then click the OK button.	
2. You should now have your class data on your Allele Server workspace.	
3. Return to the MANAGE GROUPS window and select "Reference" from the pop-up menu. This will bring up a list of different populations in the world for which the *Alu* PV92 insertion has been determined. Click the box to the left of the population group that you believe should most resemble your class and click the OK button at the bottom of the screen. This will import the data for that group onto the Allele Server workspace.	
4. Click the boxes to the left of both population groups in your workspace. Make sure the popup menu immediately to the right of the COMPARE button displays "Chi-Square". Click the COMPARE button. This will bring up the CHI SQUARE window.	

5. The CHI SQUARE window will display the Chi-square and p-value for these two population groups. **The population group you are comparing your class to is _____. What is the p-value for the Chi-square test? _____ Based on the p-value, are the genotype frequencies of your class and the other population most probably identical or significantly different? _____**



6. When you are finished, close the CHI SQUARE window. Follow the above steps again to identify a human population that your class data most resembles. Choose at least 5 populations to compare with your class. Record your p-values below. **Population** **p-value**

The population group your class most resembles is _____. Chi-square analysis gives a p-value of _____ when these two populations are compared. You may want to compare your class with another class in the database. Other classes can be found in the "Classes" section of the MANAGE GROUPS window.

Name_____

Review Question: Comparing Populations

Two populations on the small island of Sardinia in the Mediterranean Sea have significantly different genotype frequencies for the *Alu* PV92 insert. Provide a possible explanation for this observation.

Applied Biosystems/BABEC Educational PCR Kits

For research use only. Not for use in diagnostic procedures.

Notice to Purchaser: Limited Label License
A license under U.S. Patents 4,683,202, 4,683,195, and 4,965,188 or their foreign counterparts, owned by Roche Molecular Systems, Inc. and F. Hoffmann-La Roche Ltd (Roche), for use in research and development, has an up-front fee component and a running-royalty component. The purchase price of the Lambda PCR, *Alu* PV92 PCR, PCR Optimization, D1S80 PCR, and Mitochondrial PCR Kits includes limited, non-transferable rights under the running-royalty component to use only this amount of the product to practice the Polymerase Chain Reaction (PCR) and related processes described in said patents solely for the research and development activities of the purchaser when this product is used in conjunction with a thermal cycler whose use is covered by the up-front fee component. Rights to the up-front fee component must be obtained by the end user in order to have a complete license. These rights under the up-front fee component may be purchased from Applied Biosystems or obtained by purchasing an authorized thermal cycler. No right to perform or offer commercial services of any kind using PCR, including without limitation reporting the results of purchaser's activities for a fee or other commercial consideration, is hereby granted by implication or estoppel. Further information on purchasing licenses to practice the PCR process may be obtained by contacting the Director of Licensing at Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404 or at Roche Molecular Systems, Inc., 1145 Atlantic Avenue, Alameda, California 94501.

Applied Biosystems' vast distribution and service network, composed of highly trained support and applications personnel, reaches into 150 countries on six continents. For international office locations, please call the division headquarters or refer to our web site at www. Appliedbiosystems.com.