Biology 3A Laboratory  
Analysis and Interpretation of Results  

Analysis  
Compare your sample lanes with the control lanes using the DNA size marker as a reference. Mark the location and size of your fragment or fragments. By comparing your DNA migration pattern to the controls, determine whether you are homozygous +/+ , homozygous −/− , or heterozygous +/− . If your sample lane is blank, discuss the possible reasons with your classmates and teacher.  

Remember that the interpretation of this gel allows you to determine your genetic makeup only at the locus (chromosomal location), being studied. There are three possible genotypes for the Alu insert at the location you have amplified. For a class, determine the number of individuals of each genotype: homozygous +/+ , homozygous −/− , and heterozygous +/− . Tally the class results in the table on page 70.  

A major factor affecting the reliability of DNA fingerprinting evidence in forensics is population genetics and genetic statistics. In humans, there are hundreds of loci, or DNA segments, like Alu, that can be selected and used for fingerprinting analysis. Depending on demographic factors such as ethnicity and geographic isolation, some populations show much less variation in particular DNA segments than others. A lower degree of variation will increase the odds of more than one individual having the same sequence. If 33% (1 out of 3 individuals) of a given population has the same fingerprinting pattern for a certain DNA segment, then little information will be obtained from using that segment alone to identify an individual. In such a case, a positive result would only identify a person with 33% accuracy.  

In analyzing how incriminating the DNA evidence is, one needs to ask the question: Statistically, how many people in a population have the same DNA pattern as that taken from a crime scene: 1 in 1,000,000? 1 in 10,000? 1 in 10?  

For a DNA fingerprint to identify a suspect in a criminal case or a father in a paternity suit, accurate identification required not a 1 out of 3 (1/3) chance of a match in a population, but closer to a 1 in 10 million (1/10) chance of a match. The frequency of a particular DNA pattern turning up in a population drastically decreases when multiple DNA segments are selected and amplified, rather than just one segment. For DNA fingerprinting to be admissible as evidence in court, it must analyze 30 to 40 different DNA segments on multiple chromosomes from the same person.  

The Alu insert you have fingerprinted in this exercise has been used to study the migration patterns of human populations over time.  

The data from these studies have been published, and your class samples can be compared to the data collected from much larger populations.  

Analysis and Interpretation of Results  
Remember that this Alu sequence is inserted into a noncoding region of the PV92 locus on chromosome 16 and is not related to a particular disease, nor does it code for any protein sequence. It is simply a sequence that can be used to study human genotypic frequencies. Because Alu repeats appear in the general population at random, the Alu insert in chromosome 16 is very useful for the study of gene frequencies in localized human populations. Theoretically, in some small, geographically isolated populations, all individuals may be homozygous +/+ . In others, the individuals may all be homozygous −/− . In a “melting-pot” population, the three genotypes (+/+ , +/− , −/− ) may exist in equilibrium. The frequencies of genotypes and alleles are basic characteristics that population geneticists use to describe and analyze populations. The results you obtain in this exercise
provide a real-life opportunity to calculate genotypic and allelic frequencies of the Alu insert in your class and to use the Hardy-Weinberg equation.

The results of the PCR reactions reveal your and your classmates’ genotypes: +/+, +/–, and −/−. Knowing your genotypes, you can count up the alleles of your class “population” and determine their frequencies. You can then compare the allelic and genotypic frequencies of your class population to published reports of larger population sizes.

**Questions:**

1. What is your genotype for the Alu insert in your PV92 region?

2. What are the genotypic frequencies of +/+, +/–, and −/− in your class population? Fill in the table below with your class data.

**Table 1. Observed Genotypic Frequencies for the Class Category Number Frequency (# of Genotypes/Total)**

<table>
<thead>
<tr>
<th>Category</th>
<th>AM Lab #</th>
<th>PM Lab #</th>
<th>TH Lab #</th>
<th>Frequency for your lab only</th>
<th>Frequency for all sections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homozygous (+/+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterozygous (+/–)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homozygous (−/−)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Allelic frequencies can be calculated from the numbers and frequencies of the genotypes in the population. Population geneticists use the terms p and q to represent the frequencies of the (+) and (−) alleles, respectively. Allele frequencies can be calculated from either the numbers or the frequencies of the genotypes (since they are related to each other).

\[
p = \text{frequency of (+) allele} = \frac{\text{number of (+) alleles}}{\text{total number of alleles (both + and −)}}
\]

\[
= \frac{2(\text{# of } +/+ \text{ students}) + 1(\text{# of } +/– \text{ students})}{\text{total number of alleles (both + and −)}}
\]

\[
= \text{frequency of } (+/+ \text{ students) } + \frac{1}{2} \text{ (frequency of } (+/– \text{ students)}
\]

\[
q = \text{frequency of (−) allele} = \frac{\text{number of (−) alleles}}{\text{total number of alleles (both + and −)}}
\]

\[
= \frac{2(\text{# of } −/− \text{ students}) + 1(\text{# of } +/– \text{ students})}{\text{total number of alleles (both + and −)}}
\]

\[
= \text{frequency of } (−/− \text{ students)} + \frac{1}{2} \text{ (frequency of } (+/– \text{ students)}
\]
3. Calculate the frequency of each allele in your class sample and for all sections combined? Remember, a class of 32 students (N) will have a total of 64 (2N) instances of each locus.

**Calculated Allelic Frequencies for your Class Category & all sections**

<table>
<thead>
<tr>
<th>Frequency for your lab section</th>
<th>Frequency for all sections</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+) alleles</td>
<td></td>
</tr>
<tr>
<td>p =</td>
<td></td>
</tr>
<tr>
<td>(-) alleles</td>
<td></td>
</tr>
<tr>
<td>q =</td>
<td></td>
</tr>
</tbody>
</table>

Total alleles = 1.00

4. The following table presents data from a USA-wide random population study.

**Table 3. Genotypic Frequencies for Alu in a USA Sample**

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homozygous (+/+)</td>
<td>2,422</td>
<td>0.24</td>
</tr>
<tr>
<td>Heterozygous (+/–)</td>
<td>5,528</td>
<td>0.55</td>
</tr>
<tr>
<td>Homozygous (–/–)</td>
<td>2,050</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Total = 10,000 = 1.00

Now, using the data above, calculate the allelic frequencies for the USA data as you did for your class population.

**Frequency for the USA data set**

| (+) alleles | p = |
| (-) alleles | q = |

Total alleles = 1.00

5. How do your actual class data for genotypic and allelic frequencies compare with those of the random sampling of the USA population? Would you expect them to match? What reasons can you think of to explain the differences or similarities?

The **Hardy-Weinberg equation**, \( p^2 + 2pq + q^2 = 1 \), is one of the foundations of population genetics. It is the algebraic expansion of \( (p + q)^2 = 1 \), where \( p + q = 1 \). The equation describes the frequencies of genotypes in a population that is at "genetic equilibrium", meaning that the frequencies are stable from generation to generation. The Hardy-Weinberg theory states that, for a population to achieve this equilibrium, the population must be quite large, the members must mate randomly and produce offspring with equal success, and there must be no migration of individuals into or out of the
population, or an excessive mutation converting one allele to another. Given these conditions, and the allelic frequencies p and q, the Hardy-Weinberg equation says that

\[ p^2 = \text{the expected frequency of the (+/+) genotype in the population} \]
\[ 2pq = \text{the expected frequency of the (+/–) genotype in the population} \]
\[ q^2 = \text{the expected frequency of the (–/–) genotype in the population} \]

It is important to understand that \( p^2 \), \( 2pq \), and \( q^2 \) are expected, theoretical genotype frequencies of a population under Hardy-Weinberg equilibrium conditions, and they may not be realized in real-life population samples if one of the conditions is not met. These theoretical frequencies are calculated using the observed values for \( p \) and \( q \); they may or may not be the same as the observed genotypic frequencies such as those shown in Table 1. If the observed and expected genotypic frequencies are the same, this indicates that the population is in Hardy-Weinberg genetic equilibrium.

6. Using the values for \( p \) and \( q \) that you calculated in Table 2 for your class population, calculate \( p^2 \), \( 2pq \), and \( q^2 \). Do they come out to be the same as the genotype frequencies that you found in Table 1? If they do, your class resembles a Hardy-Weinberg genetic equilibrium. If your observed (actual) genotype frequencies are not the same as the expected values, what might be some of the reason(s) for the difference?
Run the chi squared analysis compared with the 0.25 (+/+), 0.50 (+/-) and 0.25 (-/-) genotypic ratio.

7. Using the values for \( p \) and \( q \) that you calculated in Table 4 for the USA population sample, calculate \( p^2 \), \( 2pq \), and \( q^2 \). Do they come out to be the same as the genotype frequencies that you found in Table 3? Does this USA-wide sample suggest that the population of the USA is in Hardy-Weinberg equilibrium?
Run the chi squared analysis comparing this data set to (0.25 +/+, 0.50 +/- and 0.25 -/-) and your data set attach a figure representing the analysis of all three data sets.