Biotechnology Explorer™

Chromosome 16: PV92 PCR Informatics Kit

Catalog #166-2100EDU

explorer.bio-rad.com

Note: Kit contains temperature-sensitive reagents. Open immediately upon arrival and store components at –20°C or at 4°C as indicated.

Please visit explorer.bio-rad.com to access our selection of language translations for Biotechnology Explorer kit curricula.

BIO-RAD

Background for Teachers

Introduction to PCR

In 1983, Kary Mullis[2] at Cetus Corporation developed the molecular biology technique that has since revolutionized genetic research, earning him the Nobel Prize in 1993. This technique, termed the **polymerase chain reaction (PCR)**, transformed molecular biology into a multidisciplinary research tool. Many molecular biology techniques used before PCR were labor intensive, time consuming and required a high level of technical expertise. Additionally, working with only trace amounts of DNA made it difficult for researchers in other biological fields (pathology, botany, zoology, pharmacy, etc.) to incorporate molecular biology into their research schemes.

PCR had an impact on four main areas of biotechnology: gene mapping, cloning, DNA sequencing, and gene detection. PCR is now used as a medical diagnostic tool to detect specific mutations that may cause genetic disease,[3] in criminal investigations and courts of law to identify suspects on a molecular level,[4] and in the sequencing of the human genome.[5] Prior to PCR the use of molecular biology techniques for therapeutic, forensic, pharmaceutical, or medical diagnostic purposes was not practical or cost-effective. The development of PCR technology changed these aspects of molecular biology from a difficult science to one of the most accessible and widely used tools in genetic and medical research.

PCR and Biotechnology — What Is It and Why Did It Revolutionize an Entire Research Community?

PCR produces exponentially large amounts of a specific piece of DNA from trace amounts of starting material (template). The template can be any form of double-stranded DNA, such as genomic DNA. A researcher can take trace amounts of DNA from a drop of blood, a single hair follicle, or a cheek cell and use PCR to generate millions of copies of a desired DNA fragment. In theory, only one single template strand is needed to generate millions of new DNA molecules. Prior to PCR, it would have been impossible to do forensic or genetic studies with this small amount of DNA. The ability to amplify the precise sequence of DNA that a researcher wishes to study or manipulate is the true power of PCR.

PCR amplification requires the presence of at least one DNA template strand. In this kit, human genomic DNA isolated from students' own cells will be the source of the template strands. One of the main reasons PCR is such a powerful tool is its simplicity and specificity. All that is required are inexpensive reaction buffers, four DNA subunits (deoxynucleotide triphosphates of adenine, guanine, thymine, and cytosine), a DNA polymerase, two DNA primers, and minute quantities of the template strand that one wants to amplify. Specificity comes from the ability to target and amplify one specific segment of DNA out of a complete genome.

PCR Makes Use of Two Basic Processes in Molecular Genetics

1. **Complementary DNA strand hybridization**
2. **DNA strand synthesis via DNA polymerase**

In the case of PCR, complementary strand hybridization takes place when two different **oligonucleotide primers** anneal to each of their respective complementary base pair sequences on the template. The two primers are designed and synthesized in the laboratory with a specific sequence of nucleotides such that they can anneal at the opposite ends and on the opposite strands of the stretch of double-stranded DNA (template strand) to be amplified.

Before a region of DNA can be amplified, one must identify and determine the sequence of a piece of DNA upstream and downstream of the region of interest. These areas are then used to determine the sequence of oligonucleotide primers that will be synthesized and used as starting points for DNA replication. Primers are complimentary to the up- and downstream regions of the sequence to be amplified, so they stick, or anneal, to those regions. Primers are needed because DNA polymerases can only add nucleotides to the end of a preexisting chain.

The DNA polymerase used in PCR must be a thermally stable polymerase because the polymerase chain reaction cycles between temperatures of 60°C and 94°C. The thermostable DNA polymerase (*Taq*) used in PCR was isolated from a thermophilic bacterium, *Thermus aquaticus*, which lives in high-temperature steam vents such as those found in Yellowstone National Park.[6]

Two new template strands are created from the original double-stranded template on each complete cycle of the strand synthesis reaction. This causes exponential growth of the number of template molecules, i.e., the number of DNA strands doubles at each cycle. Therefore, after 30 cycles there will be 2^{30}, or over 1 billion, times more copies than at the beginning. Once the template has been sufficiently amplified, it can be visualized. This allows researchers to determine the presence or absence of the desired PCR products and determine the similarities and differences between the DNA of individuals. Depending on the DNA sequence analyzed, differences among individuals can be as great as hundreds of base pairs or as small as a single base pair or single point mutation.

Genes and DNA

It is estimated that the 23 pairs of chromosomes (46 total chromosomes) of the human genome contain a total of 30,000–50,000 genes. Each gene holds the code for a particular protein. Interestingly, these 30,000–50,000 genes comprise only about 5% of chromosomal DNA. The other 95% is noncoding DNA. This noncoding DNA is found not only between, but within genes, splitting them into segments. In eukaryotes, these sequences within genes (called introns) are transcribed into RNA but in the end do not make a protein called **introns**. The sequences that do code for proteins are called **exons**. Both introns and exons are initially transcribed, then introns are spliced out of the RNA to create messenger RNA (mRNA).

In eukaryotes, genomic DNA is transcribed into RNA molecules containing both introns and exons for a particular gene. While the RNA is still in the nucleus (before being transported out of the nucleus), the introns (in = stay within the nucleus) must be removed from the RNA while the exons (ex = exit the nucleus) are spliced together to form the complete coding sequence for the protein (Figure 1). This process is called **RNA splicing**. Some genes may contain a few introns, others may contain dozens.

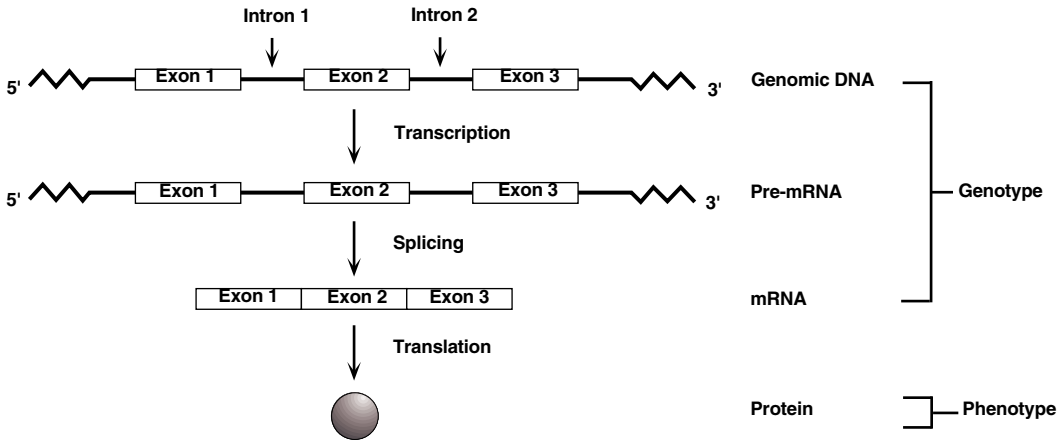

Fig. 1. Splicing of introns from genes.

As we have discussed, functional segments of genes (exons) code for proteins — molecules that carry out most cellular functions. Exon sequences are therefore similar among individuals. Introns, on the other hand, often vary in size and number among individuals. Intron sequences are thought to be the result of the differential accumulation of mutations throughout evolution that are silently passed to descendants through the hereditary code. It is this difference in intron sequences that allows us to determine human genetic diversity. The identification of these distinctive characteristics in the DNA represent the molecular basis for human identification and population genetics.

Throughout evolution, intron sequences have been the target of random insertions by short repetitive interspersed elements, also known as SINEs.[7] SINEs have become randomly inserted within our introns over millions of years. One such repetitive element is called the Alu sequence[7] (Figure 2). This is a DNA sequence about 300 base pairs long that is repeated, one copy at a time, almost 500,000 times within the human genome.[8] The origin and function of such randomly repeated sequences is not yet known. The Alu name comes from the *Alu* I **restriction enzyme** recognition site that is found in this sequence.

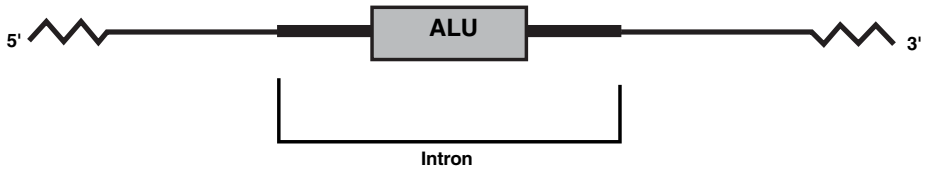

Fig. 2. Location of an Alu insertion within an intron.

Some of these Alu elements have characteristics that make them very useful to geneticists. If present within introns of genes associated with particular pathologies, they can thereby be associated with that disease. When present within the introns of genes, Alu repeats can also be used to estimate relatedness among individuals. In this activity, analysis of Alu repeats is used to estimate the frequency of an insert in a population and is a simple measure of molecular genetic variation — **with no reference to disease or relatedness among individuals**.

This kit provides a simple PCR-based screen for a single Alu sequence within the PV92 locus on chromosome 16. This particular Alu intron is dimorphic. That is, the element is present in some individuals but not others (Figure 3). Some people have the insert in the PV92 locus of one of their chromosome 16, others may have the insert in both homologous chromosomes (two alleles), and some do not have the insert in either chromosome. The presence or absence of this insert can be detected using the polymerase chain reaction followed by agarose gel electrophoresis.

In this activity, students will isolate their own genomic DNA from their cells. They will use primers that flank both the entire Alu insertion (300 base pairs in length) and 641 base pairs of the PV92 locus to amplify a 941 base pair fragment (if the Alu element is present) or a 641 base pair fragment (if the Alu element is absent). Agarose gel electrophoresis of the PCR products is sufficient to distinguish among homozygotes (+/+) for the presence of the Alu repeat (941 base pair product only), homozygotes (–/–) for the absence of the Alu repeat (641 base pair product only), and heterozygotes (+/–) having both the 641 and the 941 base pair products.

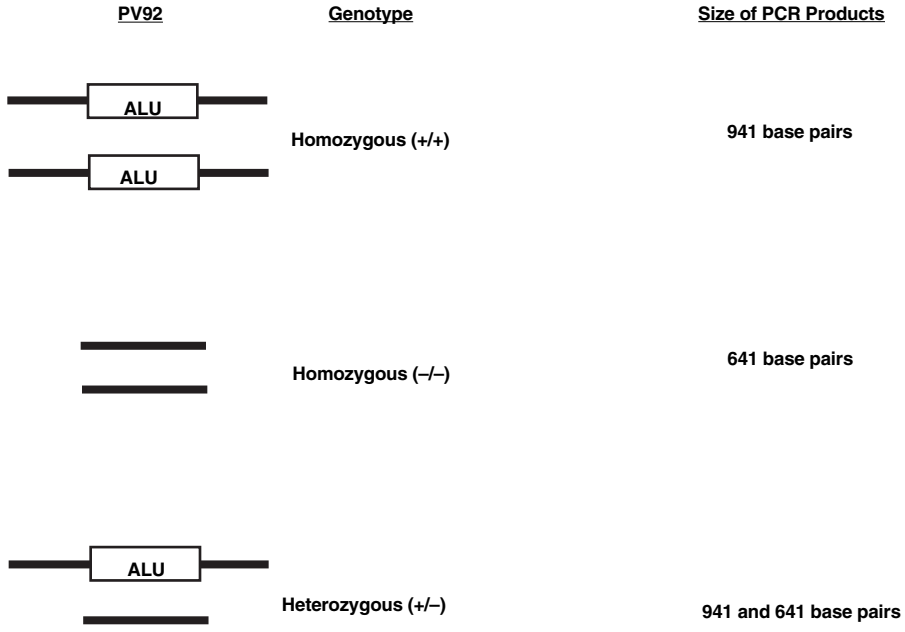

Fig. 3. The presence or absence of the Alu insert within the PV92 locus on chromosome 16.

Important Notes for the Instructor

Please note that since PV92 alleles are inherited from parents and can potentially reveal information about family relationships, we caution against generating genotypic data from multiple members of a family. If confidentiality is a concern, we suggest that the instructor mix up student samples to ensure anonymity. Student samples can be randomized at any point after cells are harvested.

Two protocols are provided for genomic DNA preparation. One involves the collection of oral epithelial cells using a saline mouthwash. The other isolates genomic DNA from hair follicles. Both methods are minimally invasive, and yield robust PCR products. Instructors may choose either protocol based on personal or student preference, or local restrictions.

PCR Step by Step

PCR involves a repetitive series of cycles, each of which consists of template denaturation, primer annealing, and extension of the annealed primer by *Taq* DNA polymerase. Before beginning DNA amplification, genomic DNA is prepared from students' cells.

Following sample preparation, the template DNA, oligonucleotide primers, thermostable DNA polymerase (*Taq*), the four deoxynucleotides (A, T, G, C), and reaction buffer are mixed in a single micro test tube. The tube is placed into the MyCycler™ thermal cycler. These thermal cyclers contain an aluminum block that holds the samples and can be rapidly heated and cooled across extreme temperature differences. The rapid heating and cooling of this thermal block is called **temperature cycling** or **thermal cycling**.

The first step of the PCR temperature cycling procedure involves heating the sample to 94°C. At this high temperature, the template strands separate (denature). This is called the **denaturation step**.

The thermal cycler then rapidly cools to 60°C to allow the primers to anneal to the separated template strands. This is called the **annealing step**. The two original template strands may reanneal to each other or compete with the primers for the primers complementary binding sites. However, the primers are added in excess such that the primers actually out-compete the original DNA strands for the primers' complementary binding sites.

Lastly, the thermal cycler heats the sample to 72°C for *Taq* DNA polymerase to extend the primers and make complete copies of each template DNA strand. This is called the **extension step**. *Taq* polymerase works most efficiently at this temperature. Two new copies of each complementary strand are created. There are now two sets of double-stranded DNA (dsDNA). These two sets of dsDNA can now be used for another cycle and subsequent strand synthesis.

At this stage, a complete temperature cycle (thermal cycle) has been completed (Figure 4).

Temperature cycle = denaturation step + annealing step + extension step

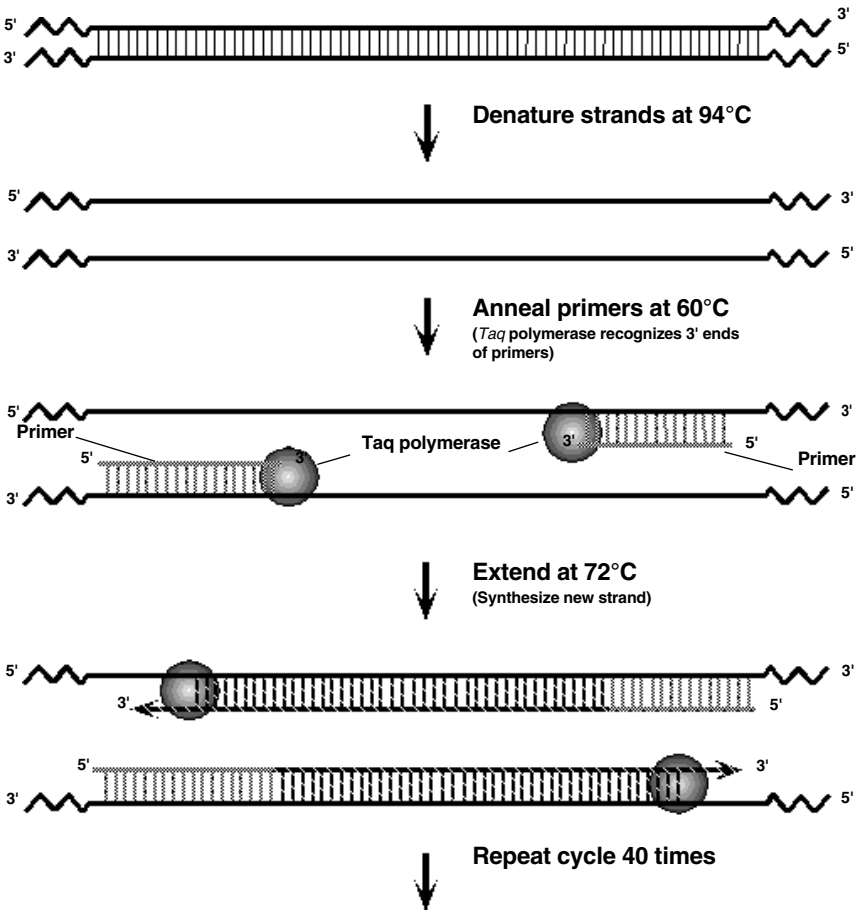

Fig. 4. A complete cycle of PCR.

Usually, thermal cycling continues for about 40 cycles. After each thermal cycle, the number of template strands doubles, resulting in an exponential increase in the number of template DNA strands. After 40 cycles there will be 1.1×10^{12} more copies of the original number of template DNA molecules.

PCR generates DNA of a precise length and sequence. On the first cycle, the two primers anneal to the original genomic template DNA strands at opposite ends and on opposite strands. After the first complete temperature cycle, two new strands are generated that are shorter than the original template strands but still longer than the length of the DNA that the researcher wants to amplify. It isn't until the third thermal cycle that fragments of the precise length are generated (Figure 5).

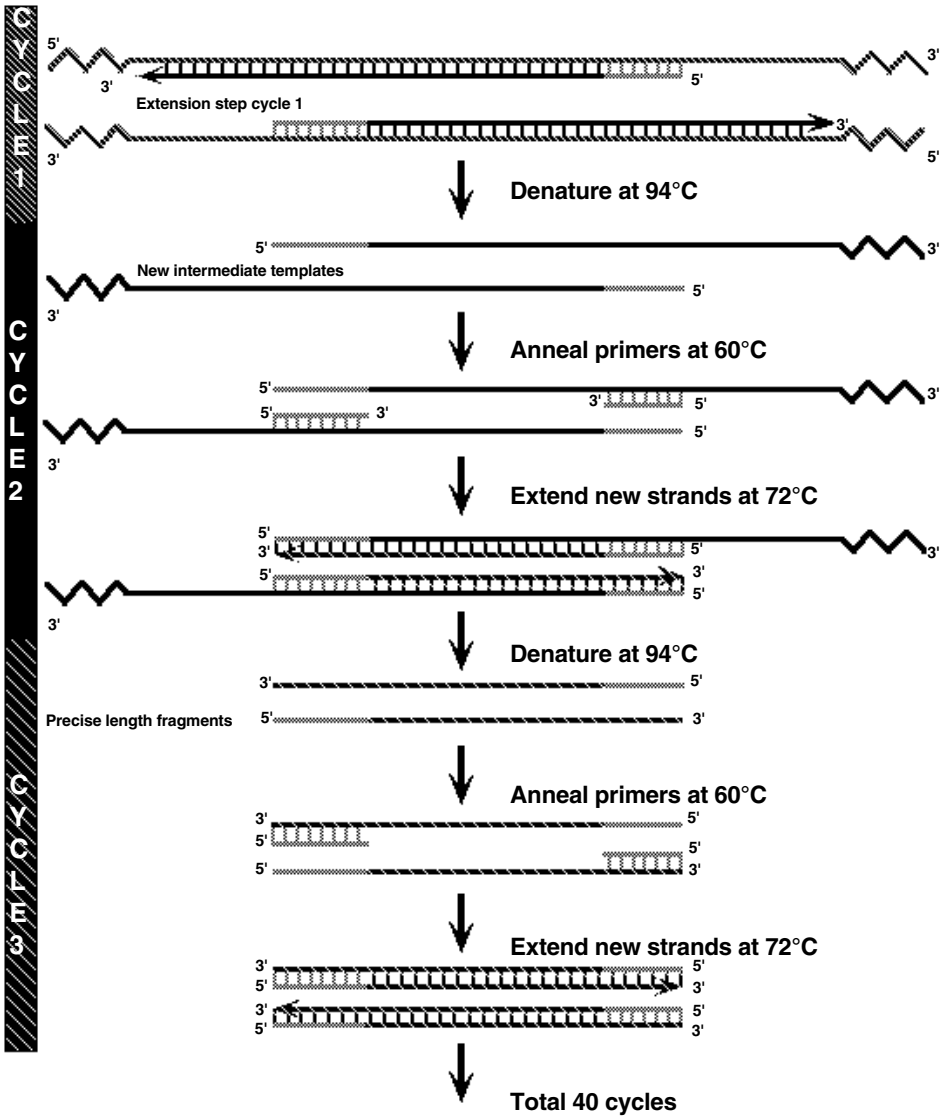

Fig. 5. Generation of precise-length fragments.

It is the template strands of the precise length that are amplified exponentially (X^n, where X = the number of original template strands and n = the number of cycles). There is always one set of original long-template DNA molecules which is never fully duplicated. After each thermal cycle, two intermediate-length strands are produced, but because they can only be generated from the original template strands, the intermediate strands are not exponentially amplified. It is the precise-length strands generated from the intermediate strands that amplify exponentially at each cycle. Therefore, if 20 thermal cycles were conducted from one double-stranded DNA molecule, there would be 1 set of original genomic template DNA strands, 20 sets of intermediate template strands, and 1,048,576 sets of precise-length template strands. After 40 cycles, there would be 1 set of original genomic template DNA strands, 40 sets of intermediate template strands, and 1.1×10^{12} sets of precise-length template strands (Figure 6).

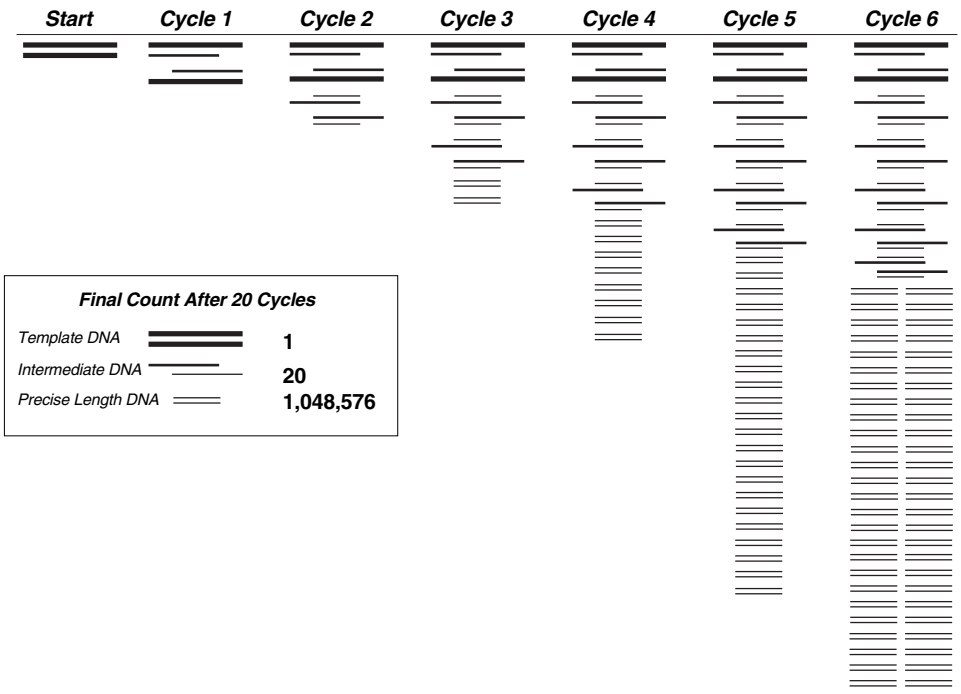

Fig. 6. Schematic of PCR amplification of DNA fragments.

The Relevance of Alu to Hardy-Weinberg Theory

The value of this Alu insert residing in a noncoding region of PV92 is that it does not code for any protein sequence and is not related to a particular disease. Because Alu repeats have inserted randomly in humans, the Alu insert in the PV92 locus can be very useful in studies of allele and genotype frequencies in the human population. In the lab exercise (Lesson 4), the principles of the Hardy-Weinberg theory can be applied to analyze genotypic and allelic frequencies of the Alu insert in populations. By determining the genotypic frequencies of the Alu genotype within your student population, the corresponding allelic frequencies can also be calculated. Additionally, the genotypic frequencies of your class population can be compared to published results of larger population sizes. This activity opens the door to teaching the principles of the Hardy-Weinberg theory. Use the student data generated in Tables 1 and 2 (pages 70–71) to discuss whether human populations are in Hardy-Weinberg equilibrium.

Analysis of Classroom Data Using Bioinformatics

Bioinformatics is a discipline that integrates mathematical, statistical, and computer tools to collect and process biological data. Bioinformatics has become an important tool in recent years for analyzing the extraordinarily large amount of biological information that is being generated by researchers around the world. In Lesson 5, students will perform a bioinformatics exercise to investigate the genotypic frequencies for the Alu polymorphism in their class population and compare them with the genotypic frequencies of other populations.

Following PCR amplification and electrophoresis of their samples, students will analyze their experimental data to determine their genotypes for the Alu insertion within the PV92 locus on chromosome 16. The classroom genotype data can then be entered into Allele Server of Cold Spring Harbor Laboratory's Dolan DNA Learning Center. Allele Server is a Web-based database that contains genotype data from populations around the world as well as other classrooms and teacher training workshops. It also provides a collection of statistical analysis tools to examine the Alu insertion polymorphism at the population level. Students can either analyze their classroom data as an individual population or compare their population with other populations in the database.

Once students enter classroom data into Allele Server, they can perform a Chi-square analysis to compare the Alu genotype frequencies within the class population with those predicted by the Hardy-Weinberg equation. The genotypic frequencies of the class population can also be compared with the genotypic frequencies of another population in the database. Using this database, students will determine if their class data are in agreement with the expected Hardy-Weinberg genotypic frequencies.

Suggested Lesson Flow

There are four student lessons in this PCR curriculum. All lessons are designed to be carried out in consecutive 50 minute periods. Lessons 1 and 2 have convenient stop points and two options. Teachers should choose cheek cell (page 41) or hair follicle (page 45) DNA preparation. Instructors may wish to offer either method as an option for students, or may elect to perform a particular protocol based on local restrictions. The samples can be stored for several days to accommodate weekends or labs that meet every other day.

Student Schedule

Lesson 1	**Cheek Cell DNA Template Preparation**
Activity	Isolate cheek cells Prepare genomic DNA from cheek cells (Stop point)
Lesson 1	**Hair Follicle DNA Template Preparation**
Activity	Isolate hairs Prepare genomic DNA from hair follicles (Stop point)
Lesson 2	**PCR Amplification**
Activity	Set up and perform PCR reactions Pour agarose gels (this may be performed by the instructor during the advance preparation) (Stop point)
Lesson 3	**Gel Electrophoresis of Amplified PCR Samples and Staining of Agarose Gels**
Activity	Load and run gels Stain gels (Note: If you are using the quick staining protocol, record the results and dry the gels)
Lesson 4	**Analysis and Interpretation of Results**
Activity	Record the results and dry the gels (if using the overnight staining protocol) Analyze and discuss results
Lesson 5	**Interpretation of Results: Bioinformatics**
Activity	Enter classroom data into PV92 Allele Server and analyze data