Tested Studies for Laboratory Teaching
Proceedings of the Association for Biology Laboratory Education
Vol. 33, 8-30, 2012

Microsatellite DNA: Population Genetics and Forensic Applications

Kuei-Chiu Chen

Cornell University, Neurobiology and Behavior, 1130 Comstock Hall, Ithaca NY 14853 USA
(**kc373@cornell.edu**)

Microsatellite DNA has been used as standard genetic markers for human identification in forensic laboratories. This multi-week lab module uses professional lab techniques to cover concepts in molecular biology and mathematics. Students who have given consent extract DNA from their cheek cells and conduct a multiplex polymerase chain reaction that amplifies ten microsatellite loci simultaneously. The PCR products are analyzed using a standard DNA analyzer and then further processed with Peak Scanner (free software). A full genetic profile can achieve a random match probability as low as one in a trillion. The analysis of a genetic profile can be further expanded to parentage determination.

Keywords: microsatellite, genetic marker, population genetics, Hardy-Weinberg equilibrium, short tandem repeats, paternity index, forensics, multiplex PCR

Introduction

The purpose of this multi-week lab sequence is to introduce to students microsatellite DNA and to apply these types of genetic markers to human identification. The microsatellites being used are the loci routinely analyzed in professional forensic laboratories in the US and abroad (Butler, 2005, 2011). The materials and processing involved minimize the costs compared to submitting human tissue samples to private DNA labs which should make this lab more feasible as long as the instructors have access to a thermal cycler and are willing to submit PCR products to a DNA analyzing center on or off campus. This lab involves students extracting their DNA for genetic profile analysis. Because each sample is coded and only the student has knowledge of his/her sample code, it is not generally necessary to receive Institutional Review Board (IRB) approval. If, however, the instructors are interested in presenting student data collected in the lab, IRB approval is normally required.

The collection of the activities described here is taught in three consecutive weeks, each three hours long, with the last week being slightly shorter. The suggested activities and time budget not including introduction or discussion are described in Table 1.

The main objectives for this lab are as follows:

1. To teach students what genetic markers are so that they are able to apply the knowledge of specific genetic markers for individual identification.
2. To use the construction of standard curves based on known DNA size standards to estimate the sizes of unknown DNA molecules.
3. To apply the principle of Hardy-Weinberg equilibrium to the calculation of allele and genotypic frequencies.
4. To apply the multiplication rule to genotypic frequencies of unlinked loci and estimate the probability of a specific combination of genotypes.
5. To apply the chi-square test to determine if a sample follows the prediction of Hardy-Weinberg equilibrium.
6. To use Mendel's Law of Segregation and Law of Independent Assortment to determine a parent-child relationship given the genetic profiles of the individuals involved.

 © 2012 by Kuei-Chiu Chen

Table 1. Sequence of activities and time budget for the human microsatellite lab module.

Week	Exercise	Approximate Time (minutes)
Week 1	Practice pipeting	15
	Student DNA extraction	60
	DNA loading for multiplex PCR	20
Week 2	Electrophoresis of student DNA PCR products	20
	Standard curve construction and DNA size estimation	40
	Discuss and calculate random match probability(RMP)	40
Week 3	Test of Hardy-Weinberg equilibrium using student data	45
	Parentage determination	30
	Family tree of Tutankhamun	20

Student Outline

Background

A genetic marker is a silent DNA variation associated with a particular gene or trait. These markers can be used to locate a gene of interest because the markers are physically linked to that gene. At the same time the genetic markers themselves may be the focus of study for variation between individuals in a population. The purpose of this lab is to use 10 established genetic markers to understand variation between individuals and provide a quantitative measure of individual identification. To reach that goal, you will also learn some of the most commonly used molecular techniques in this lab.

Questions To Prepare You For This Laboratory

Week One

1. What is a genetic marker?
2. What are the essential molecular components of the polymerase chain reaction? What is a temperature profile and what is the purpose of each temperature of this technique? What is the relationship between the number of temperature cycles and the final number of copies of PCR products?
3. What is multiplex PCR?
4. When choosing genetic markers to conduct individual identification, why is it important to choose markers located on separate chromosomes?

Week Two

1. What physical feature of DNA molecules determines the direction they will move in an electrical field?
2. What is the relationship between DNA fragment size and migration distance in an electrophoresis gel?
3. Name one advantage and one disadvantage in using ethidium bromide for staining DNA bands on an electrophoresis gel?
4. What is random match probability in a forensic analysis? What is the relationship between random match probability and genetic profile frequency?
5. In deducing the genotype of each of the two DNA donors in a mixture sample, what do we assume about the amount of final PCR product and the amount of DNA template added in the PCR?

Week Three

1. Given the genotypes of microsatellite DNA alleles of all individuals in a population, how do you calculate the frequency of each allele of a specific locus?
2. What are the two components that determine the paternity index?

Introduction – Week One

Genetic Markers Used in Human Identification

The Human genome, like the genome of other eukaryotic species, is characterized by the presence of repetitive DNA sequences, many of which have no known functions (Li, 1997; Goldstein & Schloitterer, 2001). These repetitive sequences may be short with only a few repeats or long with millions of repeats. When the genome of an eukaryote is analyzed based on its density gradient using centrifugation, these repetitive sequences often appear as distinctive bands known as satellite DNA, separate from other, non-repeat or single copy DNA. One type of satellite DNA is known as microsatellites, in which the repeat unit ranges from 1 to 6 base pairs of nucleotides. In the forensic community the microsatellites used for human identification are also called short tandem repeats, or STRs.

An example of a microsatellite sequence is shown below (the complementary strand is not shown):

----GCTTGCC**AATGAATGAATGAATGAATGAATGAATG**CATTAAG----

Notice that the example above has AATG repeated eight times shown in box and the repeat sequence is flanked by non-repeating sequences. The AATG itself if not repeating will not be considered as microsatellite; the entire repetitive sequence is a microsatellite DNA where AATG is its repeat unit.

The origin of microsatellites is not fully understood. Genetics studies have suggested that the predominant mechanism for variation in the number of repeat units in microsatellites involves slippage of base-pairing DNA strands during replication.

One of the microsatellites with this repeat motif of AATG in humans is known as the TPOX locus. This locus, found on chromosome 2, was so named because it is located in an intron of the thyroid peroxidase gene. It has been observed that the TPOX repeat number ranges from 4 to 16 in the human populations. Naming the various forms of this locus is straightforward, simply based on the number of repeats, and each type is considered an **allele** for the TPOX **locus**. If an allele is repeated eight times, it is allele 8. If an allele has 9 complete units and a partial repeat of 2 nucleotides, this allele will be designated as 9.2. The **genotype** of a person's microsatellite locus may be **homozygous** thus carrying the same number of repeats in both of his/her homologous chromosomes, for example 8, 8. A person **heterozygous** at this locus will show a different number of repeats on his/her homologous chromosomes, for example 8, 11, or 7, 9.3, etc.

Because of the broad range of variation, and the observation that they do not outwardly express any phenotype, many STRs have been used as genetic markers for the purpose of human identification. STR markers used for forensic purposes typically are those with tetra-nucleotide repeats. In the U.S., 13 loci are routinely used for human identification in forensic laboratories.

In this lab you will see the allelic pattern of 10 STR markers amplified from your own DNA. If all 10 loci show a clear pattern, the probability of a random match can be as low as 1 in a trillion.

Nomenclature for STR marker loci

The STR markers are named based on their chromosomal location or derived from the name of the gene with which they are associated. In the example of TPOX, it is found inside the human thyroid peroxidase gene as described above. The STR markers not associated with genes are designated by their chromosomal position. A marker named D3S1358 stands for DNA, located on chromosome number 3, single copy (not to be confused with the "single copy DNA" above; here it means "not found in another chromosomal location"), and it is the 1358th locus described on this chromosome.

In order to know your genetic profile using the designated loci shown in Table 2, DNA extracted from your cheek cells will be subjected to the **polymerase chain reaction** (PCR), a procedure similar to DNA replication that is conducted in a cell-free condition.

To conduct a PCR we will need the following molecular constituents: 1) the original DNA as template, 2) a pair of short, single stranded DNA primers; the sequences of the primers are complementary to a portion of the flanking sequences of the targeted locus, 3) DNA polymerase to conduct the synthesis reaction, and 4) four deoxyribonucleotides dATP, dTTP, dCTP, and dGTP as the raw materials for synthesis of the new DNA strands. A complete PCR product should be double-stranded DNA that includes the forward and reverse primer sequences and the flanking sequences at both sides of the repeat sequences (Figure 1).

In this study we will use the primer sequences published by Promega® Corporation shown in Table 2. A fluorescent dye molecule is covalently attached to one of each pair of primers when they are commercially synthesized (see DNA Detection Method below).

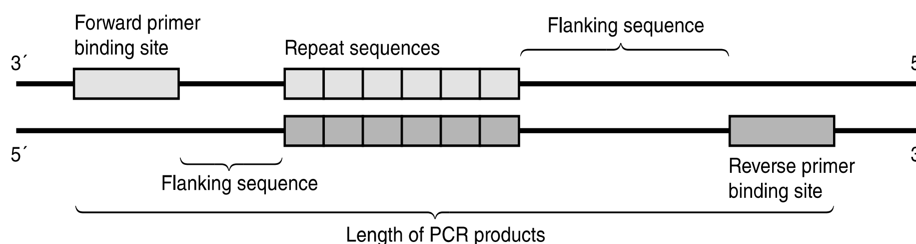

Figure 1. The composition of a PCR product from one microsatellite marker.

A PCR procedure involving amplification of multiple loci simultaneously in one reaction is known as **multiplexing**. In a multiplex reaction, multiple pairs of primers, each complementary to the flanking sequences of the target locus, are added in the reaction mixture. The remaining constituents of the reaction are the same as a typical PCR on single locus. Each of the primers has its unique nucleotide sequence so it can find its target, complementary sequence in the genome. Because multiple reactions have now been reduced to one, multiplexing saves time and costs and has been the preferred method in forensic practice in the U.S. and many other countries. One of the major challenges in multiplexing is the differentiation of products from separate loci that have the same length. This will be discussed in the DNA Detection Method section.

When designing a primer we need to include several important considerations. First, the ideal length of a primer is usually about 17 to 35 bases long to start the DNA synthesis. Second, because there are two hydrogen bonds between the pairing of adenine and thymine and three hydrogen bonds between guanine and cytosine, the composition of the appropriate number of nucleotide types is critical to provide adequate stability of the hybrid. A good primer normally requires 50-60% of cytosine and guanine in the sequence. In addition, the sequence of the primer should avoid being complementary so as to prevent formation of self-dimers. As shown in the example below, if a substantial number of bases is complementary to each other within a primer, primer molecules will preferentially bind to one another and will not bind to the template sequence, so the PCR will not be successful.

```
          5’GGGAAAATTCCAGGATCTAT 3’
              ¦ ¦ ¦ ¦   ¦ ¦ ¦ ¦
3’TATCTAGGACCTTAAAAGGG 5’
```

Furthermore, a structure commonly known as a hairpin may form if portions of the primer sequence are complementary to each other. In the example below, the end of the molecule folds over and forms base-pairing with another part of the primer.

```
        5’   GGAAA ⌒
               ¦ ¦ ¦
3’   TATCTAGGACCTTA ⌒
```

Table 2. Primer sequences of 10 STR markers used in this lab’s protocol. Notice that one of each primer sequence pair is labeled with a fluorescent dye indicated in brackets.

Locus	Primers
TPOX	5’-[Yellow]-GCACAGAACAGGCACTTAGG-3’ 5’-CGCTCAAACGTGAGGTTG-3’
D3S1358	5’-[Blue]-ATGAAATCAACAGAGGCTTGC-3’ 5’-ACTGCAGTCCAATCTGGGT-3’
FGA	5’-[Yellow]-GGCTGCAGGGCATAACATTA-3’ 5’-ATTCTATGACTTTGCGCTTCAGGA-3’
CSF1PO	5’-CCGGAGGTAAAGGTGTCTTAAAGT-3’ 5’-[Blue]-ATTTCCTGTGTCAGACCCTGTT-3’
D7S820	5’-[Green]-ATGTTGGTCAGGCTGACTATG-3’ 5’-GATTCCACATTTATCCTCATTGAC-3’
D8S1179	5’-ACCAAATTGTGTTCATGAGTATAGTTTC-3’ 5’-[Blue]-ATTGCAACTTATATGTATTTTTGTATTTCATG-3
TH01	5’-[Yellow]-GTGATTCCCATTGGCCTGTTC-3’ 5’-ATTCCTGTGGGCTGAAAAGCTC-3’
VWA	5’-[Red]-GCCCTAGTGGATGATAAGAATAATCAGTATGTG-3’ 5’-GGACAGATGATAAATACATAGGATGGATGG-3’
D13S317	5’-[Green]-ATTACAGAAGTCTGGGATGTGGAGGA-3’ 5’-GGCAGCCCAAAAAGACAGA-3’
D16S539	5’-[Green]-GGGGGTCTAAGAGCTTGTAAAAAG-3’ 5’-GTTTGTGTGTGCATCTGTAAGCATGTATC-3’

Choose one primer from Table 2. How many nucleotides are there in this primer? Calculate the G+C ratio. Is it within 50-60% of the length of the sequence?

It is important to ensure the working condition of PCR ingredients by including a template sample known to generate products. This is known as the positive control. It is equally crucial to include a negative control, in which no template DNA is added to the reaction, to detect any contamination.

One of the common challenges in PCR is proper adjustment of template DNA concentration in relation to other ingredients in the reaction. In your DNA purification process you will use resin beads to attempt to overcome this hurdle. You will use a

fixed amount of resin beads that provides a constant total surface area. You will then add enough cheek cells to supply a sufficient amount of template DNA to saturate the surfaces of the resin beads. The excess amount of DNA will not adsorb to the beads and will be washed away. When the DNA is released from the resin beads using an appropriate solution, its yield should fall within the range of ideal concentrations for use in the PCR process.

DNA Detection Methods

To visualize the PCR product, a molecule that produces a fluorescent color when excited by laser radiation is covalently attached to the 5' end of one of the primer pairs when the primers are chemically synthesized (see Table 2). Different color dyes are incorporated in primers for STR markers whose PCR products overlap in size range. This allows differentiation of alleles from different loci that are similar in length. The PCR products are then analyzed using an instrument with capillary tubes similar to those used in DNA sequencing analysis. This capillary analysis uses a mechanism similar to the gel electrophoresis you will conduct in week 2 of this lab, but is much more sensitive to the sizes of the DNA fragments. The instrument can determine the number of base pairs if size standards are provided.

The completed analysis can be demonstrated in an **electropherogram** (or **chromatogram**) in which the color of signal peaks reflects the dye label attached to the primers. In an electropherogram the horizontal axis indicates the size (length) of the DNA product in base pairs, and the height of the peaks indicates the relative amount of this molecule. Examples of a genotyping electropherogram will be provided in the lab and a partial electropherogram is shown in Figure 2.

Interpretation of Microsatellite Profiles

As described earlier, the STR loci chosen for this study are markers with four base-pair repeat units; each locus has a specific primer sequence flanking the target sequence. Although a repeat unit may have the same sequence in the repeat unit, each of these loci is located at a different chromosomal location and is considered a different marker. Further details on how to interpret an STR profile can be found in the example in Table 3.

Table 3. Example of a hypothetical STR profile.

Locus	Genotype
TPOX	8, 11
D3S1358	14, 15
FGA	20, 22
CSF1PO	11, 11
D7S820	9, 10
D8S1179	13, 13
TH01	9, 9.3
VWA	15, 16
D13S317	12, 13
D16S539	11, 12

In the example shown in Table 3 the allele 12 in D13S317 and allele 12 in D16S539 are different alleles from different markers even though they appear as the same number. In CSF1PO allele 11 is recorded twice. This allele showed as a single peak on the hypothetical electropherogram (see Figure 2 below). The most common interpretation for a single allelic peak is that the individual from whom the DNA was obtained is homozygous at this locus. The locus TH01 has the allele 9.3, which indicates that there is a partial repeat of three base pairs in addition to nine complete repeats.

Population Genetics Using Microsatellite Markers

One of the advantages of using microsatellite DNA for human identification is its power of discrimination. By using many highly polymorphic loci, it is extremely unlikely that two individuals will share the entire suite of genotypes except in the case of identical twins. The match of genotype at each locus is determined based on Hardy-Weinberg equilibrium. Details on how to mathematically derive the allele and genotypic frequencies of a population are available in the assigned pages in the textbook (Sadava *et al*., 2011, p. 444-448). The following are abbreviated descriptions of this principle.

In short, under Hardy-Weinberg equilibrium if a particular locus has only two alleles, *A* and *B,* and their respective frequencies are *p*, and *q*, then

$$0 < p < 1; \ 0 < q < 1; \text{ and } p + q = 1$$

and the frequency of the genotype *AA* in the population will be p^2, the genotypic frequency of *BB* will be q^2 and the frequency of the heterozygote *AB* will be *2pq*. This relationship can be described in the mathematical equation shown below:

$$(p + q)^2 = p^2 + 2\,pq + q^2 = 1$$

When there are more than two alleles for each gene (locus), such as n alleles, each noted as A_1, A_2... and A_n, the Hardy-Weinberg equilibrium still applies as shown in the following expanded equation:

$$(p_1 + p_2 + p_3 + \ldots\ldots + p_n)^2$$
$$= p_1^2 + p_2^2 + \ldots + p_n^2 + 2p_1p_2 + 2p_1p_3 + \ldots + 2p_1p_n + 2p_2p_3 + \ldots + 2p_2p_n + \ldots + 2p_n\text{-}1p_n$$
$$= 1$$

where the frequency of allele A_i is represented as p_i. In the expanded equation, the frequency of each homozygous genotype is represented by the square of the frequency of the allele in question, and the frequency of a heterozygous genotype is the product of the two corresponding allelic frequencies multiplied by 2.

In forensic practices, it is customary to use the appropriate database of allele frequency for a specific ethnical group. These frequencies are obtained from published studies and you will choose appropriate frequencies for calculations in week two (Ashma & Kashyap, 2002; Budowle *et al*., 2002; Butler *et al*, 2003; Hashiyada, Itakura, & Nata, 2003; Kim *et al*., 2003; Wang *et al*., 2003). Table 4 lists allele frequencies of the Caucasian population in the U.S. (Butler *et al*, 2003) and is used in the examples below.

For the TPOX locus, because the frequency of allele 8 is 0.535 and the frequency of allele 11 is 0.243, the frequency of encountering a Caucasian who has the genotype of 8 and 11 on this locus will be 2 x 0.535 x 0.243 = 0.260.

In STR analysis multiple loci are often analyzed simultaneously. One important requirement is that the alleles assort independently, i.e., the occurrence of one genetic marker does not influence the occurrence of another. From a biological standpoint, if genetic markers are located on different chromosomes (i.e., unlinked), they will assort independently during meiosis. Ten the probability of encountering one genetic marker is not influenced by the probability of encountering another. If two markers assort independently, the frequency of encountering both of them simultaneously is the product of the frequency of each of the two markers. This is commonly known as the **product rule** or **multiplication rule**.

Let's consider the second locus D3S1358 in Table 4. Allele 14 has a frequency of 0.103 and allele 15 has a frequency of 0.262. The probability of the genotype 14, 15 in D3S1358 will be 0.054. Thus, to encounter a Caucasian who carries alleles 8 and 11 for TPOX locus AND alleles 14 and 15 for D3S1358 locus the probability will be 0.260 x 0.054 = 0.014.

Table 4 lists the formulae for obtaining the frequency for each of genotypes in Table 3.

From Table 4, the probability of encountering an individual genetic profile by random chance in a population based on the product rule will be about 3.737 x 10-11. This is known as the **random match probability**. The inverse of this probability equals 2.675 x 1010, and is known as the **genetic profile frequency**, an expression easier to comprehend by audiences when forensic biologists testify in court. When this frequency is expressed verbally, it is said that the chance of a random match is 1 in more than twenty-six billion.

Table 4. Genotypic frequencies of hypothetical genotypes of ten STR markers.

Locus	Genotype	Allele frequencies	Genotypic frequency formula	Genotypic frequency
TPOX	8, 11	0.535, 0.243	$2p_ip_j$	0.260
D3S1358	14, 15	0.103, 0.262	$2p_ip_j$	0.054
FGA	20, 22	0.127, 0.219	$2p_ip_j$	0.056
CSF1PO	11, 11	0.301, 0.301	p_i^2	0.091
D7S820	9, 10	0.177, 0.243	$2p_ip_j$	0.086
D8S1179	13, 13	0.305, 0.305	p_i^2	0.093
TH01	9, 9.3	0.114, 0.367	$2p_ip_j$	0.084
VWA	16, 17	0.248, 0.242	$2p_ip_j$	0.060
D13S317	12, 13	0.248, 0.124	$2p_ip_j$	0.062
D16S539	11, 12	0.321, 0.326	$2p_ip_j$	0.209

Introduction – Week Two

In week two of the microsatellite DNA lab you will practice loading a sample into the well of a practice agarose gel before loading the PCR product from the last lab into a precast gel. This gel allows you to visualize DNA migration in real time. To minimize logistic complications you may not be getting your own PCR product. This should not impact how you learn this DNA separation technique based on the size (number of base pairs) of the molecules. You will construct a standard curve based on known DNA fragment sizes and their migration distances. Using this curve you will estimate the sizes of the PCR fragments of your sample. You will receive your STR results and calculate the probability of finding another individual with the same genotypic profile as yours in the population. The last exercise of this week will be a forensic scenario in which two individuals left their DNA at the crime scene. The method to deduce the individual DNA donors will be discussed in lab.

Agarose Gel Electrophoresis

The essential features of electrophoresis are simple; an electrical field is established in a medium through which charged molecules can diffuse. Nucleic acids are acids because the phosphate groups of the sugar phosphate chains ionize (by loss of a H+ from the oxygen) when in solution. The resulting negatively charged phosphate ions give nucleic acids a net negative charge. As a result of their negative charge, nucleic acids migrate toward the positive pole in an electrical field.

Smaller molecules (fragments) move more rapidly through an electrophoresis gel than larger molecules and end up nearer the positive end of the gel. Molecules of the same size move at the same rate so electrophoresis separates populations of molecules of the same size into "bands" on the gel. Because DNA is colorless, you must add a dye, called a "loading dye" that will move as fast as the smallest fragments. To visualize DNA we need to add a stain that can be associated with DNA, either in the buffer solution during electrophoresis to allow DNA to pick up the stain, or to soak the gel in the staining solution after electrophoresis is complete. Ethidium bromide is a commonly used stain. This highly mutagenic substance binds to DNA by intercalating between the stacked bases of DNA, producing highly concentrated patches of the dye in the exact location of DNA bands on the gel. In your lab the gel is a commercially available precast gel containing a proprietary stain of similar properties as ethidium bromide. Because this stain is already in the gel and the gel run takes place on the transilluminator, we can visualize DNA migration in real time while the gel is being run.

Because smaller molecules move faster, you expect to see an inverse relationship between the distance moved from the sample well and the size of the DNA fragment. DNA size is usually expressed as the number of nucleotide base pairs. Specifically, if you plot the migration distance against the log of the number of base pairs, a linear relationship will result. To "calibrate" this curve, one typically includes a sample producing bands whose fragment sizes are known. These **size standards** then allow you to construct a standard curve defining the relationship between the migration distances of DNA fragments and the log of fragment size in base pairs for the electrophoresis system being used. The standard curve can then be used to estimate the base pair lengths of the fragments in any bands observed from the unknowns on the gel. Recall from last week you used one pair of primers to amplify each locus of microsatellite marker. Between the primers and the repeat sequences there are also flanking sequences, so the total length of the PCR products include the entire length of nucleotides comprised of primers, the repeat sequence and its flanking sequences (see Figure 2).

Interpreting a Chromatogram

After you have gained an understanding of how DNA sizes determine the distance of their migration in electrophoresis, you will expand that concept in the analysis of an electropherogram (or chromatogram), which essentially is a multicolor graphic result of an extremely fine-tuned electrophoresis. After presenting your unique sample code recorded in week one, you will receive your genetic profile in an electropherogram in which color peaks represent the color of the fluorescent labels on the primers which can indicate the possible loci. The strengths of signals, shown as the height of the peaks, indicate the amount of molecules with that color label. Each peak also indicates the base pairs of that molecule, which indicate the allele of that locus. There are four panels in your chromatogram and one of the panels is shown in Figure 2.

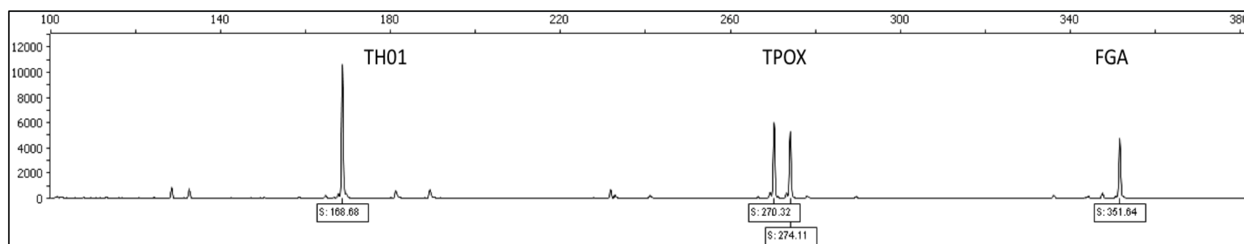

Figure 2. One of the four panels of a chromatogram from a microsatellite fragment analysis. The y-axis is the fluorescent unit as indication of quantity of DNA fragments carrying that specific color label; the x-axis is the size of fragments in base pairs shown at top of the panel. The name of each locus is shown next to the corresponding peak and the base pair size of each allele is shown below the peak in the box. In this example both TH01 and TPOX are heterozygous while FGA is homozygous.

Introduction – Week Three

In week three of this lab sequence you will start by revisiting the concept of Hardy-Weinberg equilibrium and conducting a χ^2 test on genotypic data from 150 students randomly sampled from your class. The purpose of this exercise is to test if the genotypic frequencies of students in your class follow the prediction of Hardy-Weinberg principle. You will also conduct parentage analysis where you may resolve paternity given the mother and the child's genotypes are known (Evett & Weir, 1998). Using published data (Hawaas *et al*., 2010), the last exercise of the lab will expand the concept of family trio (father-mother-child) in order to resolve four generations of kinships involving the Egyptian King Tutankhamun.

Test of Hardy-Weinberg Equilibrium Using Student Data

In this exercise you will use one microsatellite locus to determine if genotypic frequencies of students in this class are consistent with the prediction of Hardy-Weinberg equilibrium. You will count the number of each allele of this locus, TH01, and calculate the frequency (percentage) of each allele of this locus. Based on the prediction of Hardy-Weinberg equilibrium, there is an expected frequency of each genotype determined by the allele frequency. The comparison between the observed genotypic frequency in students and the predicted genotypic frequency based on Hardy-Weinberg principle can be assessed using a χ^2 test.

Parentage Analysis

In recent years microsatellite markers have been widely used to determine parentage in legal disputes. Because most of the parentage cases have the established mother-child relationship, the relationship in question is the paternity. So in a typical paternity case involves the trio of the mother, the child and the potential father (PF). Assuming the mother is indeed the biological mother of the child, there are two possible outcomes of the paternity tests, exclusion or inclusion of the potential father as the biological father of the child. The determination of this outcome is based on the obligate allele found in the child at each locus. The **obligate allele** is the allele of a locus of the child that is passed down from the father. If this allele is not found in the PF, he is considered excluded from consideration. An example of excluded paternity is shown in Table 5 below.

Table 5. An example of exclusion in paternity.

Locus	Child	Mother	Obligate allele	Potential Father	Inclusion?
TPOX	8, 11	8, 9	11	9, 13	No
D3S1358	14, 15	15, 16	14	9, 9	No

Only in the condition of inclusion will an assessment of likelihood be conducted. To determine the likelihood of the potential father being the biological father based on each microsatellite locus, we will consider the ratio of two competing possibilities:

$$PI = \frac{X}{Y}$$

in which, X is the probability for the potential father to pass the obligate allele to the child, and Y is the probability of a randomly selected man unrelated to the child to pass the obligate allele to the child, and PI is the **paternity index**. Paternity index is a type of **likelihood ratio** that weighs the strengths of two competing hypotheses, the higher the PI value, the stronger the implication that the PF is the father. To simplify, if the PF is homozygous for the obligate allele, X is equal to 1. If the PF is heterozygous, the probability of passing the allele to the child is 0.5. The probability of a randomly selected man unrelated to the child having that allele is the probability of that allele in the population of the ethnic group of the potential father. Using the first two loci mentioned in Table 3 and some of the same alleles, PI is demonstrated in an example in Table 6 below.

Table 6. An example of inclusion of paternity.

Locus	Child	Mother	Obligate allele	Potential Father	Inclusion?	PI
TPOX	8, 11	8, 9	11	11, 11	Yes	1/0.243 = 4.115
D3S1358	14, 15	15, 15	14	14, 16	Yes	0.5/0.103 = 4.854

In a more complicated situation where the child and the mother are both heterozygous and share both alleles of that locus, the paternity index is determined by a different denominator. In the probability rule where there are two mutually exclusive events, the probability for either one of the events to occur is the sum of the individual probabilities of the event. Because there is no knowledge of which of the two alleles was passed on from the mother and which was from the biological father, the denominator will include the probability of either one of the alleles. Table 7 describes this situation.

Table 7. Determination of paternity index when mother and child are in common with both heterozygous alleles.

Locus	Child	Mother	Obligate allele	Potential Father	Inclusion?	PI
TPOX	8, 11	8, 11	8 or 11	11, 11	Yes	1/(0.535+0.243)= 1/0.778 = 1.285
D3S1358	14, 15	14, 15	14 or 15	14, 16	Yes	0.5/(0.103+0.262) = 0.5/0.365 = 1.370

The above examples did not address all the possible allelic combinations in a paternity dispute trio. Further, more complex determination of PI will not be emphasized here.

Because the marker loci are unlinked, we can combine the PI's by multiplying the individual PI's and generate a **combined paternity index** (CPI). In Table 7, the CPI is calculated by multiplying 4.115 by 4.854, which is equal to 19.974. A CPI can potentially reach infinity if a very large number of loci are considered together. A common practice of acceptance of paternity is when the CPI is above 100.

Genealogy of Tutankhamun's Family

As the last activity of the microsatellite DNA lab sequence, this exercise allows you to use published microsatellite data (Hawass *et al*., 2010) to resolve seven members and four generations of relations in King Tutankhamun's family based on Mendel's Law of Segregation, Law of Independent Assortment, and the concept of parentage trio you explored in the previous exercise.

Nucleotides just like other molecules degrade over time due to exposure to physical, chemical, and biological conditions, so successful extractions of DNA and PCR amplifications of alleles from ancient tissues from subjects such these three-thousand-year old mummies were not as immediate as you experienced using your cheek cell sample. It took the authors two years before PCR products were amplified to large enough quantity so that the genotypes were resolvable. However, the fact that there were successes suggests that, because the amplified sequences were short, even degraded samples may produce reliable results.

Procedures – Week One

DNA Extraction and Purification

Although in real forensic analysis you would be extracting unknown samples related to criminal cases, to gain an appreciation of your unique genotypic profile, you will be extracting your own DNA from your cheek cells.

Many genetic studies require DNA isolation so that other cellular components do not interfere with the analysis. One commonly used component in the commercial DNA extraction kit is silica beads. DNA molecules bind to these resin beads while other cellular components are washed away. Eventually a proper solution is added to the beads to release DNA into the solution. You will extract your DNA using silica coated magnetic beads. The magnetism in the beads used in conjunction with a magnetic stand allows the DNA-covered beads to quickly pool at the bottom of the tube so that the solution can be removed readily without centrifugation. In addition, the extraction procedure includes three buffer systems. The **lysis buffer** contains chemicals to disrupt cell membranes, break down proteins associated with DNA in the chromosomes, and denature DNases that can cleave DNA. The released DNA will bind to the surfaces of the beads. The **wash buffer** contains high salt and ethanol to remove organic components such as lipids and proteins from DNA or RNA molecules while allowing DNA molecules to remain adsorbed to the beads. Finally the low ion concentration in the **elution buffer** allows the release of DNA molecules from the surfaces of the beads into the solution. The 65°C incubation further facilitates the release and increases the total yield of DNA.

DNA Extraction Using Liquid Saliva Sample

Deionized water should be used for students who choose not to extract their DNA. These samples may be used as a negative control to detect any contamination.

1. Label your initials on the cap of the microcentrifuge tube containing magnetic beads and lysis buffer.
2. Rub your tongue around your inner cheeks for 30 seconds then spit your saliva into the small cup.
3. Add 100 µl of fresh saliva to the tube containing resin beads and lysis buffer. Vortex 3 sec. per min. for a 5-min. incubation on the bench top.
4. Conduct a final vortexing, then immediately place the tube on the magnetic stand. Resin beads will settle at the bottom and side of tube. Your DNA has been released from cheek cells and is binding on the surface of the resin beads.
5. While the tube is on the stand, pipette out the liquid without disturbing the beads. Discard the liquid into the saliva cup.
6. Add 100 µl of Lysis Buffer (LB), vortex 2 sec. and return the tube to the magnetic stand.
7. Remove and discard all LB into the saliva cup.
8. Add 100 µl prepared Wash Buffer (WB); vortex 2 sec.
9. Return to the magnetic stand; pipette out and discard all WB.

   Repeat step 8 and step 9 two more times. Make sure all liquid has been removed in from the last wash.
10. Return the tube to the magnetic stand, lid open, and air dry for 10-15 min. or until the inside of the tube is free of wash buffer.
11. Add 100 µl of Elution Buffer (EB); close the lid.
12. Vortex the tube for 2 sec.; immediately put the tube in a 65ºC bath for 5 min.
13. Remove the tube from the water bath, vortex 2 sec. and immediately place in the magnetic stand.
14. Now the solution has your DNA so DO NOT DISCARD. Carefully transfer this solution to a new tube but DO NOT label it. Bring this new tube to your TA, who will instruct you on how to transfer your DNA to conduct a multiplex PCR.

Multiplex Polymerase Chain Reaction

After you have completed your DNA extraction, bring your sample tube to the instructor's desk. Two or three separate sample plates will be there labeled with your lab section code. On each of the plates is a strip of aluminum foil covering a row of wells containing all ingredients for PCR except the template DNA. By adding your DNA, the mixture is ready for PCR.

Having two or three separate plates reduces congestion.

1. Choose either one of the plates. Observe the foil – the punctured foil means DNA from another student has already been added so you should not add your DNA in that well.
2. Use the pipet at the TA desk to add 15 µl of your DNA into a fresh well by puncturing the foil. Pipette up and down a couple of times. Depress the plunger while removing the pipet from the well.
3. Record the well number where your DNA was added here: __________________. You will need to use it to retrieve your sample next time.

When the class is ready, your lab instructor will bring all the samples to the thermal cycler located in one of the prep rooms. The thermal cycler is programmed to run the temperature profile described in the Introduction-Week one, with a total run time of about 3 hours. This instrument has a capacity of running 96 samples simultaneously. To save time, samples from concurrent lab sections will be run together. The prep staff will submit the PCR results to the DNA analysis facility on campus. The results of your profile will be returned to you next lab.

Procedures – Week Two

Gel Electrophoresis

This week we will examine the PCR results using two different methods, the common gel electrophoresis and the chromatogram. In the gel electrophoresis exercise one volunteer from your lab section (or your TA) will pour an agarose gel on the gel electrophoresis apparatus. Students will let the gel solidify and practice loading samples on it.

Loading Sample on Lonza™ FlashGel

The precast gel material is sandwiched between a plastic casing with the buffer sealed inside. There is no gel material in the well thus no fear of puncturing it. Each student will receive a PCR product sample with loading dye in a microcentrifuge tube and the samples from the entire class will be loaded on one gel.

1. Set the micropipet to 5 µL and fill it with the PCR product. When it is your turn dispense the PCR product into the well. Record your well number here ___________.
2. When all samples have been loaded, your TA will place the gel cartridge on the transilluminator and insert the end with the electrodes to connect with the illuminator dock.
3. Plug in the power unit, select the "Hi" voltage setting, turn on the unit, and adjust the voltage to about 250 volts.
4. Allow electrophoresis to proceed for about 4-5 minutes. During this time observe the migration of DNA bands on the projected screen. The intercalator of DNA (similar to ethidium bromide) is in the buffer in the gel cartridge. When DNA is added, it picks up the intercalator and this stain is illuminated to show the position of the DNA bands.
5. Your TA will take a photo of the gel and print one copy for each student.

Using a Standard Curve to Estimate the Length of PCR Product

The size standards are manufactured DNA fragments with known base pair lengths. By knowing their base pairs and the migration distances, we can determine the relationship between the two by constructing a standard curve. After measuring the migration distance of a DNA fragment of an unknown size and locating its travel distance on the standard curve, we will be able to estimate its size in base pairs. Measure all but the largest and the smallest size standards.

Use a small plastic ruler to make measurements directly on the printout of your gel. For each lane, measure to the nearest 0.5 mm the distance from the front edge of the well to the front edge of each band. Enter band location data for each lane in Table 8 from smallest to largest measurement.

The best linear relationship between fragment size and migration distance is obtained by plotting on the distance migrate on a linear scale and the size of the fragment on a logarithmic scale.

Observe Figure 3, a semilog graph. Notice that the x-axis (migration distance) follows a linear scale while the y-axis (DNA fragment size) includes one complete cycle ranging from 100 to 1,000 base pairs.

Table 8. Fragment sizes in base pairs (bp) for size standard and their migration distance.

Fragment Size (bp)	Migration Distance (cm)
1500	- - - - - - -
800	
500	
300	
200	
150	
100	
50	- - - - - - -

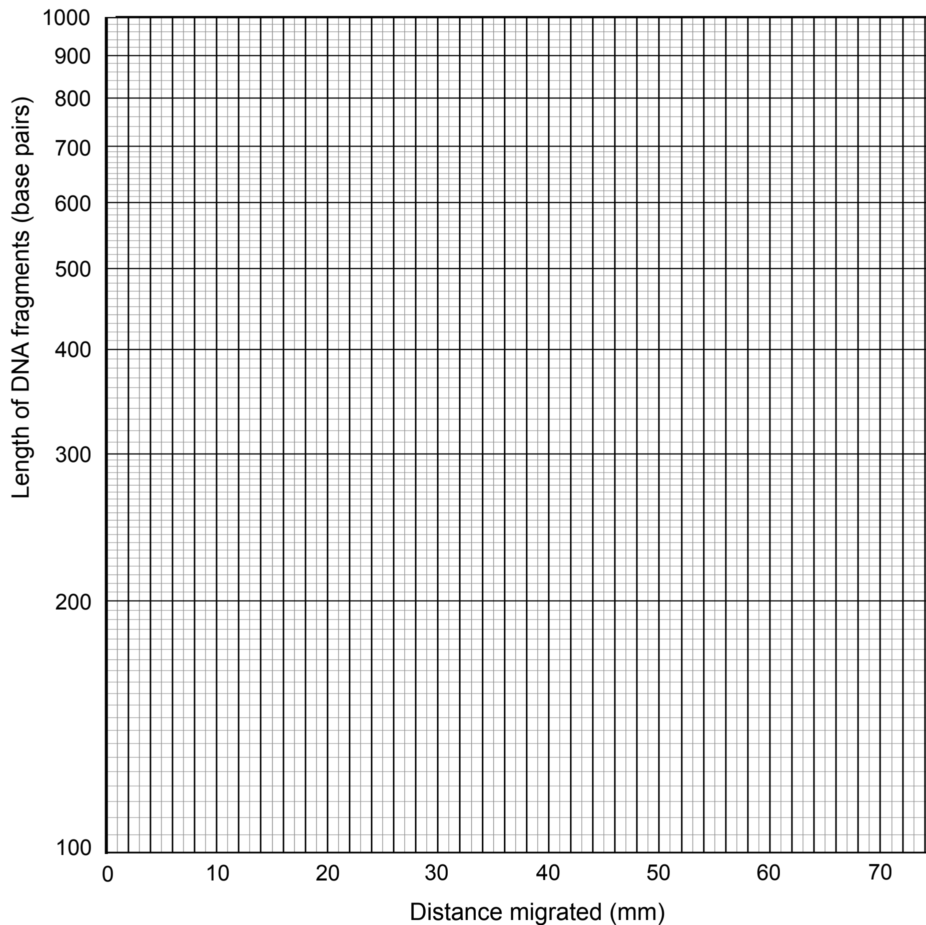

Figure 3. A standard curve showing the relationship between the migration distance of DNA fragments from PCR of microsatellites and the log of fragment sizes in base pairs.

Plot the data in Table 8 on the semilog graph in Figure 3. Place the edge of a ruler along the line that marks the size of a fragment and then plot the migration distance of that fragment.

Draw a straight line that best fits the data in Figure 3. There should be an approximate linear relationship between the log of fragment sizes and the migration distances.

Use the standard curve in Figure 3 to estimate the base pair size of each of the fragments in the lane where you loaded your sample and record your deduced fragment size and possible locus in Table 9.

Table 9. The travel distance and deduced fragment size of your assigned DNA band.

Student fragment	Distance traveled (mm)	Deduced fragment size (bp)	Likely locus
1			
2			
3			
4			

Estimation of Random Match Probability of Your DNA Type

1. Present your sample code to your lab instructor to retrieve your electropherogram. If you did not extract your DNA, an example will be provided for you.
2. Your TA will provide you with a table to convert fragment size to allele. Use this information to convert the size of each of your alleles and record your genotype in the Table 10 below.
3. Look up the allele frequencies of your ethnic group from your TA and record the allele frequencies in Table 10. Circle the genotypic frequency formula and calculate the genotypic frequency of each locus.

Table 10. Genotypes and genotypic frequencies from individual student.
Choose one table of your ethnic group and record appropriate allele frequencies in the table below.

Locus (in order of panel arrangement)	Genotype	Allele frequencies	Genotypic frequency formula (circle one)	Genotypic frequency
			$2p_ip_j$ or p_i^2	
			$2p_ip_j$ or p_i^2	
			$2p_ip_j$ or p_i^2	
			$2p_ip_j$ or p_i^2	
			$2p_ip_j$ or p_i^2	
			$2p_ip_j$ or p_i^2	
			$2p_ip_j$ or p_i^2	
			$2p_ip_j$ or p_i^2	
			$2p_ip_j$ or p_i^2	
			$2p_ip_j$ or p_i^2	

4. Multiply all genotypic frequencies to estimate the random match probability of your profile.

   What is your random match probability? ______________________________

5. Genetic profile frequency is the inverse of random match probability. Convert your random match probability to genetic profile frequency.

   What is your genetic profile frequency? ______________________________

6. In the following space verbally describe the likelihood of encountering your genetic profile in the population.

   ____

Procedure – Week Three

Test of Hardy-Weinberg Equilibrium From Student Data

In this exercise you will use genotypic data collected randomly from your class. To minimize the number of possible genotypes we have chosen the TH01 locus for its lower number of alleles among all markers we have analyzed. You will be supplied with data from 150 students (300 alleles total) for easier calculation. Your TA will provide you with data for Table 11. Complete Table 12 and 13 accordingly as your TA guides you with the procedure.

1. Your TA will provide the numbers of each genotype in the class sample. Record these numbers in Table 11. These are observed number of students (O).
2. Calculate the direct count of alleles and allele frequencies in Table 12. Each student homozygous for an allele will contribute two of that allele; a heterozygous student contributes one of each of the two alleles carried by this student to the total number of alleles in the sample.
3. Calculate each of the Expected genotypic frequencies for each genotype based on Hardy-Weinberg equilibrium and record it in Table 13. To do so, square the allele frequency if homozygous for that allele (p_i^2). If heterozygous, multiply the two corresponding frequencies by two ($2p_ip_j$).
4. Multiply Expected genotypic frequency by 150 to derive at the Expected number of students (E) and record it in the last column in Table 13.

You will use a χ^2 test to compare the observed genotypic frequencies and expected genotypic frequencies from your class in Table 13. The purpose of this analysis is to see if the genotypic frequencies of the student samples are consistent with the prediction from Hardy-Weinberg equilibrium. Read the procedures for conducting a χ^2 test. Since fractions are not suitable for χ^2 tests, you will compare the observed number of students and expected number of students of each genotype. As a convention, you should state your two competing hypotheses when conducting a statistical test.

H_o: Observed genotypic frequencies do not differ from the expected genotypic frequencies predicted from Hardy-Weinberg equilibrium.

H_a: Observed genotypic frequencies do differ from the expected genotypic frequencies predicted from Hardy-Weinberg equilibrium.

Table 11. Observed number of each of the genotypes in students at TH01 locus.

Genotype	Observed number (O)
6, 6	
6, 7	
6, 8	
6, 9	
6, 9.3	
7, 7	
7, 8	
7, 9	
7, 9.3	
8, 8	
8, 9	
8, 9.3	
9, 9	
9, 9.3	
9.3, 9.3	
Total	150

Table 12. Allele frequencies on TH01. Data for direct count of alleles are calculated from Table 11.

Allele	Direct count (DC) of alleles.	Allele Frequency (p_i =DC/300)
6		
7		
8		
9		
9.3		
Total	300	1.0

Table 13. Observed genotypic frequencies for locus TH01 from students in this class.
The expected genotypic frequencies are calculated based on Hardy-Weinberg equilibrium.

Genotype	Expected genotypic frequency (Use $p_i^{2,}$ or $2p_ip_j$ as formula)	Expected number of students (E) (= $p_i^{2,}$, or $2p_ip_j$ multiplied by 150)
6, 6		
6, 7		
6, 8		
6, 9		
6, 9.3		
7, 7		
7, 8		
7, 9		
7, 9.3		
8, 8		
8, 9		
8, 9.3		
9, 9		
9, 9.3		
9.3, 9.3		
Total	**1**	**150**

Table 14. χ^2 test of difference between observed and expected genotypic frequencies predicted from Hardy-Weinberg principle on TH01 in your class.

i	Categories (genotypes)	Observed (O)	Expected (E)	(O-E)	$(O-E)^2$	$(O-E)^2/E$
2	6, 7					
3	6, 8					
4	6, 9					
5	6, 9.3					
6	7, 7					
7	7, 8					
8	7, 9					
9	7, 9.3					
10	8, 8					
11	8, 9					
12	8, 9.3					
13	9, 9					
14	9, 9.3					
15	9.3, 9.3					

k = χ^2 =

Test Statistic = χ^2 = ____________

Degree of freedom (d.f.) = k-1 = ___________

Tabular Statistic (α = .05, d.f. =) = __________________

Test Statistic $\geq$ or < Tabular statistic

Conclusion: ____

Class Exercise on Parentage Analysis and Paternity Index Determination

After your TA has completed the introduction of parentage determination complete the following procedure.

1. Provide the name of the obligate allele of each of the loci in the child in Table 15.

Table 15. Genotype of microsatellite markers in the child, mother, and potential fathers (PF1 & PF2).

	Genotype of Locus			
	D3S1358	**D7S820**	**D8S1179**	**D13S317**
Child	15, 17	8, 12	13, 14	12, 12
Mother	15, 16	8, 8	13, 14	11, 12
Obligate allele in child				
PF1	14, 16	8, 11	10, 15	11, 11
PF2	17, 18	12,12	13, 15	11, 12

2. Determine whether each of the potential fathers is included or excluded for paternity (circled one).

PF1: **Included** or **Excluded** PF2: **Included** or **Excluded**

3. Complete the individual PI of the included PF in Table 16.

Table 16. Paternity index of the potential father implicated by obligate alleles.

Locus	D3S1358	D7S820	D8S1179	D13S317
PI of (circle one) PF1 or PF2				

4. Calculate the Combined Paternity Index of the included PF. Because the loci are not linked you can apply the multiplication rule to generate a CPI.

CPI = ____

Tutankhamun Genealogy

The class will be divided into groups of three students in this exercise. Genotype of each of the seven family members and their sex are printed on a strip of paper. A blank family tree is there for you to determine the relationship among family members. To simplify the exercise, you are given that Tutankhamun was the last member in the family tree. In addition, recall the following information:

1. Only one of the two alleles of each locus is passed on to the offspring (Mendel's Law of Segregation).
2. These loci are located on separate chromosomes so alleles sort independently (Mendel's Law of Independent Assortment).
3. Some alleles are not passed on to descendants but all alleles from a descendent must have come from ancestors (similar to a parentage trio relationship).

Materials

Quantity	Item
1 per class	Thermal cycler (Applied Biosystems ABI 2720 or equivalent) with programmable touchdown PCR capability
1 per pair of student	20-200 µL micropipets
1 per pair of student	2-20 µL micropipets
1 per class	Water bath or heat block for 65° Celsius
1	DNA sequencing facility on or off campus
1 per 3 pairs of students	vortex mixer
1 per class	microcentrifuge for 0.2 mL tubes (optional)
1 per student	2.0 mL microcentrifuge tubes (DNase and RNase free)
4 per student	1.5 mL microcentrifuge tubes
1 box per pair of students	micropipet tips with barrier
1 per pair of students	Magnetic stands (Promega, two-position)
1 per student	Disposable cups
1 per 100 or 400 student samples	Promega DNA IQ System (100 or 400 reactions)
1 per 300 student samples (for 15 µL reaction)	Qiagen Multiplex PCR Kit (100-reaction)
4.4 µL per student sample	Primer mix (see Student Outline for sequences)
1 vial per class	Mineral oil for sealing PCR products
18 µL per student sample	Hi Di formamide Applied Biosystems)
30 µL per student sample	Molecular grade water
2-3 for up to 96 student samples	96-well skirted PCR plate
0.2 µL per student	LIZ 500 DNA size standards (Applied Biosystems)
1 per 20 students	Lonza FlashGel System

Notes for the Instructor

1. To ensure accurate conversion from a PCR fragment size to an allele, instructor should have his or her own genotyping conducted by a professional DNA typing lab or contact the author for further information. When amplifying student samples, instructors should use their own sample with known genotypes as a reference to convert student DNA fragment sizes to alleles.
2. Because of the lower resolution of multiplex DNA bands on the electrophoretic gel, the instructor may consider a separate PCR using only two to three loci in addition to the 10-locus multiplex. The ideal loci of choice should include those with alleles that are quite different in sizes for easier distinction on the gel. The sizes of the fragments deduced from the standard curve complement the information learned from analyzing an electropherogram.
3. For parentage analysis, coverage of paternal index may only apply to students who are more prepared in mathematics. Identifying the obligate allele will be sufficient in helping the student understand the Mendel's Laws of Segregation and Independent Assortment.
4. If the class size is large enough, you may use student DNA data to conduct a chi square test to see if student population follows the prediction of Hardy-Weinberg equilibrium (which is always the case). To minimize the possible number of genotypes it is advisable to choose a locus with the lowest number of alleles, such as TH01 or D16S539. Example results based on data collected from 150 students at Cornell University in Fall 2011 are shown in the following tables.
5. Contamination may be a very serious issue because human DNA is ubiquitous. However, this may be prevented by using filter pipet tips and wearing gloves during DNA extraction and PCR.
6. For the Tutankhamun family tree exercise I used a pedigree with slight modifications from the convention (Fig. 4). To make the exercise more challenging, the instructor may withhold the information of where Tutankhamun is situated in the family tree. One of the example paper strips showing genotypic profile is shown in Figure 5. The graphic file of these genotypes may be provided upon request.

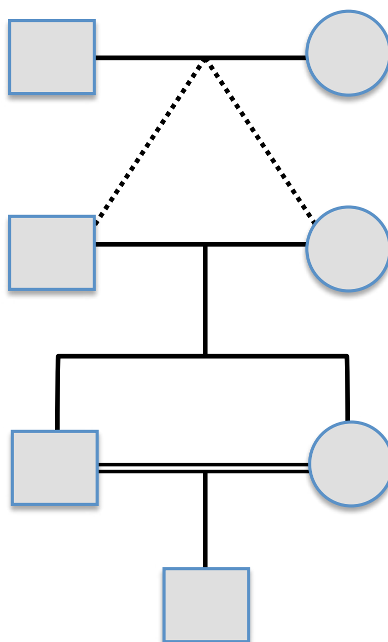

Figure 4. The blank family tree provided to student groups. The dotted lines shows uncertain descendant given rise by the first generation. The single horizontal line indicates marriage and the double horizontal line indicates sibling marriage
.

	D13S317		D7S829		D2S1338		D21S11		D16S539		D18S51		CSF1PO		FGA	
TIYE	11	12	10	15	22	26	26	29	6	11	19	22	9	12	20	26

Figure 5. An example of a genotypic profile of a member in the Tutankhamun family printed on paper for use in the student exercise. The circle at the left indicates a female member (where there will be a square if male). The name of the individual is shown below the circle. Eight genetic markers are shown on the top row and the two alleles of each marker are shown below the name of the marker. The color shading helps tracking the alleles between generations.

Acknowledgements

The author would like to thank the help of Martha Lyon, Steven Bogdanowicz, Paul Goncharoff for their assistance to make this lab possible in a large biology lab course.

Literature Cited

Ashma, R., & Kashyap, V. K. (2002). Genetic polymorphism at 15 STR loci among three important subpopulation of Bihar, India. *Forensic Science International*, 130(1), 58-62.

Budowle, B., Chidambaram, A., Strickland, L., Beheim, C. W., Taft, G. M., & Chakraborty R. (2002). Population studies on three Native Alaskan population groups using STR loci. *Forensic Science International*, 129(1), 51-57.

Butler, J. M. (2005). *Forensic DNA typing* (2nd ed.). London: Elsevier Academic Press.

Butler, J. M. (2011). S*hort tandem repeat DNA Internet DataBase*. Retrieved September 12, 2011, from **http://www.cstl.nist.gov/biotech/strbase/**

Butler, J. M., Schoske, R., Vallone, P. M., Redman, J. W., & Kline, M. C. (2003). Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. *Journal of Forensic Science,* 48(4), 908-911.

Evett, I. W., & Weir, B. S. (1998). Parentage testing. *Interpreting DNA evidence: Statistical genetics for forensic scientists.* (pp. 163-187). Sunderland, Massachusetts: Sinauer Associates, Inc.

Goldstein, D. B., & Schloitterer, C. (Eds.). (2001). Microsatellites: Evolution and applications. Oxford ; New York: Oxford University Press.

Hashiyada, M., Itakura, Y., & Nata, M. (2003). PowerPlex 16 analysis in the Japanese population. *International Congress Series*, 1239, 125-130.

Hawass, Z., Gad, Y. Z., Ismail, S., Khairat, R., Fathalla, D., Hasan, N., *et al*. (2010). Ancestry and pathology in Tutankhamun's family. *The Journal of the American Medical Association*, 303(7), 638-647.

Kim, Y., Hwang, J., Kim, Y., Lee, S., Chung, N., Goh, H., *et al*. (2003). Allele frequencies of 15 STR loci using AmpF/STR Identifiler kit in a Korean population. *Forensic Science International*, 136(1-3), 92-95.

Li, W. (1997). Genome organization and evolution. *Molecular evolution* (pp. 379). Sunderland, MA: Sinauer Associates, Inc.

Sadava, D., Hillis, D. M., Heller, H. C., & Berenbaum, M. R. (2011). *Life, the science of biology* (9th ed.). Sunderland, MA: Sinauer Associates, inc.

Wang, W., Jia, H., Wang, Q., Cai, Z., Wei, L., Wang, D., *et al*. (2003). STR polymorphisms of "forensic loci" in the northern Han Chinese population. *Journal of Human Genetics*, 48(7), 337-341.

About the Author

Kuei-Chiu Chen received her Ph.D. in biology from New York University. She has worked in New York City's Forensic Biology Department of the Office of Chief Medical Examiner. She was the lab coordinator for three introductory biology courses at UCLA. In 2005 she joined Cornell University as Director of Introductory Biology Laboratories, which was later changed to Investigative Biology Laboratory after a curricular reform. She is interested in further developing project-based lab modules and assessment of student learning.

Appendix

Material preparation and products/vendors/catalog numbers

1. Pre- PCR preparation

Use 2 mL microcentrifuge tube

Per student: Lysis buffer :200 µL, Resin – 7 µL

Put each of these into the tube individually. Because resin beads are heavy and sink to the bottom of the tube, do not make a large mix and then aliquot 207 µL into each tube.

Use 1.5 ml microcentrifuge tubes – use a different color for each solution

Lysis buffer –	220 µL
Wash buffer –	660 µL
Elution buffer–	220 µL

Total of 1000 µL/one pair of students

To collect student DNA, use 0.2 mL non-skirted low profile 96-well PCR plates (ABgene PCR plates AB-0700 from Fisher Scientific)

To seal wells on the 96-well plate, usealuminum sealing tape (Fisher Scientific 07-200-684).

Put the strips onto a Preparation Rack for PCR thin-walled tubes (Fisher Scientific 05-541-color code).

Seal the PCR plate strip with a strip of the sealing tape. Use a plate roller to press the sealing tape onto the plate strip. (Plate roller, if needed, from Sigma-Aldrich R1275-1EA)

Done in lab by students: Using the pipet tip loaded with their purified DNA, the student punctures the strip on the well and dispensed DNA into the well. The punctured strips will be sealed by the same aluminum strip.

Save the tubes of student DNA and store at -20°C in case students did not pipet enough of their DNA into the sample well.

Evaporate the alcohol from the student DNA samples. Leave the plate strips in the preparation rack. Remove the sealing tape. Dry at 46°C for 5 minutes.

2. To run multiplex PCR

For 2 µL of DNA/well use 18 µL of the following/well

2X Master Mix –	10 µL/well
Q solution –	2 µL/well
H_2O from Qiagen kit –	1.6 µL/well
Primer mix –	4.4 µL/well
	18 µL

Store the above Super Master Mix at -20°C

<u>Other Items</u>

Qiagen multiplex PCR kit – Cat #206143 – for 100 reactions

For the water you can also use Sigma-Aldrich Cat#W4502

Use a PCR plate appropriate for your PCR machine.

<u>Steps</u>

1. Add 18 µL of the Super Master Mix to the appropriate wells.
2. Add 2 µL of student DNA to the appropriate wells.
3. Add 1 drop of mineral oil to each well being used. (Sigma-Aldrich M5904-5 mL mineral for molecular biology).
4. Seal the plate with a sheet of the thermo well sealing tape. Use the plate roller to press the tape firmly in place.

5. Run PCR based on the program specified below:

95°C	15 min	
94°C	50 sec	
60-54°C	45sec	(Touchdown, 1-7)
72°C	1 min	
94°C	50sec	
54°C	45sec	28x
72°C	1 min	
60°C	45 min	
4°C	Hold	

3. Dilution of PCR product

[Note: Make 1:10 dilution from original PCR product. If small sample size, say 20 samples or less, add 2 µL of PCR product first, then add 18 µL of water for better mixing. If large sample size, see below]

Sigma water –	18 µL/well (Cat #W4502-1L)
PCR product –	2 µL/well
	20 µL/well

BD Falcon 96-well assay plate is used for the dilution. (BD#35390. Fisher Scientific cat#08-772-5)

Figure 6. A typical chromatogram produced from a multiplex PCR on ten microsatellite loci. The y-axis in each of the four panels represents the fluorescent unit as indication of quantity of DNA fragments carrying that specific color label. The x-axis is the size of fragments in base pairs and the color bar at top of each panel shows the range of the allele sizes in the population. The allele of each locus is shown as a peak with base pair size indicated in the box below the peak. The loci with two peaks are heterozygous and the loci with one peak are homozygous.

For Fragment Analysis: It depends on the sequencing facility. At Cornell University's Life Sciences Core Laboratories Center the following is used.

		50 wells
LIZ 500 –	0.2 µL/well	10 µL
Hi-di formamide –	17.8 µL/well	890 µL
	18.0 µL	900 µL

Store at -20°C

(Hi-di formamide from Applied Biosystems, genetic analysis grade (cat #4311320)for 25 mL. Store at -20°C.

GeneScan-500 LIZ size standard Applied Biosystems (cat#4322682)

Hi-di formamide + LIZ-500	18 µL/well
Diluted PCR product	2 µL/well
	20 µL/well

4. Run Peak Scanner® on PCR results

Download Peak Scanner from the Applied Biosystems website. They will require you complete a short survey before allowing you to download this program for free. The instructions are also available for download. I have condensed their instructions and can share with ABLE members upon request. After the results are shown on the computer screen, I take a screen shot and paste the image onto a blank Power Point slide. On the slide I added the label of each locus. See Figure 6 below for one example.

Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with biology learning and teaching in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit **http://www.ableweb.org/**

Papers published in *Tested Studies for Laboratory Teaching: Proceedings of the Conference of the Association for Biology Laboratory Education* are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board.

Citing This Article

Chen, K. 2012. Microsatellite DNA: Population Genetics and Forensic Applications. *Tested Studies for Laboratory Teaching,* Volume 33 (K. McMahon, Editor). Proceedings of the 33rd Conference of the Association for Biology Laboratory Education (ABLE), 390 pages. http://www.ableweb.org/volumes/vol-33/?art=2

Compilation © 2012 by the Association for Biology Laboratory Education, ISBN 1-890444-15-4. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright owner.

ABLE strongly encourages individuals to use the exercises in this proceedings volume in their teaching program. If this exercise is used solely at one's own institution with no intent for profit, it is excluded from the preceding copyright restriction, unless otherwise noted on the copyright notice of the individual chapter in this volume. Proper credit to this publication must be included in your laboratory outline for each use; a sample citation is given above.