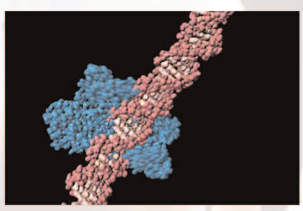
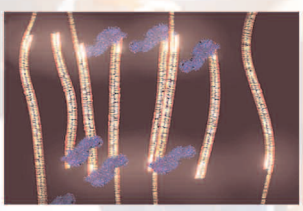
21-1230 21-1230A 21-1231
21-1231A 21-1232 21-1232A

Using an *Alu* Insertion Polymorphism to Study Human Populations

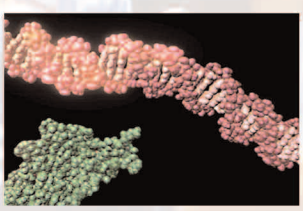
DNA KITS
Learning Center



DOLAN
DNA
LEARNING
CENTER



CAROLINA



Using an *Alu* Insertion Polymorphism to Study Human Populations

IMPORTANT INFORMATION

Storage: Upon receipt of the kit, store proteinase K, PV92B primer/loading dye mix, and DNA marker pBR322/*Bst*NI in a freezer (approximately –20°C). All other materials may be stored at room temperature (approximately 25°C).

Use and Lab Safety: The materials supplied are for use with the method described in this kit only. Use of this kit presumes and requires prior knowledge of basic methods of gel electrophoresis and staining of DNA. Individuals should use this kit only in accordance with prudent laboratory safety precautions and under the supervision of a person familiar with such precautions. Use of this kit by unsupervised or improperly supervised individuals could result in injury.

Limited License: Polymerase chain reaction (PCR) is protected by patents owned by Hoffman-La Roche, Inc. The purchase price of this product includes a limited, non-transferable license under U.S. Patents 4,683,202; 4,683,195; and 4,965,188 or their foreign counterparts, owned by Hoffmann-La Roche Inc. and F. Hoffmann-La Roche Ltd. (Roche), 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, educational, and training activities of the purchaser when this product is used either manually or in conjunction with 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 The Perkin-Elmer Corporation, 850 Lincoln Center Drive, Foster City, California 94404 or at Roche Molecular Systems, Inc., 1145 Atlantic Avenue, Alameda, California 94501.

Printed material: The student instructions, pages 5–24, as well as the *Carolina*BLU™ staining protocol on page 32 may be photocopied as needed for use by your students.

REAGENTS, SUPPLIES, AND EQUIPMENT CHECKLIST

Included in the kit:

DNA extraction and amplification (all kits):

- ☐ 1.5 g Chelex® resin
- ☐ 5 mL proteinase K (100 µg/mL)
- ☐ 700 µL PV92B primer/loading dye mix
- ☐ 25 *Ready-to-Go™ PCR Beads
- ☐ 5 mL mineral oil
- ☐ 130-µL tube pBR322/*Bst*NI markers (0.075 µg/µL)
- ☐ Instructor's manual with reproducible Student Lab Instructions
- ☐ *Alu* CD-ROM

***Electrophoresis kits with ethidium bromide staining (Kits 21-1231and 21-1231A) also include:*

- ☐ 5 g agarose
- ☐ 150 mL 20× TBE
- ☐ 250 mL ethidium bromide, 1 µg/mL
- ☐ 4 latex gloves
- ☐ 6 staining trays

***Electrophoresis kits with Carolina*BLU™ *staining (Kits 21-1232 and 21-1232A) also include*:

- ☐ 5 g agarose
- ☐ 150 mL 20× TBE
- ☐ 7 mL *Carolina*BLU™ Gel & Buffer Stain
- ☐ 250 mL *Carolina*BLU™ Final Stain
- ☐ 4 latex gloves
- ☐ 6 staining trays

Needed but not supplied:

- ☐ 0.9% saline solution (NaCl), 10 mL per student, in 15-mL tube
- ☐ Micropipets and tips (1 µL to 1000 µL)
- ☐ 1.5-mL microcentrifuge tubes, polypropylene, 2 per student
- ☐ Microcentrifuge tube racks
- ☐ Microcentrifuge for 1.5-mL tubes
- ☐ 0.2-mL or 0.5-mL PCR tubes, 1 per student (1.5-mL microcentrifuge tubes may also be used.)
- ☐ 0.2-mL or 0.5-mL tube adapters for microcentrifuge (can be made from 0.5-mL and/or 1.5-mL tubes)
- ☐ Thermal cycler, programmable
- ☐ Electrophoresis chambers
- ☐ Electrophoresis power supplies
- ☐ Gel-staining trays
- ☐ UV transilluminator (ethidium bromide staining)
- ☐ White light box (*Carolina*BLU™ staining, optional)
- ☐ Camera or photo-documentary system (optional)
- ☐ Paper cup, 1 per student
- ☐ Permanent markers
- ☐ Container with cracked or crushed ice
- ☐ Boiling water bath (optional, see instructions)

*Ready-to-Go™ PCR Beads incorporate *Taq* polymerase, dNTPs, and $MgCl_2$. Each bead is supplied in an individual 0.5-mL tube or a 0.2-mL tube.

**Electrophoresis reagents must be purchased separately for Kits 21-1230 and 21-1230A.

Using an *Alu* Insertion Polymorphism to Study Human Populations

CONTENTS

STUDENT LAB INSTRUCTIONS

INTRODUCTION

Although DNA from any two people is more alike than different, many chromosome regions exhibit sequence differences between individuals. Such variable sequences are termed "polymorphic" (meaning many forms) and are used in the study of human evolution, as well as for disease and identity testing. Many polymorphisms are located in the estimated 98% of the human genome that does not encode protein.

This experiment examines a polymorphism in the human genome that is caused by the insertion of an *Alu* transposon, or transposable element. *Alu* is a member of the family of short interspersed elements (SINEs) and is approximately 300 nucleotides in length. *Alu* owes its name to a recognition site for the endonuclease *Alu*I in its middle. Although *Alu* is sometimes called a "jumping gene," it is not properly a gene, because it does not produce a protein product.

Alu transposons are found only in primate genomes and have accumulated in large numbers since primates diverged from other mammals. Human chromosomes contain more than one million *Alu* copies, equaling about 10% of the genome by mass. This accumulation was made possible by a transposition mechanism that reverse transcribes *Alu* mRNAs into mobile DNA copies. Another transposon, the long interspersed element (LINE) L1, supplies a specialized reverse transcriptase enzyme needed for *Alu* to jump. Hence, *Alu* and L1 exist in a sort of molecular symbiosis.

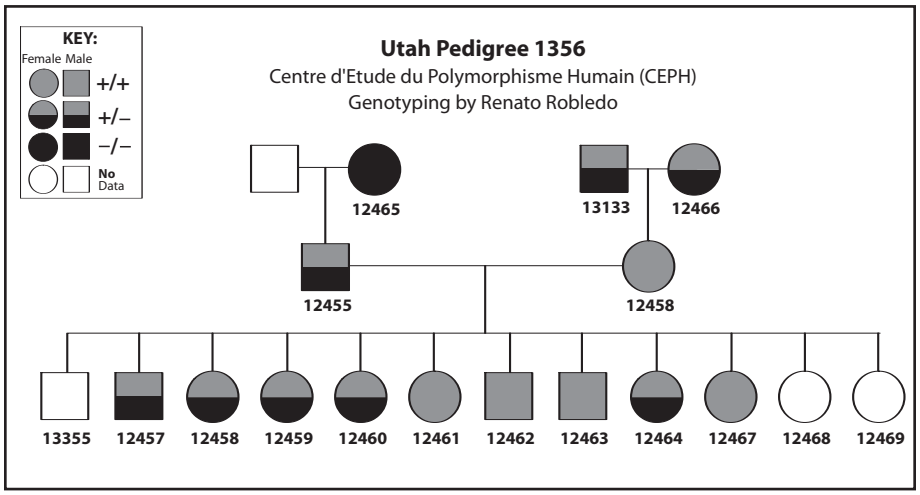
At any point in evolutionary time, only one or several *Alu* "masters" were capable of transposing. Although the rate of transposition was once much higher, a new *Alu* jump is estimated to now occur once per 200 live human births.

There is lively debate about whether *Alu* serves some larger purpose in primate genomes or is merely "selfish DNA" that has been successful in its mode of replication. *Alu* insertions in coding exons are implicated in a number of human diseases, including neurofibromatosis, thalassemia, cancer, and heart attack. However, the vast majority of *Alu*s are located in introns or intergenic regions, where they appear to have no phenotypic effect. *Alu*s in introns have had a potentially important impact on protein evolution: they provide alternative splice sites in approximately 5% of genes that produce multiple protein products.

Each *Alu* is the "fossil" of a unique transposition event that occurred once in primate history. After the initial jump, an *Alu* is inherited from parents by offspring in a Mendelian fashion. The vast majority of *Alu* insertions occurred millions of years ago and are "fixed." This means that, for a particular locus, all primates have inherited *Alu*s on each of the paired chromosomes.

However, several thousand *Alu*s have inserted in our genome since humans branched from other primates. Some of these are not fixed, meaning the *Alu* insertion may be present or absent on each of the paired

chromosomes, thus creating two possible alleles (+ and –). These "dimorphic" *Alu*s inserted within the last several hundred thousand years, reaching different allele frequencies in different human populations. Thus, *Alu* insertion polymorphisms are useful tools for reconstructing human evolution and migration.



Mendelian inheritance of the *Alu* insertion (+) at the PV92 locus.

This experiment examines a human *Alu* dimorphism at the PV92 locus. A sample of human cells is obtained by saline mouthwash (alternatively DNA may be isolated from hair sheaths). DNA is extracted by boiling with Chelex® resin, which binds contaminating metal ions. Polymerase chain reaction (PCR) is then used to amplify a chromosome region that contains the PV92 *Alu* dimorphism. The *Alu* insertion allele (+) is 300 nucleotides longer than the non-insertion allele (–), so the two alleles are readily separated by agarose gel electrophoresis.

Each student scores his or her genotype, and the compiled class results are used as a case study in human population genetics. Tools for testing Hardy-Weinberg equilibrium, comparing the PV92 insertion in world populations, and simulating the inheritance of a new *Alu* insertion are found on the included CD-ROM or at the *BioServers* Internet site of the Dolan DNA Learning Center (www.BioServers.org).

Batzer, M.A., Stoneking, M., Alegria-Hartman, M., Barzan, H., Kass, D.H., Shaikh, T.H., Novick, G.E., Iannou, P.A., Scheer, W.D., Herrera, R.J., and Deininger, P.L. (1994). African Origin of Human-specific Polymorphic *Alu* Insertions. *Proceedings of the National Academy of Sciences*. USA 91: 12288-12292.

Comas, D., Plaza, S., Calafell, F., Sajantila, A., and Bertranpetit, J. (2001). Recent Insertion of an *Alu* Element Within a Polymorphic Human-specific *Alu* Insertion. *Molecular Biology and Evolution* 18: 85-88.

Deininger, P.L. and Batzer, M.A. (1999). *Alu* Repeats and Human Disease. *Molecular Genetics and Metabolism* 67(3): 183-193.

Mullis, K. (1990). The Unusual Origin of the Polymerase Chain Reaction. *Scientific American* 262(4): 56-65.

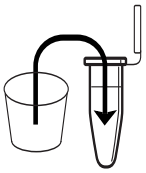
Prak, E.T.L. and Kazazian, H.H. (2000). Mobile Elements and the Human Genome. *Nature Reviews Genetics* 1(2): 134-144.

LAB FLOW

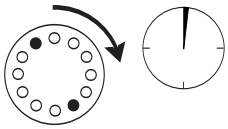
I. ISOLATE DNA FROM CHEEK CELLS

RINSE mouth with saline

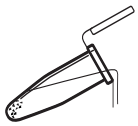
TRANSFER saline



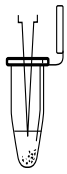
CENTRIFUGE



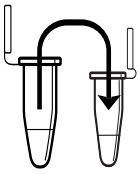
POUR OFF supernatant



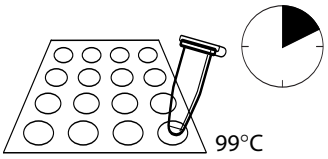
RESUSPEND



TRANSFER cell suspension

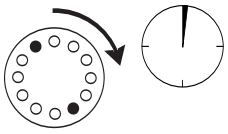


BOIL in thermal cycler

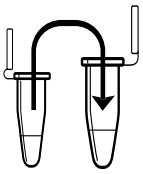


SHAKE vigorously

CENTRIFUGE



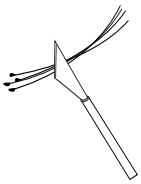
TRANSFER supernatant



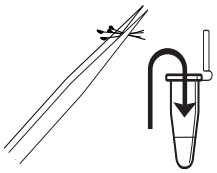
STORE on ice

(ALTERNATE) I. ISOLATE DNA FROM HAIR SHEATHS

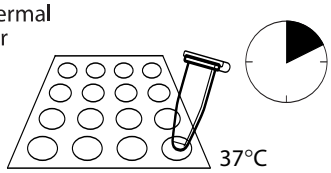
CUT hairs



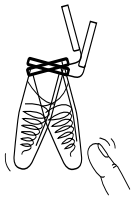
TRANSFER hairs



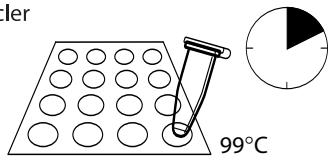
INCUBATE in thermal cycler



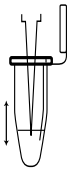
VORTEX



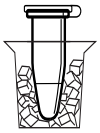
BOIL in thermal cycler



MIX by pipetting in and out

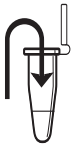


STORE on ice

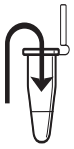


II. AMPLIFY DNA BY PCR

ADD primer/ loading dye mix

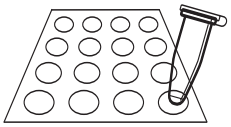


ADD DNA



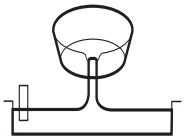
ADD mineral oil (if necessary)

AMPLIFY in thermal cycler



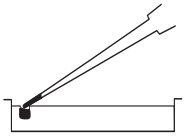
III. ANALYZE PCR PRODUCTS BY GEL ELECTROPHORESIS

POUR gel

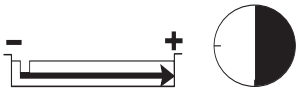


SET

LOAD gel



ELECTROPHORESE 130V



CAROLINA

METHODS

I. ISOLATE DNA FROM CHEEK CELLS

Reagents	Supplies and Equipment
0.9% Saline solution, 10 mL 10% Chelex®, 100 µL (in 0.2- or 0.5-mL PCR tube)	Permanent marker Paper cup Micropipets and tips (10–1000 µL) 1.5-mL microcentrifuge tubes Microcentrifuge tube rack Microcentrifuge adapters Microcentrifuge Thermal cycler (or water bath or heat block) Container with cracked or crushed ice Vortexer (optional)

1. Use a permanent marker to label a 1.5-mL tube and paper cup with your assigned number.
2. Pour saline solution into your mouth, and vigorously rinse your cheek pockets for 30 seconds.
3. Expel saline solution into the paper cup.
4. Swirl cup gently to mix cells that may have settled to the bottom. Use micropipet with fresh tip to transfer 1500 µL of the solution into your labeled 1.5-mL microcentrifuge tube.
5. Place your sample tube, along with other student samples, in a balanced configuration in a microcentrifuge, and spin for 90 seconds at full speed.
6. Carefully pour off supernatant into the paper cup. Try to remove most of the supernatant, but be careful not to disturb cell pellet at the bottom of the tube. (The remaining volume will approximately reach the 0.1 mark of a graduated tube.)

Before pouring off supernatant, check to see that pellet is firmly attached to tube. If pellet is loose or unconsolidated, carefully use micropipet to remove as much saline solution as possible.

Food particles will not resuspend.

Alternatively, you may add the cell suspension to Chelex in a 1.5-mL tube, and incubate in a boiling water bath or heat block.

Your teacher may instruct you to collect a sample of cell suspension to observe under a microscope.

7. Set micropipet to 30 µL. Resuspend cells in the remaining saline by pipetting in and out. Work carefully to minimize bubbles.
8. Withdraw 30 µL of cell suspension, and add to a PCR tube containing 100 µL of Chelex®. Label the cap and side of the tube with your assigned number.
9. Place your PCR tube, along with other student samples, in a thermal cycler that has been programmed for one cycle of the following profile. The profile may be linked to a 4°C hold program.

   Boiling step: 99°C 10 minutes

The near-boiling temperature lyses the cell and nuclear membranes, releasing DNA and other cell contents.

10. After boiling, vigorously shake the PCR tube for 5 seconds.

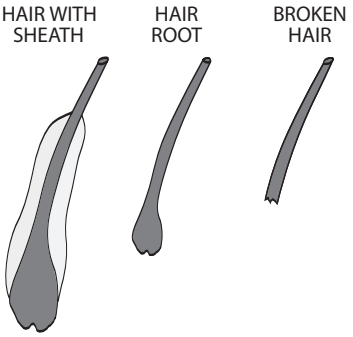
To use adapters, "nest" the sample tube within sequentially larger tubes: 0.2 mL within 0.5 mL within 1.5 mL. Remove caps from tubes used as adapters.

11. Place your tube, along with other student samples, in a balanced configuration in a microcentrifuge, and spin for 90 seconds at full speed. *If your sample is in a PCR tube, one or two adapters will be needed to spin the tube in a microcentrifuge designed for 1.5-mL tubes.*
12. Use a micropipet with fresh tip to transfer 30 µL of the clear supernatant into a clean 1.5-mL tube. Be careful to avoid pipetting any cell debris and Chelex® beads.
13. Label the cap and side of the tube with your assigned number. This sample will be used for setting up one or more PCR reactions.
14. Store your sample on ice or at –20°C until you are ready to continue with Part II.

I. (ALTERNATE) ISOLATE DNA FROM HAIR SHEATHS

Reagent	Supplies and Equipment
100 mg/mL proteinase K, 100 µL (in 0.2- or 0.5-mL tube)	Permanent marker Scalpel or razor blade Forceps or tweezers Thermal cycler (or water bath or heat block) Container with cracked or crushed ice Vortexer (optional)

Your teacher may instruct you to prepare a hair sheath to observe under a microscope.



1. Pull out several hairs and inspect for presence of a sheath. The sheath is a barrel-shaped structure surrounding the base of the hair, and can be readily observed with a hand lens or dissecting microscope. The glistening sheath can be observed with the naked eyes by holding the hair up to a light source. (Sheaths are most easily observed on dark hair.)
2. Select one to several hairs with good sheaths. Alternately, select hairs with the largest roots. *Broken hairs, without roots or sheaths, will not yield enough DNA for amplification.*
3. Use a fresh razor blade or scalpel to cut off hair shafts just above the sheath.
4. Use forceps to transfer hairs to a PCR tube containing 100 µL of proteinase K. *Make sure sheath is submerged in the solution and not stuck on the test tube wall.* Label the cap and side of the tube with your assigned number.

Alternatively, you may add the hairs to proteinase K in a 1.5-mL tube, and incubate in a water bath or heat block.

5. Place your PCR tube, along with other student samples, in a thermal cycler that has been programmed for one cycle of the following profile.

   Incubation Step: 37°C 10 minutes

6. Remove sample tube to room temperature. Vortex by machine or vigorously with finger for 15 seconds to dislodge cells from hair shaft.

7. Place your PCR tube, along with other student samples, in a thermal cycler that has been programmed for one cycle of the following profile. The profile may be linked to a 4°C hold program.

   Boiling step: 99°C 10 minutes

8. Remove sample tube to room temperature, and mix by pipetting in and out for 15 seconds.

9. Store your sample on ice or in the freezer until ready to begin Part II.

II. AMPLIFY DNA BY PCR

Reagents (at each student station)	**Supplies and Equipment**
*Cheek cell or hair sheath DNA 2.5 µL (from Part I) *PV92B primer/loading dye mix, 25 µL Ready-To-Go™ PCR beads (in 0.2-mL or 0.5-mL PCR tube)	Permanent marker Micropipet and tips (1-100 µL) Microcentrifuge tube rack Thermal cycler Container with cracked or crushed ice
Shared Reagent	
Mineral oil, 5 mL (depending on thermal cycler)	
*Store on ice	

1. Obtain a PCR tube containing a Ready-To-Go™ PCR Bead. Label with your assigned number.

The primer/loading dye mix will turn purple as the PCR bead dissolves.

2. Use a micropipet with fresh tip to add 22.5 µL of PV92B primer/loading dye mix to the tube. Allow the bead to dissolve for a minute or so.

If the reagents become splattered on the wall of the tube, pool them by pulsing in a microcentrifuge or by sharply tapping the tube bottom on the lab bench.

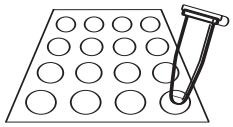
3. Use a micropipet with fresh tip to add 2.5 µL of your DNA (from Part I) *directly into* the primer/loading dye mix. Insure that no cheek cell DNA remains in the tip after pipetting.

4. *Store your sample on ice until your class is ready to begin thermal cycling.*

5. Place your PCR tube, along with other student samples, in a thermal cycler that has been programmed for 30 cycles of the following profile. The profile may be linked to a 4°C hold program after the 30 cycles are completed.

   Denaturing step: 94°C 30 seconds
   Annealing step: 68°C 30 seconds
   Extending step: 72°C 30 seconds

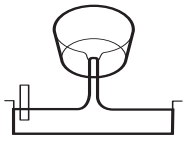
If your thermal cycler does not have a heated lid: Prior to thermal cycling, you must add a drop of mineral oil on top of your PCR reaction. Be careful not to touch the dropper tip to the tube or reaction, or the oil will be contaminated with your sample.



6. After cycling, store the amplified DNA on ice or at –20°C until you are ready to continue with Part III.

III. ANALYZE PCR PRODUCTS BY GEL ELECTROPHORESIS

Reagents	Supplies and Equipment
*PCR product (from Part II), 25 µL **Shared Reagents** *pBR322/*Bst*NI marker 1.5% agarose in 1× TBE, 50 mL 1× TBE, 300 mL Ethidium bromide (1 µg/mL), 250 mL *or* *Carolina*BLU™ Gel & Buffer Stain, 7 mL *Carolina*BLU™ Final Stain, 250 mL *Store on ice	Micropipet and tips (1–100 µL) Microcentrifuge tube rack Gel electrophoresis chamber Power supply Staining trays Latex gloves UV transilluminator (for use with ethidium bromide) White light transilluminator (for use with *Carolina*BLU™) Digital or instant camera (optional) Water bath (60°C) Container with cracked or crushed ice



Avoid pouring an overly thick gel, which is more difficult to visualize. The gel will become cloudy as it solidifies.

1. Seal the ends of the gel-casting tray with masking tape, and insert a well-forming comb.
2. Pour 1.5% agarose solution to a depth that covers about 1/3 the height of the open teeth of the comb.
3. Allow the gel to solidify completely. This takes approximately 20 minutes.
4. Place the gel into the electrophoresis chamber, and add enough 1× TBE buffer to cover the surface of the gel.

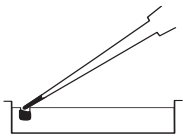
Do not add more buffer than necessary. Too much buffer above the gel channels electrical current over the gel, increasing running time.

5. Carefully remove the comb, and add additional 1× TBE buffer to just cover and fill in wells, creating a smooth buffer surface.

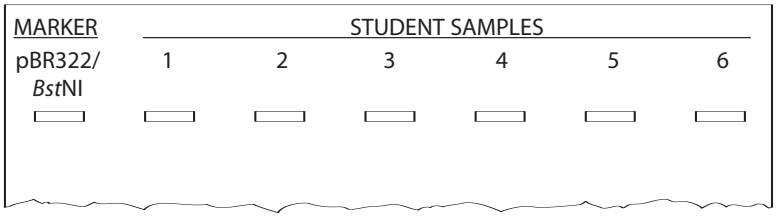
100-bp ladder may also be used as a marker.

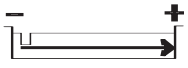
6. Use a micropipet with a fresh tip to load 20 µL of pBR322/*Bst*NI size marker into the far left lane of the gel.

Expel any air from the tip before loading. Be careful not to push the tip of the pipet through the bottom of the sample well.



7. Use a micropipet with a fresh tip to add 25 µL of your sample/loading dye mixture into your assigned lane of a 1.5% agarose gel, according to the diagram below. (If you used mineral oil during PCR, pierce your pipet tip through the layer of mineral oil to withdraw the PCR sample and leave the mineral oil behind in the original tube.)





8. Run the gel at 130 V for approximately 30 minutes. Adequate separation will have occurred when the cresol red dye front has moved at least 50 mm from the wells.

Destaining the gel for 5–10 minutes in tap water leaches unbound ethidium bromide from the gel, decreasing background and increasing contrast of the stained DNA.

9. Stain the gel using ethidium bromide or *Carolina*BLU™:
   a. For ethidium bromide, stain 10-15 minutes. Decant stain back into storage container for reuse, and rinse gel in tap water. *Use gloves when handling ethidium bromide solution and stained gels or anything that has ethidium bromide on it. Ethidium bromide is a known mutagen and care should be taken when using and disposing of it.*
   b. For *Carolina*BLU™, follow directions in the Instructor Planning section.

Transillumination, where the light source is below the gel, increases brightness and contrast.

10. View gel using transillumination, and photograph using a digital or instant camera.

BIOINFORMATICS

For a better understanding of the experiment, do the following bioinformatics exercises before you analyze your results.

Biological information is encoded in the nucleotide sequence of DNA. Bioinformatics is the field that identifies biological information in DNA using computer-based tools. Some bioinformatics algorithms aid the identification of genes, promoters, and other functional elements of DNA. Other algorithms help determine the evolutionary relationships between DNA sequences.

Because of the large number of tools and DNA sequences available on the Internet, experiments done *in silico* ("in silicon," or on the computer) now complement experiments done *in vitro* (in glass, or test tube). This movement between biochemistry and computation is a key feature of modern biological research.

In Part I you will use the Basic Local Alignment Search Tool (BLAST) to identify sequences in biological databases and to make predictions about the outcome of your experiments. In Part II you will identify additional alleles at the PV92 locus. In Part III you will discover the chromosome location of the PV92 insertion.

NOTE: The links in these bioinformatics exercises were correct at the time of printing. However, links and labels within the NCBI Internet site change occasionally. When this occurs, you can find updated exercises at http://bioinformatics.dnalc.org.

I. Use BLAST to Find DNA Sequences in Databases (Electronic PCR)

The following primer set was used in the experiment:

5'-GGATCTCAGGGTGGGTGGCAATGCT-3' (Forward Primer)
5'-GAAAGGCAAGCTACCAGAAGCCCCAA-3' (Reverse Primer)

1. Initiate a BLAST search.
   a. Open the Internet site of the National Center for Biotechnology Information (NCBI) www.ncbi.nlm.nih.gov/.
   b. Click on *BLAST* in the top speed bar.
   c. Click on the link *nucleotide BLAST* under the heading *Basic BLAST.*
   d. Enter the sequences of the primers into the *Search* window. These are the query sequences.
   e. Omit any non-nucleotide characters from the window, because they will not be recognized by the BLAST algorithm.
   f. Under *Choose Search Set*, select the *Nucleotide collection (nr/nt) database* from the drop-down menu.

g. Under *Program Selection*, optimize for somewhat similar sequences by selecting *blastn*.

h. Click on *BLAST!* and the query sequences are sent to a server at the National Center for Biotechnology Information in Bethesda, Maryland. There, the BLAST algorithm will attempt to match the primer sequences to the millions of DNA sequences stored in its database. While searching, a page showing the status of your search will be displayed until your results are available. This may take only a few seconds, or more than a minute if a lot of other searches are queued at the server.

2. The results of the BLAST search are displayed in three ways as you scroll down the page:

   a. First, a graphical overview illustrates how significant matches, or hits, align with the query sequence. Matches of differing lengths are coded by color. What do you notice?

   b. This is followed by a list of *significant alignments*, or hits, with *Accession* information.

   c. Next, is a detailed view of each primer sequence (*query*) aligned to the nucleotide sequence of the search hit (*subject*). Notice that a match to the forward primer (nucleotides 1–25), and a match to the reverse primer (nucleotides 26–51) are within the same *Accession*.

3. What is the predicted length of the product that the primer set would amplify in a PCR reaction (*in vitro*)?

   a. In the list of *significant alignments*, notice the scores in the *E-value* column on the right. The *Expectation* or *E-value* is the number of alignments with the query sequence that would be expected to occur by chance in the database. The lower the *E-value* the higher the probability that the hit is related to the query.

   b. Note the names of any *significant alignments* that have *E-values* less than 0.1. Do they make sense?

   c. Scroll down to the *Alignments* section to see exactly where the two primers have landed in this subject sequence.

   d. The lowest and highest nucleotide positions in the subject sequence indicate the borders of the amplified sequence. Subtracting one from the other gives the difference between the two coordinates.

   e. However, the actual length of the fragment *includes* both ends, so add 1 nucleotide to the result to determine the exact length of the PCR product amplified by the two primers.

   f. Is this the + or the – allele?

4. Now, take a closer look at this database hit, and copy its sequence for future use.

   a. Click on the *Accession* link at the left to open the sequence datasheet for this hit.

b. At the top of the report, note basic information about the sequence, including its basepair length, database accession number, source, and references.

c. The bottom section of the report lists the entire nucleotide sequence of the gene or DNA sequence that contains the PCR product. Highlight all the nucleotides between the beginning of the forward primer and end of reverse primer. Paste this sequence into a text document. Then, trim any extra nucleotides from the ends, and delete all non-nucleotide characters and spaces. This is the amplicon, or amplified product.

II. Use BLAST to Identify Additional Alleles at the PV92 Locus

1. Return to the nucleotide BLAST page.
2. Paste the 416-bp PV92 amplicon, from 4.c. above, into the search window. Ensure that *Nucleotide collection (nr/nt)* and *blastn* are selected, then click on *BLAST!*
3. Wait until the BLAST results are displayed.
4. What do you notice about the *E-values* obtained by this search? Why is this so?
5. Why does the first hit have an *E-value* of 0?
6. Now focus on the hit named "Human *Alu* repeat"; this is the *Alu* insertion at PV92.
   a. Follow the Accession link, then click on repeat_region 77..384/rpt_family="Alu" in the Features section . What do you notice about the 3' end of the *Alu* repeat?
   b. Also in the *Features* section, look at the "insertion target sequence" on either side of the *Alu* repeat. What appears to be going on?
7. What is the length of the *Alu* inserted at PV92?
8. If you assume that the amplicon in Part I is the – allele, what is the length of the + allele?
9. Now look carefully at the hit named "Homo sapiens isolate BAS101 AluPV92 repeat sequence." Examine the *Features* and follow links. What is going on here? How are the three hits related to one another?

III. Use *Map Viewer* to Determine the Chromosome Location of the PV92 Insertion

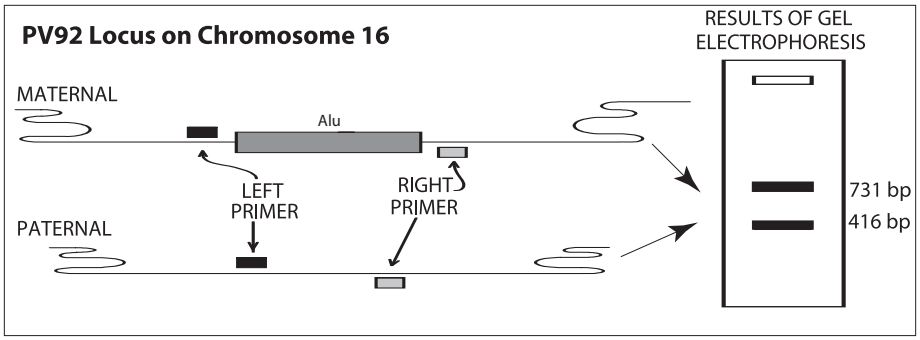
1. Return to the NCBI home page, then click on *Map Viewer* located in the *Hot Spots* column on the right.
2. Find *Homo sapiens (humans)* in the table to the right and click on the "B" icon under the *Tools* header. If more than one build is displayed,

select the one with the highest number, as this will be the most recent version.

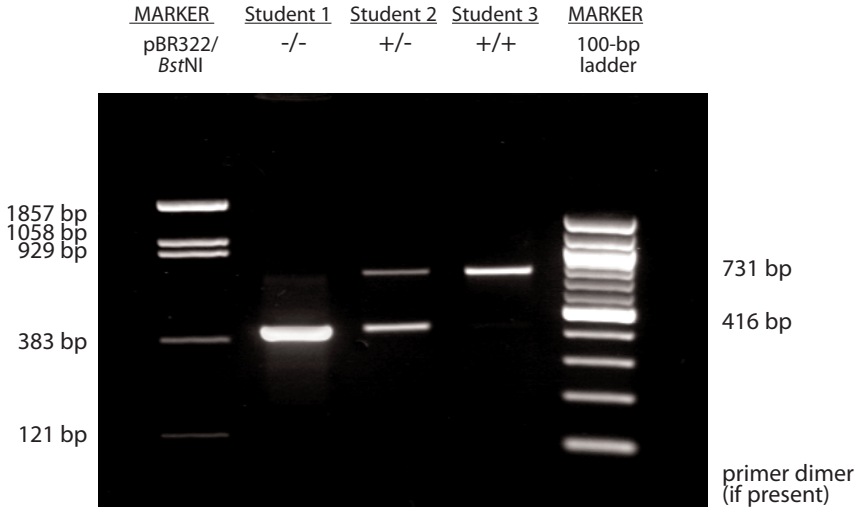
3. Paste the 416-bp amplicon (from Part I) into the search window. (Primers usually are not long enough to produce a result in the map BLAST.)
4. Select *BLASTN* from the drop-down menu under *Program* and click on *Begin Search*.
5. Click on *View report* to retrieve the results.
6. Click on *[Human genome view]* in the list of *Other reports* at the top of the page to see the chromosome location of the BLAST hit. On what chromosome have you landed?
7. Click on the marked chromosome number to move to the PV92 locus. Click on the small blue arrow labeled *Genes seq* to display genes. The 416-bp amplicon (red) occupies the whole field of the default view. What can you say about the gene that contains the amplicon? Click on the name under the *Symbol* track, and then follow links to find out.
8. Use the *zoom out* toggle on the left to get a better perspective on the CDH13 gene. Introns and noncoding sequences are denoted by a thin line, while exons are denoted by thick bar.
   a. Determine the size of the CDH13 gene using the map coordinates to the left of the contig map.
   b. How many introns and exons does CDH13 gene have?
   c. Where in the CDH13 gene is PV92 *Alu* inserted: an exon or intron?
   d. How does this explain the fact that the PV92 insertion is believed to be neutral, i.e., to have no phenotypic effect?

RESULTS AND DISCUSSION

The following diagram shows how PCR amplification identifies the *Alu* insertion polymorphism at the PV92 locus.



1. **Determine your PV92 genotype.** Observe the photograph of the stained gel containing your PCR samples and those from other students. Orient the photograph with the sample wells at the top. Use the sample gel shown below to help interpret the band(s) in each lane of the gel.



   a. Locate the lane containing the pBR322/*Bst*NI markers on the left side of the sample gel. Working from the well, locate the bands corresponding to each restriction fragment: 1857 bp, 1058 bp, 929 bp, 383 bp, and 121 bp. The 1058-bp and 929-bp fragments will be very close together or may appear as a single large band. The 121-bp band may be very faint or not visible. *(Alternatively, use a 100-bp ladder as shown on the right-hand side of the sample gel. These DNA markers increase in size in 100-bp increments starting with the fastest migrating band of 100 bp.)*

   b. Scan across the row of student results that contains your sample. You should notice that virtually all student lanes contain one or two prominent bands.

c. To "score" your genotype, compare your PCR product with the markers and other types in your row. The analysis will be simple if your row contains a heterozygous type (+/–) that shows the positions of both alleles. Homozygotes of each type (+/+ and –/–) will also help. If your row contains only a single homozygous type, you will need to rely entirely on markers to determine which allele it is.

+/– (heterozygous) Shows two prominent bands. The + allele (731 bp) should be slightly ahead of the 929-bp marker. The – allele (416 bp) should be about even with the 383-bp marker.

+/+ (homozygous) Shows a single prominent band slightly ahead of the 929-bp marker.

–/– (homozygous) Shows a single prominent band about even with the 383-bp marker.

d. It is common to see a diffuse (fuzzy) band that runs ahead of the 121-bp marker. This is "primer dimer," an artifact of the PCR reaction that results from the primers overlapping one another and amplifying themselves. The presence of primer dimer, in the absence of other bands, confirms that the reaction contained all components necessary for amplification.

e. Additional faint bands at other positions occur when the primers bind to chromosomal loci other than the PV92 locus and give rise to "nonspecific" amplification products.

2. An *Alu* insertion has only two states: + and –. How does this relate to information stored in digital form by a computer? What equivalent in digital information is provided by an *Alu* genotype?

3. **Determine the observed genotype and allele frequencies for your class.** Use the chart below to record your answers to the questions that follow.

Genotype Frequency	# Students	Genotype	+ Allele (#)	– Allele (#)
		+/+		
		+/–		
		–/–		
TOTALS>				
		Allele Frequency>		

a. Count the number of students of each genotype: +/+, +/–, and –/–. Exclude from the analysis any students whose genotypes could not be determined.

b. Calculate the frequency of each genotype, where

$$\text{genotype frequency (\%)} = \frac{\text{number of students of X genotype}}{\text{total student samples}}$$

c. Calculate the frequency of each allele, where

$$\text{allele frequency (\%)} = \frac{\text{number of X alleles}}{\text{total alleles in sample}}$$

First, multiply the number of students of each genotype by the number of + or – alleles in that genotype. Remember that each +/+ or –/– student contributes 2 copies of that allele, while each +/– student contributes one of each allele. Then add up the total number of copies of each allele. The TOTAL number of alleles in the sample is twice the number of students.

4. Is the + allele confined to any particular racial or ethnic group? What can you say about people in the class who have at least one + allele?

5. **Calculate genotype frequencies expected for your class under Hardy Weinberg Equilibrium.** Under certain conditions a population comes into genetic equilibrium, where the genotype frequencies at a single locus remain constant over time. The Hardy-Weinberg equation describes the genotype frequencies that are expected in a population at equilibrium:

$$p^2 + 2pq + q^2 = 1$$

where p and q represent the allele frequencies; p^2 and q^2 are the homozygote frequencies; and 2pq is the heterozygote frequency.

a. Use the allele frequencies calculated for your class in Step 2 to determine the genotype frequencies expected under Hardy-Weinberg equilibrium. Make + = p and – = q in the equation.

b. How do genotype frequencies you observed in your experiment compare with those expected by the Hardy-Weinberg equation? Would you say they are very similar or very different?

6. **Enter your class data into the *Allele Server* Database**. Population statistics are tedious to calculate by hand, but are easily accomplished by algorithms at the *BioServers* Internet site. First, you need to enter your data into a class file that has been set up by your teacher.

For the teacher: To enter student data, you must first register with *Allele Server* and set up a class account.

Click on *Manage Groups*, then wait while the existing data loads. This may take a moment. Select *Your Groups* from the pull-down menu. Click *ADD GROUP*. Provide the requested information, and be sure to make the group *Public*. Then create a password, and enter the number of students who will submit data. Click OK. The class now appears in the list of *Your Groups* and can now be accessed by class members.

a. Open the *BioServers* Internet site at the Dolan DNA Learning Center www.BioServers.org.

b. Enter *Allele Server*. You can register if you want to save your work for future reference, but it is not required.

c. The interface is simple to use: add or obtain data using the top buttons and pull-down menus, then work with the data in the workspace below.

d. Click on the *ADD DATA* at the top of the page, and find your group in the pull-down menu. Enter the password supplied by your teacher and your sample number. Then click *OK*.

e. Use the pull-down menus to add your sex, descent, and genotype. Then click *OK*. Your data has been added to your group.

7. **Test Hardy-Weinberg Equilibrium in your class.** A Chi-square test is used to compare observed genotype frequencies with those predicted by the Hardy-Weinberg equation.
   a. Click on *Manage Groups*, then wait while the existing data loads. This may take a moment.
   b. Find your class in the list, and click on the check box to select it.
   c. Click *OK*, and your class data are moved into the workspace.
   d. Click *OPEN* to get basic information on your population: number in the sample, frequencies of the + and – alleles, and frequencies of the three genotypes +/+, +/–, and –/–.
   e. Mark the dot to the right of your group name, and click ANALYZE.
   f. The pie chart provides a visual comparison of your observed *versus* expected results. When you ask yourself if the sections of the two pies are substantially similar or rather different, you are doing an informal Chi-square analysis.
   g. The Chi-square statistic tests the "null hypothesis"—that there is no significant difference between observed and expected genotype frequencies. The Chi-square result at the top of the page is associated with a p-value or probability that observed and expected frequencies are substantially alike and that frequency differences are merely due to chance. Scientists generally accept that the results are statistically significant at a p-value of 0.05 or less. This technically means there is only a five percent chance that such results could be obtained by chance, or, more to the point, that the observed differences in genotype frequencies are likely real.
   h. Is your p-value greater or less than the 0.05 cut off? What does this mean?
   i. What conditions are required for a population to come into genetic equilibrium? Does your class satisfy these requirements?
8. **Compare genotype frequencies in world populations.** The Chi-square statistic is also used to compare the genotype frequencies of two populations. A p-value of 0.05 or less indicates that two populations have significantly different genetic structure.
   a. Click on *Manage Groups*, then wait for the existing data to load.
   b. Select *Reference* from the pull-down menu, to get a list of PV92 experiments that have been conducted by scientists with people from a number of relatively distinctive populations from around around the world.
   c. Browse the list, and click on the check boxes of a number of populations that interest you. Take samples that represent different continents and regions of the world.
   d. Press *OK* to move the populations into the workspace.
   e. Test Hardy-Weinberg equilibrium in any population by marking the

dot in the right-hand column and clicking *ANALYZE*. (Only one population can be tested at a time.)

f. Next, compare your class to one of the world populations, by checking the appropriate boxes in the left-hand column and clicking *COMPARE*. (Only two populations can be compared at a time.)

g. Do the pie charts look similar or different? Does the Chi-Square statistic and associated p-value support your visual impression?

h. Continue on comparing your class to other world populations. Also compare any two reference populations. Uncheck populations you are finished with.

i. Which groups have significantly different genotype frequencies? What is the most frequent genotype in each group?

9. **Compare allele frequencies in world populations.** Genetic distance is a relatively simple statistic that uses differences in allele frequency to gauge the relative distance that separates two populations in genetic space, 0 being the least distance and 1 being the greatest.

a. Click on the check boxes to select any two populations you selected in Question 8 above.

b. Select *Fst Genetic Distance* from the pull-down window next to the *COMPARE* button.

c. Then click *COMPARE*.

d. Compare the pie charts with the calculated genetic distance.

e. Continue comparing populations you selected in Question 8 above, and *note the + allele frequency for each*. (You can also obtain the + allele frequency by clicking the *OPEN* button next to each population.)

f. Now, plot the + allele frequency for each group on the map of world populations (page 24).

g. Do you notice any pattern in the allele frequencies?

h. Suggest a hypothesis about the origin and dispersal of the *Alu* allele that accounts for your observation.

i. Calculations suggest that the original *Alu* insertion at the PV92 locus occurred about 200,000 years ago. If this is so, in what sort of hominid did the jump occur, and what implications does this have for your hypothesis from h. above?

10. **Simulate a new *Alu* jump in an ancient hominid population.** In this experiment, you will simulate the sort of populations in which the PV92 insertion occurred about 200,000 years ago. A Hardy-Weinberg simulator will allow you to model population changes over time. In each generation, parents are chosen at random and offspring are generated using an approach similar to a Punnett Square analysis. The survival rate of a particular genotype (+/+, +/-, or -/-) determines

the probability that an individual will reproduce in his/her generation. This process is repeated in each generation, producing enough offspring to maintain the population at a constant size.

a. Enter *Simulation Server* from the *BioServers* homepage. Wait while the Java applet loads on your computer.

b. Create a node (#1) by clicking in the white workspace. The node represents a human population.

c. The red circle indicates that the parameters for Node #1 are available for editing in the right-hand control panel. Think about how to represent this population at the start of the simulation.

d. How did hominids live 200,000 years ago, and what size population group would be supported? Enter this number into the *Starting pop.* Window at the top right.

e. What would be the allele frequency if a new *Alu* jump occurred in a group of this size? Enter this number into the *Starting %* "+" window.

f. Leave the *# Generations* at 100.

g. Assume that this *Alu* jump is neutral and has no effect on gene expression. So, leave the *Survival %* for each genotype at 100%. This means that individuals with each of the three genotypes have equal chance of surviving to reproduce.

h. At the top of the window, set the *# Runs* to 100. The computer will do 100 experiments with these parameters. You can think of this as 100 different population groups in which a new *Alu* jump occurs. These 100 groups would be equivalent to estimates of the size of the entire hominid population in Africa during bottlenecks before the advent of agriculture.

i. Click the *Enter Values* button to program the node.

j. Click on the *Begin Run* button at the top left. Don't touch or move the screen until the calculations are complete, or the application may freeze. The progress of the run is indicated in *% Complete* at the top of the window.

k. Scroll down to see the results of the simulation. The histogram is difficult to interpret, so click on the *Graph* tab at the upper left. Then check *Node #1*, and click on *Press here to graph*.

l. Allele frequency is on the Y axis and generations are on the X axis. Each blue line traces one population over 100 generations.

m. What happens to the new *Alu* insertion in the 100 populations?

n. Follow the allele frequency in one population over 100 generations. What happens to the allele frequency, and what causes this?

o. Try another experiment with the same parameters. Scroll to the top of the page, click on the *Restart* and *Begin Run* button.

11. **Simulate population expansion.** Next, find out what happens to an *Alu* insertion when a small population expands dramatically. This simulates what happened to neutral alleles when hunter-gatherer groups became agriculturalists and settled down to form the first urban centers. It also illustrates the so called "founder effect," the effect on an allele frequency when a large population is derived from a small group of original settlers.
    a. Click restart, then click on the workspace to add *Node #2*.
    b. With *Node #2* active, change one parameter in the right-hand column. Enter 2000 in the *Starting pop. Window*. Then click *Enter Values* to program the node.
    c. Change the second window in the lower right corner to read *Link 1 to 2*. Click on the *Link* button, and a red line will appear between *Nodes 1* and *2*.
    d. In the link mode, *Node #1* feeds its results into *Node #2*. So the initial population mates randomly for 100 generation then feeds the resulting + allele frequency into an expanded population, which mates for an additional 100 generations at *Node #2*. (This is why the *Starting %* "+" is inactivated in *Node #2*.)
    e. Click on the *Begin Run* button at the top left. The calculations take longer with the larger population, so be patient.
    f. When the calculations are complete, scroll down to see the results.
    g. In the graph mode, check *Node #1*, *Node #2*, and *Graph Linked*. Then click on *Press here to graph*.
    h. The left-hand side of the graph shows the first 100 generations of the small population, and the right-hand side shows the next 100 generations as a larger population.
    i. What do you notice about the allele frequency in those populations that maintain the + allele over 200 generations?
    j. Click on the *Restart* and *Begin Run* button to see another set of experiments with the same parameters
12. Add additional nodes to simulate other effects, such as population bottlenecks, or create scenarios in which the + allele confers some survival advantage or disadvantage.

INFORMATION FOR INSTRUCTOR

CONCEPTS AND METHODS

This laboratory can help students understand several important concepts of modern biology:

- How to collect and analyze genetic information in populations.
- The use of allele and genotype frequencies to test Hardy-Weinberg equilibrium.
- The use of DNA polymorphisms in the study of human evolution.
- Identity by descent from a common ancestor.
- The movement between *in vitro* experimentation and *in silico* computation.

The laboratory uses several methods for modern biological research:

- DNA extraction and purification.
- Polymerase chain reaction (PCR).
- Gel electrophoresis.
- Bioinformatics.

LAB SAFETY

The National Association of Biology Teachers recognizes the importance of laboratory activities using human body samples and has developed safety guidelines to minimize the risk of transmitting serious disease. ("The Use of Human Body Fluids and Tissue Products in Biology," *News & Views*, June 1996.) These are summarized below:

- Collect samples only from students under your direct supervision.
- Do not use samples brought from home or obtained from an unknown source.
- Do not collect samples from students who are obviously ill or are known to have a serious communicable disease.
- Have students wear proper safety apparel: latex or plastic gloves, safety glasses or goggles, and lab coat or apron.
- Supernatants and samples may be disposed of in public sewers (down lab drains).
- Have students wash their hands at the end of the lab period.
- Do not store samples in a refrigerator or freezer used for food.

The risk of spreading an infectious agent by this lab method is much less likely than from natural atomizing processes, such as coughing or sneezing. Several elements further minimize any risk of spreading an infectious agent that might be present in mouthwash samples:

- Each experimenter works only with his or her sample.
- The sample is sterilized during a 10-minute boiling step.
- There is no culturing of the samples that might allow growth of pathogens.
- Samples and plasticware are discarded after the experiment.

INFORMED CONSENT AND DISCLOSURE

Student participation in this experiment raises real-life questions about the use of personal genetic data: What is my DNA sample being used for? Does my DNA type tell me anything about my life or health? Can my data be linked personally to me?

There is consensus that a human DNA sample should be obtained only with the willing consent of a donor, who understands the purpose for which it is being collected. Thus, this experiment should be explained ahead of time and students given the option to refrain from participating. (Some teachers may wish to have parents sign a consent form, such as those filled out for a field trip.) There is also consensus that a DNA sample be used only for the express purpose for which it is collected. Thus, student DNA samples should be thrown away after completing the experiment.

The PV92 polymorphism was specifically selected for this experiment because it is phenotypically neutral—it has no known relationship to any trait, disease state, or sex determination.

PV92 alleles are inherited in a Mendelian fashion and can give indications about family relationships. To avoid the possibility of suggesting inconsistent inheritance, it is best not to generate genotypes from parent-child pairs. In any event, this two-allele system would be less likely to turn up an inconsistency than the ABO blood groups. Furthermore, the chance that student samples can be mixed up when isolating DNA, setting up PCR reactions, and loading electrophoresis gels provides no certainty to any of the genotypes obtained in the experiment. (A forensic laboratory would use approved methods for maintaining "chain of custody" of samples and for tracking samples.)

INSTRUCTOR PLANNING, PREPARATION, AND LAB FINE POINTS

The following table will help you to plan and integrate the four parts of the experiment.

Part		Day	Time	Activity	
I.	Isolate DNA	1	60 min.	**Pre-lab:**	Prepare and aliquot saline solution Prepare and aliquot 10% Chelex® Aliquot proteinase K (alternate) Make centrifuge adapters Set up student stations
			30 min.	**Lab:**	Isolate student DNA
II.	Amplify DNA by PCR	1	15 min. 15 min. 60–150 min.	**Pre-lab:** **Lab:** **Post-lab:**	Aliquot PV92B primer/loading dye mix Set up PCR reactions Amplify DNA in thermal cycler
III.	Analyze PCR Products by Gel Electrophoresis	2 3	15 min. 30 min. 15 min 20+ min. 20+ min. 30–45 min. to overnight 20 min.	**Pre-lab:** **Lab:** **Post-lab:**	Dilute TBE electrophoresis buffer Prepare agarose gel solution and cast gels. Load DNA samples into gel Electrophorese samples Stain gels De-stain gels (for *Carolina*BLU™) Photograph gels
	Results and Discussion	4	30-60 min.	Score PV92 genotypes; determine class genotype and allele frequencies	

I. ISOLATE DNA FROM CHEEK CELLS

Saline mouthwash is the most reproducible of the simple methods to obtain human DNA for PCR. The mouthwash gently loosens a large number of single cells and small clusters of cheek cells. This maximizes the surface area of cells, allowing for virtually complete lysis during boiling. Cheek brushes and swabs generally yield larger clumps of cells, which are less effectively lysed by boiling.

With careful lab management, up to 90% of students should be able to "score" their *Alu* genotypes using the mouthwash method. Be especially watchful after the initial centrifugation step. Most students will have compact pellets that stay attached to the tube when the supernatant is poured off. However, about 10% of students will have diffuse or slimy masses that do not pellet well. Centrifuge these samples again, then carefully pipet out as much supernatant as possible. Surprisingly, food particles rinsed out with the mouthwash have little effect on PCR amplification. Still, it is best to avoid eating before the experiment, because food particles, especially from fruits, may block the pipet tip and make pipetting difficult.

It is worth a diversion to allow students to view their own squamous epithelial cells under a compound microscope. Add several µL of suspension remaining after Step I. 8. to a microscope slide, add a drop of 1% methylene blue (or other stain), and add a cover slip.

DNA is liberated from cheek cells by boiling in 10% Chelex®, which binds contaminating metal ions that are the major inhibitors of PCR. The boiling step is most easily accomplished using the same thermal cycler used for PCR. To do this, provide each student with 100 µL of 10% Chelex® suspension in a PCR tube that is compatible with the thermal cycler you will be using: either 0.2 mL or 0.5 mL. It is not necessary to use a "thin-walled" tube. Alternatively, use 1.5-mL tubes in a heat block or a boiling water bath. *Watch out for lids opening as the tubes heat*. (Make a simple water bath by maintaining a beaker of water at a low boil on a hot plate. Place 1.5-mL tubes in a floating rack or in holes punched in a double layer of aluminum foil over the top. If using aluminum foil, insure that tubes are immersed, and add hot water as necessary to maintain water level.)

Pre-lab Preparation

Prepare saline by dissolving 0.9 g NaCl in 100 mL distilled or deionized water. For each student, aliquot 10 mL into a 15-mL polypropylene tube.

Prepare 10% Chelex® by adding 15 mL distilled or deionized water to 1.5 g of Chelex®. For each student, aliquot 100 µL of 10% Chelex® into either a 0.2-mL or 0.5-mL tube (whichever format is accommodated by your thermal cycler). Alternatively, use a 1.5-mL microcentrifuge tube if you are planning to use a heat block or water bath instead of a thermal cycler. *The Chelex® resin quickly settles, so be sure to shake the stock tube to re-suspend the Chelex® each time before pipetting a student aliquot.*

Remove caps from 1.5-mL tubes to use as adapters in which to centrifuge the 0.5-mL PCR tubes used for Chelex® extraction. Two adapters are needed to spin 0.2-mL PCR tubes—a capless 0.5-mL PCR tube is nested within a capless 1.5-mL tube.

Pre-lab Set Up for DNA Isolation from Cheek Cells (per student station)

Saline solution (0.9% NaCl) tubes, 10 mL (in 15 mL tube)
10% Chelex®, 100 µL (in 0.2 or 0.5 mL tube, depending on thermal cycler)
2 1.5-mL microcentrifuge tubes
Permanent marker
Micropipets and tips (10–1,000 µL)
Microcentrifuge tube rack
Container with cracked or crushed ice
Paper cup

Shared Items

Microcentrifuge
Microcentrifuge adapters for 0.2-mL or 0.5-mL PCR tubes
Thermal cycler
Vortexer (optional)

I. (ALTERNATE) ISOLATE DNA FROM HAIR SHEATHS

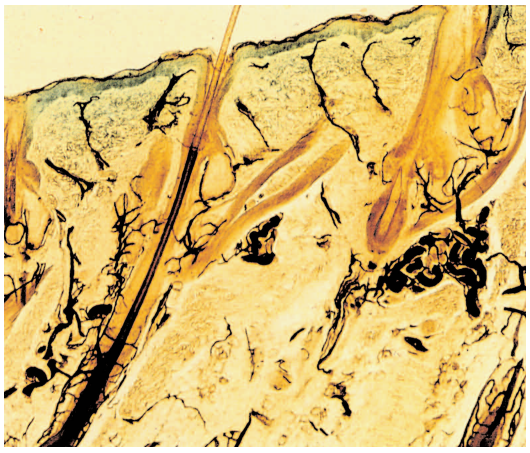
Hair roots provide the simplest source of DNA for PCR amplification; no special equipment is required for extraction. Hairs also are an extremely safe source of cells. Risk of spreading an infectious agent is minimized by "dry" collection, which does not involve any body fluid or generate any supernatant. This method also stresses the power of PCR in forensic cases—even one growing hair root provides enough DNA for excellent amplification.

HOWEVER, forensic biologists generally rate hair as a poor source of DNA for analysis, for the same reason that it can prove difficult in the classroom. Most plucked or shed hairs are broken off from the root, which is the source of cells for DNA extraction.

The success of this method is entirely dependent upon finding large roots from growing hairs. This can be tricky and time consuming—if often hilarious. With vigilance, up to 80% of students may find hairs with good roots from which to isolate DNA. However, it is more likely that only about 60–70% of students ultimately will be able to score their *Alu* genotypes using this method.

A hair is anchored in the skin by a follicle, or "root," whose growing cells produce the hair shaft. Hair goes through a growth cycle with alternating periods of growth and quiescence during which the follicle increases and decreases in size. During the growth phase, the follicle extends up the hair shaft in a structure called the sheath. The sheath is a rich source of cells. The sheath membrane is easily digested by treatment with proteinase K, releasing sqaumous cells singly or in small clusters. A high percentage of these cells are lysed by boiling and release DNA.

The sheath decreases in size as the hair follicle enters a resting stage (see drawing and micrograph of growing and resting follicles). The withered bulb of a resting follicle is, in fact, what most people would consider a "root." Resting follicles usually yield little DNA for analysis. First, there are fewer cells. Second, proteinase K treatment does not effectively digest the shriveled root mass, and only cells at the edge are lysed by boiling.



Successful amplification of the PV92 locus, which is available in only two copies per cell, is closely correlated to presence of a sheath on the hair shaft. One or two hairs with long sheaths will provide plenty of DNA for PCR amplification. Three or four good sized roots will usually work, especially if they have at least small sheaths.

A good sheath is unmistakable. Especially contrasted on a dark hair, it glistens when held up to the light and extends several mm up the hair shaft. Make sure to show off the first several

good sheaths that turn up, so other students will know what to look for. Because of the hair growth cycle, most people find sheaths only on some hairs. Students whose hair grows slowly may have difficulty finding sheaths, and thin or brittle hair is likely to break off before the root. If students are having difficulty finding sheaths on hairs pulled from their scalps, have them try hairs from the eyebrow or arm.

Sheaths are the most underrated source of squamous cells for microscopic examination. Give them a try! Simply place a sheath on a microscope slide and add a drop of proteinase K (100 mg/mL). Let stand for several minutes, to allow the proteinase K to digest the sheath membrane. Then add a drop of methylene blue or other cell stain, add a cover slip, and gently press to disrupt the sheath membrane. Observe under medium power and at several time points, to see the effect of enzyme digestion. If you gently press the cover slip while the slide is on the microscope stage, you should be able to observe squamous cells squirting out of tears in the sheath membrane.

Prelab Preparation

For each student, aliquot 100 µL of 100 mg/mL proteinase K into either a 0.2-mL or 0.5-mL tube (whichever format is accommodated by your thermal cycler). Alternatively, use a 1.5-mL microcentrifuge tube if you are planning to use a heat block or water bath instead of a thermal cycler.

Pre-lab Set Up for DNA Isolation from Hair Sheaths (per student station)

100 mg/mL proteinase K, 100 µL (in 0.2- or 0.5-mL PCR tube)
Permanent marker
Scalpel or razor blade
Forceps or tweezers

Shared Items

Thermal cycler (or water bath or heat block)
Container with cracked or crushed ice
Vortexer (optional)

II. AMPLIFY DNA BY PCR

The primer/loading dye mix incorporates the appropriate primer pair (0.26 picomoles/µL of each primer), 13.8% sucrose, and 0.0081% cresol red. The inclusion of the loading dye components, sucrose and cresol red, allows the amplified product to be directly loaded into an agarose gel for electrophoresis. Each Ready-To-Go™ PCR Bead contains reagents so that when brought to a final volume of 25 µL, the reaction contains 2.5 units of *Taq* DNA polymerase, 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 1.5 mM $MgCl_2$, and 200 µM of each dNTP.

The lyophilized *Taq* DNA polymerase in the bead becomes active immediately upon addition of the primer/loading dye mix and template DNA. In the absence of thermal cycling, "nonspecific priming" at room temperature allows the polymerase to begin generating erroneous products, which can show up as extra bands in gel analysis. *Therefore, work quickly. Be sure the thermal cycler is set and have all experimenters set up their PCR reactions as a coordinated effort. Add primer/loading dye mix to all reaction tubes, then add each student template, and begin thermal cycling as quickly as possible. Hold reactions on ice until all student samples are ready to load into the thermal cycler.*

PCR amplification from crude cell extracts is biochemically demanding, and requires the precision of automated thermal cycling. However, amplification of the PV92 locus is not complicated by the presence of repeated units. Therefore, the recommended amplification times and temperatures will work adequately for most common thermal cyclers, which ramp between temperatures within a single heating/cooling block. *IMPORTANT*: A different cycling profile is required for Robocycler or other brands of thermal cyclers

that physically move PCR reaction tubes between multiple temperature blocks. Because there is no ramping time between temperatures, these machines require the longer cycling times listed below:

Denaturing step:	94°C	1 minute
Annealing step:	68°C	2 minutes
Extending step:	72°C	2 minutes

Pre-lab Preparation

Aliquot 25 µL of PV92B primer/loading dye mix per student. The primer/loading dye mix may collect in the tube cap during shipping; pool the reagent by spinning the tube briefly in a microcentrifuge or by sharply tapping the tube bottom on the lab bench.

Pre-lab Set Up for DNA Amplification (per student station)

Cheek cell DNA. 2.5 µL (from Part I)
PV92B primer/loading dye mix, 25 µL
Ready-To-Go™ PCR beads (in 0.2-mL or 0.5-mL PCR tube)
Permanent marker
Micropipet and tips (1–100 µL)
Microcentrifuge tube rack
Container with cracked or crushed ice

Shared Items

Mineral oil, 5 mL (depending on thermal cycler)
Thermal cycler

III. ANALYZE AMPLIFIED DNA BY GEL ELECTROPHORESIS

The cresol red and sucrose in the primer mix function as loading dye, so that amplified samples can be loaded directly into an agarose gel. This is a nice time saver. However, since it has relatively little sugar and cresol red, this loading dye is more difficult to use than typical loading dyes. So, encourage students to load carefully.

Plasmid pBR322 digested with the restriction endonuclease *Bst*NI is an inexpensive marker and produces fragments that are useful as size markers in this experiment. The size of the DNA fragments in the marker are 1,857 bp, 1,058 bp, 929 bp, 383 bp, and 121 bp. Use 20 µL of a 0.075 µg/µL stock solution of this DNA ladder per gel. Other markers or a 100-bp ladder may be substituted.

View and photograph gels as soon as possible after appropriate staining/destaining. Over time, the small-sized PCR products will diffuse through the gel and lose sharpness. Refrigeration will slow diffusion somewhat, but for best results view and photograph gels as soon as staining/destaining is complete.

Pre-lab Preparation

Prepare a 1× concentration of TBE by adding 75 mL of 20× concentrated stock into 1,425 mL of deionized or distilled water. Mix thoroughly.

Prepare a 1.5% agarose solution by adding 1.5 g of agarose to 100 mL of 1× TBE in a 500-mL flask or beaker. Heat the flask or beaker in a boiling water bath (approximately 15 minutes) or in a microwave oven (approximately 4 minutes) until the agarose is completely dissolved. You should no longer see agarose particles floating in solution when the beaker is swirled. Allow the agarose to cool to approximately 60°C,

and hold at this temperature in a hot water bath. Cover beaker or flask with aluminum foil, and skim any polymerized "skin" off the top of the solution before pouring.

Pre-lab Set Up for Gel Analysis (per student station)

Amplified human DNA PCR products from Part III (store on ice)
Container with cracked or crushed ice

Shared Items

pBR322/*Bst*NI markers, 20 µL per row of gel (thaw and store on ice)
1.5% agarose in 1× TBE (hold at 60°C), 50 mL per gel
1× TBE buffer, 300 mL per gel
Ethidium bromide (1 µg/mL), 250 mL
or
*Carolina*BLU™ Gel & Buffer Stain, 7 mL
*Carolina*BLU™ Final Stain, 250 mL
Micropipet and tips (1–100 µL)
Microcentrifuge tube rack
Gel electrophoresis chambers
Power supplies
Water bath for agarose solution (60°C)
Latex gloves
Staining tray
Transilluminator with digital or instant camera (optional)

*Carolina*BLU™ STAINING

POST-STAINING

1. Cover the electrophoresed gel with the *Carolina*BLU™ Final Stain and let sit for 20–30 minutes. Agitate gently (optional).
2. After staining, pour the stain back into the bottle for future use. (The stain can be used 6–8 times.)
3. Cover the gel with deionized or distilled water to destain. Chloride ions in tap water can partially remove the stain from the DNA bands and will cause the staining to fade.
4. Change the water 3 or 4 times over the course of 30–40 minutes. Agitate the gel occasionally.
5. Bands that are not immediately present will become more apparent with time and will reach their maximum visibility if the gel is left to destain overnight in just enough water to cover the gel. Gels left overnight in a large volume of water may destain too much.

PRE-STAINING

*Carolina*BLU™ can also be used to stain the DNA while it is being electrophoresed. Pre-staining will allow students to visualize their results prior to the end of the gel run. However, post-staining is still required for optimum viewing.

To pre-stain the gel during electrophoresis, add *Carolina*BLU™ Gel and Buffer Stain in the amounts indicated in the table below. Note that the amount of stain added is dependent upon the voltage used for electrophoresis. *Do not use more stain than recommended. This may precipitate the DNA in the wells and create artifact bands.*

Gels containing *Carolina*BLU™ may be prepared one day ahead of the lab day, if necessary. However, gels stored longer tend to fade and lose their ability to stain DNA bands during electrophoresis.

Use the table below to add the appropriate volume of *Carolina*BLU™ stain to the agarose gel:

Voltage	Agarose Volume	Stain Volume
<50 Volts	30 mL	40 µL (1 drop)
	200 mL	240 µL (6 drops)
	400 mL	520 µL (13 drops)
>50 Volts	50 mL	80 µL (2 drops)
	300 mL	480 µL (12 drops)
	400 mL	640 µL (16 drops)

Use the table below to add the appropriate volume of *Carolina*BLU™ stain to 1× TBE buffer:

Voltage	Agarose Volume	Stain Volume
<50 Volts	500 mL	480 µL (12 drops)
	3000 mL	3 mL (72 drops)
>50 Volts	500 mL	960 µL (24 drops)
	2600 mL	5 mL (125 drops)

BIOINFORMATICS

Have students do the bioinformatics exercises before starting the experiment—or analyzing results. This should improve conceptual and practical understanding.

The onscreen Bio-i Guide can be played from the included CD-ROM or from the Internet site http://bioinformatics.dnalc.org/pv92/. The default version (640 x 480 pixels) allows one to follow along with an open browser window. The full screen version (1024 x 768 pixels) is best for demonstrations.

ANSWERS TO BIOINFORMATICS QUESTIONS

I.2. a. Matches of different lengths are coded by color. What do you notice? **There is only one complete match to the forward and reverse primers, followed by a number of partial matches.**

I.3. b. Note the names of any *significant alignments* that have *E-values* less than 0.1. Do they make sense? **There is only one hit with an *E-value* of less than 0.1. It makes sense, because it is from human Chromosome 16.**

I.3. d. The lowest and highest nucleotide positions in the subject sequence indicate the borders of the amplified sequence. Subtracting one from the other gives the difference between the two coordinates. **57137 – 56722 = 415.**

I.3. e. However, the actual length of the fragment includes both ends, so add 1 nucleotide to the result to determine the exact length of the PCR product amplified by the two primers. **416 nucleotides.**

I.3. f. Is this the + or the – allele? **There is not enough information to tell yet.**

II.4. What do you notice about the *E-values* obtained by this search? Why is this so? **Three hits have extremely low *E-values* (have many decimal places). This is because the query sequence is longer.**

II.5. Why does the first hit have an *E-value* of 0? **This hit completely matches the query, because it is the same Chromosome 16 clone identified in Part I.**

II.6. a. What do you notice about the 3′ end of the *Alu* repeat? **There is a poly-A tail composed of a string of 28 As (adenines).**

II.6. b. What appears to be going on? **The target sequence gaaagaa is duplicated during the insertion of the *Alu* element.**

II.7. What is the length of the *Alu* inserted at PV92? **The PV92 *Alu* is 308 bp long.**

II.8. If you assume that the amplicon in Part I is the – allele, what is the length of the + allele? **The + allele would appear to be the sum of 416 bp + 308 bp = 724 bp. However, the + allele also includes the 7-bp duplication of the target sequence. So the actual length of the + allele is 731 bp.**

II.9. Now look carefully at the third low *E-value* hit. Examine the *Features* and follow links. What is going on here? How are the three hits related to one another? **There are annotations for *Alu*s belonging to two different subfamilies: Ya5 is the older group that includes PV92, and Yb8 is a younger group. The younger *Alu* jumped inside the original *Alu* at the PV92 locus. One can easily see two poly-A tails in the sequence – one belonging to each *Alu*. This *Alu* within an *Alu* allele is rare and inserted so recently that it has only been found in a few people, notably from the Basque region of Spain and northern Morocco.**

III.4. On what chromosome have you landed? **Chromosome 16.**

III.7. What can you say about the gene that contains the amplicon? Click on the name in the *Genes_seq* track, then follow links to find out. **The amplicon lies within the cadherin H 13 (CDH13) gene. This gene produces a cell adhesion protein that mediates interactions between cells in the heart.**

III.8. a. Determine the size of the CDH13 gene using the map coordinates to the left of the contig map. **CDH13 is approximately 1.2 million nucleotides in length.**

III.8. b. How many introns and exons does CDH13 gene have? **CDH13 has 13 exons and 12 introns.**

III.8. c. Where in the CDH13 gene is PV92 *Alu* inserted: an exon or intron? **PV92 *Alu* is inserted within the 2nd intron.**

III.8. d. How does this explain the fact that the PV92 insertion is believed to be neutral—having no phenotypic effect? **Mutations within introns generally have no phenotypic effect.**

ANSWERS TO DISCUSSION QUESTIONS

Instructions on how to set up your class data are on the website and the CD-ROM.

2. An *Alu* insertion has only two states: + and –. How does this relate to information stored in digital form by a computer? What equivalent in digital information is provided by an *Alu* genotype? ***Alu* +/– is equivalent to a digital 0/1, or one bit of information. An *Alu* genotype (+/+, –/–, or +/–) contains two bits of information.**

4. Is the + allele confined to any particular racial or ethnic group? What can you say about people in the class who have at least one + allele? **The + allele is not exclusive to any racial or ethnic group. All people who have at least one + allele inherited their allele(s) from a common ancestor.**

5. b. How do genotype frequencies you observed in your experiment compare with those expected by the Hardy-Weinberg equation? Would you say they are very similar or very different? **Observed genotype frequencies typically are quite similar to the expected frequencies.**

7. h. Is your p-value greater or less than the 0.05 cut-off? What does this mean? **Class results typically have p-values greater than 0.05. This means that there is not a significant difference between observed and expected frequencies and that the observed frequencies are consistent with Hardy-Weinberg equilibrium.**

7. i. What conditions are required for a population to come into genetic equilibrium? Does your class satisfy these requirements? **Genetic equilibrium requires a relatively large population, no migration in or out of the group, no new mutations at the locus under study, and random mating in relation to the locus. While the class itself would be a very small population, its members are more or less representative of a larger population in your town or region. There is probably a relatively small amount of migration in and out of your town or region. There is no evidence of very recent, new mutations at the PV92 locus that would influence genotypes. There is no way of telling a person's PV92 genotype by looking at them, so people mate randomly in relation to this polymorphism. So, perhaps surprisingly, the class may generally fulfill the requirements for Hardy-Weinberg equilibrium.**

8. i. Which groups have significantly different genotype frequencies? What is the most frequent genotype in each group? **European, African, Australian, and American populations typically have similar genotype frequencies, with the –/– genotype being most common. The +/+ genotype is most common in Asian populations.**

9. g. Do you notice any pattern in the allele frequencies? **The + allele frequency is high in all Asian groups (up to 90%) and generally decreases moving westward through the Middle East, with European and African populations having frequencies of 10-35%. High + allele frequencies are also found in American Indian populations: Yanamamo (96%) and Maya (70%).**

PV92 "+" Allele Frequencies

Kung	20
African American	20
Alaska Native	29
Australia Aborigine	15
Breton	27
Cajun	21
Chinese	86
Euro-American	18
Filipino	80
French	23
German	10
Greek, Cyprus	18
Hispanic American	51
Hungarian	12
India Christian	48
India Hindu	52
Indian Muslim	30
Italian	24
Java	84
Malay	72
Maya	70
Moluccas	69
Mvsoke	53
Nguni	24
Nigerian	9
Pakistani	30
Papua New Guinea	14
Papua New Guinea (Coastal)	19
Pushtoon	33
Pygmy (Central Africa)	26
Pygmy (Zaire, now Congo)	35
Sardinian (Aritzo)	17
Sardinian (Marrubiu)	00
Sardinian (Ollolai)	00
Sardinian (San Teodoro)	27
Sotho	29
South India	56
Swiss	20
Syrian	18
Taiwanese	90
Turkish, Cyprus	58
United Arab Emirates	30
Yanomamo	96

9. h. Suggest a hypothesis about the origin and dispersal of the *Alu* allele that accounts for your observation. **Most students conclude that this pattern is consistent with the PV92 *Alu* insertion arising in Asia and then being diluted by gene flow to the west. Well-studied students, especially after doing exercise 10 (below), may understand that the pattern could also be the product of migration and genetic drift.**

9. i. Calculations suggest that the PV92 insertion occurred about 200,000 years ago. If this is so, in what sort of hominid did the jump occur, and what implications does this have for your hypothesis from h. above? **The PV92 insertion would have occurred in a population of *Homo erectus*, which then survived to give rise to modern humans (us). If this jump occurred in Asia, then *Homo erectus* must have survived in Asia to give rise to modern populations there. This would be consistent with the regional development hypothesis. The accepted replacement hypothesis—also called "Out of Africa"—supports the PV92 insertion occurring in a *Homo erectus* population in Africa. The worldwide frequencies of approximately 20% suggest that the + allele drifted to approximately this frequency in Africa prior to the migrations that gave rise to European, Asian, and Australian populations. The frequency then drifted much higher among the migrants that founded Asian populations, several of which may have carried a high + allele frequency when they migrated across the Bering Strait to found American Indian populations.**

10. d. How did hominids live 200,000 years ago, and what size population group would be supported? **Our hominid ancestors existed only by hunting and gathering, so that would limit the size of each group to around 50 individuals.**

10. e. What would be the allele frequency if a new *Alu* jump occurred in a group of this size? **For example, 50 people in the hunter-gatherer group would have 100 alleles, with one having the new *Alu* insertion – for an allele frequency of 1%.**

10. m. What happens to the new *Alu* insertion in the 100 populations? **The + allele frequency decreases from 1 percent to 0 percent in most of the populations within about 10 generations. The new *Alu* mutation is lost from these populations. Typically, the + allele is maintained in several populations at the end of 100 generations. Occasionally, the + allele will be fixed in a population, when the frequency rises to 100%.**

10. n. Follow the allele frequency in one population over 100 generations. What happens to the allele frequency, and what causes this? **The + allele frequency changes dramatically within one population. This random fluctuation in allele frequency is termed genetic drift.**

11. i. What do you notice about the allele frequency in those populations that maintain the + allele over 200 generations? **The + allele frequency drifts during the first 100 generations, but stabilizes in the expanded population. The larger population is nearing Hardy-Weinberg equilibrium.**

CD-ROM CONTENTS

The valuable companion CD-ROM is for exclusive use of purchasers of this DNA Learning Center Kit. To accommodate home or computer lab use by students, all materials may also be reached at the companion Internet site http://bioinformatics.dnalc.org/pv92/.

- **Protocol:** a unique online lab notebook with the complete experiment, as well as printable PDF files.
- **Resources:** 13 animations on key techniques of molecular genetics and genomic biology, from the award-winning Internet site, *DNA Interactive*.

Carolina Biological Supply Company
2700 York Road, Burlington, North Carolina 27215
Phone: 800.334.5551 • Fax: 800.222.7112
Technical Support: 800.227.1150 • www.carolina.com
CB272440608