INTRODUCTION

Although the DNA from different individuals is more alike than different, there are many regions of the human chromosomes that exhibit a great deal of diversity. Such variable sequences are termed "polymorphic" (meaning many forms) and provide the basis for genetic disease diagnosis, forensic identification, and paternity testing. In this experiment, polymerase chain reaction (PCR) is used to amplify a DNA sequence from chromosome 16 to look for the insertion of a short nucleotide sequence called *Alu* within a noncoding region of the chromosome.

The *Alu* family of repeated DNA sequences are found throughout primate genomes. *Alu* elements are approximately 300-bp in length and derive their name from the presence of two recognition sites for the endonuclease *Alu*I within the *Alu* element sequence. Over the past 65 million years, the *Alu* sequence has amplified to about 1,000,000 copies, comprising an estimated 10% of the human genome. Many of these *Alu* insertions are only found in humans. Some of these are so recent (1-2 million years ago) that they are not fixed in human populations.

In 1994, Batzer et al. (Proc. Natl. Acad. Sci. U.S.A., vol. 91, pp.12288-12292) described a human-specific insertion of an *Alu* element located at a region (or locus) on chromosome 16 called PV92. This locus is dimorphic; meaning that the *Alu* insertion is present in some individuals but not in others. The presence of the *Alu* sequence does not seem to affect the expression of any gene. As Batzer et al. did, we will use polymerase chain reaction (PCR) to screen for the presence of the PV92 *Alu* insertion. Oligonucleotide primers, flanking the insertion site, were selected to amplify a 715-bp fragment when the *Alu* insertion is present and a 415-bp fragment when it is absent. Each of the three possible genotypes—homozygote for the presence of the *Alu* insertion (715-bp fragment only), homozygote for the absence of the *Alu* insertion (415-bp fragment only), and heterozygote (715-bp and 415-bp fragments)—can be distinguished following electrophoresis in agarose gels.

The source of template DNA for amplification is a sample of several thousand cells obtained by saline mouthwash (bloodless and noninvasive). The cells are collected by centrifugation and resuspended in a solution containing the resin "Chelex," which binds metal ions that inhibit the PCR. The cells are lysed by boiling and centrifuged to remove cell debris. A sample of the supernatant containing chromosomal DNA is mixed with Taq DNA polymerase, oligonucleotide primers, the four deoxynucleotides, and the proper buffer for the enzyme. Temperature cycling is used to denature the target DNA, anneal the primers, and extend a complementary DNA strand. The size of the amplification product(s) depends on the presence or absence of the *Alu* insertion at the PV92 locus on each copy of chromosome 16.

In order to compare the genotypes from a number of individuals, aliquots of the amplified sample and a DNA molecular ladder are loaded into the wells of an agarose gel. Following electrophoresis and staining, amplification products appear as distinct bands in the gel—the distance moved from the well is inversely proportional to the size of the DNA fragments. One or two bands are visible in each lane, indicating the genotype for an individual.

PROCEDURE

I. DNA Isolation by Saline Mouthwash

1. Pour 10 ml of the saline solution (0.9% NaCl) into mouth and vigorously swish for 30 seconds.
2. Expel saline solution into a paper cup.
3. Swirl to mix cells in the cup and use a micropipette to transfer 1 ml (1000 µl) of the liquid to 1.5-ml tube.
4. Place your sample tube, together with other student samples, in a balanced configuration in a microcentrifuge, and spin for 1 minute.
5. Carefully pour off supernatant into paper cup or sink. Be careful not to disturb the cell pellet at the bottom of the test tube. A small amount of saline will remain in the tube.
6. Resuspend cells in remaining saline by pipetting in and out. (If needed, 30 µl of saline solution may be added to facilitate resuspension.)
7. Use a micropipette to withdraw 30 µl of cell suspension, and add to tube containing 100 µl of Chelex. Shake well to mix.
8. Boil cell sample for 10 minutes. Use boiling water bath, heat block, or program thermal cycler for 10 minutes at 99° C. Then, cool tube briefly on ice (optional).
9. After boiling, shake tube. Place in a balanced configuration in a microcentrifuge, and spin for 1 minute.
10. Use a micropipette to transfer 30 µl of supernatant (containing the DNA) to clean 1.5-ml tube. Avoid cell debris and Chelex beads. This sample will be used for setting up one or more PCR reactions.

II. DNA Amplification by PCR

1. Use a micropipette with a fresh tip to add 22.5 µl of the appropriate primer/loading buffer mix to a PCR tube containing a Ready-To-Go PCR Bead. Tap tube with finger to dissolve bead.
2. Use a micropipette with a fresh tip to add 2.5 µl of human DNA (from Part I) to reaction tube, and tap to mix. Pool reagents by pulsing in a microcentrifuge or by sharply tapping tube bottom on lab bench.
3. Label the cap of your tube with a number, as assigned by your teacher. In this way, your results will be anonymous.
4. If the thermal cycler requires, add one drop of mineral oil on top of reactants in the PCR tube. Be careful not to touch the dropper tip to the tube or reactants, or subsequent reactions will be contaminated with DNA from your preparation. Note: Thermal cyclers with heated lids do not require use of mineral oil.

5. Store all samples on ice until ready to amplify according to the following profile. Program thermal cycler for 30 cycles according to the following cycle profiles. Each program may be linked to a 4° C cycle to hold samples after completing the cycle profile, but amplified DNAs also hold well at room temperature.

Denaturing time and temperature	30 sec - 94° C
Annealing time and temperature	30 sec - 68° C
Extending time and temperature	30 sec - 72° C

III. DNA Analysis by Gel Electrophoresis

1. Use a micropipette with a fresh tip to add entire PCR sample/loading dye mixture into your assigned well of a 2% agarose gel. Expel any air from the tip before loading, and be careful not to push the tip of the pipet through the bottom of the sample well.

2. Load 5 µl of the pBR322-BstNI size markers into one lane of gel.

3. Electrophorese at 140 volts for 20-30 minutes. Adequate separation will have occurred when the cresol red dye front has moved at least 50 mm from the wells.

4. Gels will contain DNA stain and be ready to view with a UV transilluminator immediately. DNA fragments of identical size in the same lane will appear as a discrete band.

RESULTS AND DISCUSSION

(follow-up activities and questions here: http://www.geneticorigins.org/pv92/aluframeset.htm)

I.-3. What can you say about any person in the class who has at least one "+" allele?

II.-1. What is the p-value for your Chi-Square? Is it less than .05? If so, can you suggest any factors that might account for why your observed population is not in Hardy-Weinberg equilibrium?

III.-1. Based on the results you recorded, how useful is the PV-92 *Alu* polymorphism in distinguishing populations from each other? Why do you think that the PV-92 allele frequencies differ significantly between some populations, while not between others? Do you think you could use PV-92 data to answer the questions of where humans originated and the paths by which they spread throughout the world? What other population data sets or types of information might you need to accurately answer this question?

III.-2. Use the Heterozygosity Feature of the Analyze Function to determine the + allele frequency in a number of populations representing differing parts of the world. Select 'Heterozygosity' from the pulldown menu on the RIGHT. Click the round checkbox underneath to select a group for analysis. (You can only analyze one group at a time). Print the world map (http://www.bioservers.org/map.html) and plot the + allele frequencies on it.

a. Do you notice any pattern in the allele frequencies?
b. Suggest a hypothesis about the origin and dispersal of the *Alu* allele that accounts for your observation.
c. Which of these mechanisms is consistent with the statistical evidence that PV92 first inserted about 200,000 years ago?

III.-3. Considering your results, do you think this protocol could be used forensically to link a suspect with a crime?

III.-4. *Alu* can be considered parasites of retroviruses, which produce the enzyme reverse transcriptase needed for transposition. Why is this so? Propose a transposition mechanism requiring reverse transcriptase.

III.-5. An *Alu* insertion has only two states: + and -. How does this relate to information stored in digital form by a computer? How much digital information is provided by an *Alu* genotype?

IV.-1. Examine the graph that shows the frequency of the "+" allele versus generation number. In general, what happens to the "+" allele (i.e., the *Alu* insertion) over time? Are there any runs in which the *Alu* insertion begins to spread throughout the population during the first 100 generations? Is this a common or rare occurrence? For the runs where the *Alu* allele is not lost from the population, what happens to the allele frequency when the population size "explodes" (generation 101-200 in Population 2)? Conversely, what happens to the *Alu* allele when the population is kept small (generation 101-200 in Population 3)? In addition to population size, can you think of any other factors that would help an *Alu* insertion event spread? (Hint: What would happen if inheriting an *Alu* insertion conferred a selective advantage to the individual carrying the insertion?)