EDVOTEK®

THE **BIOTECHNOLOGY EDUCATION** COMPANY®

AP BIOLOGY

BIG IDEA 1 INVESTIGATION 2

Edvo-Kit #AP02

Mathematical Modeling: Hardy-Weinberg

REVISED & UPDATED

Experiment Objective:

In this exercise, students will determine whether they are PTC tasters. They will use the Hardy-Weinberg equation to analyze the collected class data. Finally, students will use a computer spreadsheet to build a mathematical model that describes the behavior of the PTC gene in a hypothetical gene pool. After performing these exercises, students will understand how mathematics and model populations have enhanced the study of population genetics.

See page 2 for storage instructions.

Table of Contents

Safety Data Sheets are available on our website: www.edvotek.com/safety-data-sheets

Experiment Components and Requirements

COMPONENTS

- PTC taste paper
- Control taste paper

REQUIREMENTS

- Computer with spreadsheet software (e.g. Microsoft® Excel)
- Calculator with square root function

Store the entire experiment at room temperature.

This experiment is designed for 10 lab groups.

Background Information

The study of genetics often focuses on the small changes in genes that can affect the appearance, health, and well being of organisms. These different gene forms, called alleles, are passed from parent to offspring following the rules of Mendelian genetics. The combination of alleles, or genotype, determines an organism's observable characteristics, or phenotype. Each offspring has a 50/50 chance of inheriting a specific allele from the parental generation. Many alleles are dominant or recessive. When a dominant allele is inherited, it will mask the trait coded by the recessive allele. With these rules and the parental genotypes, we can predict an offspring's genotype and phenotype.

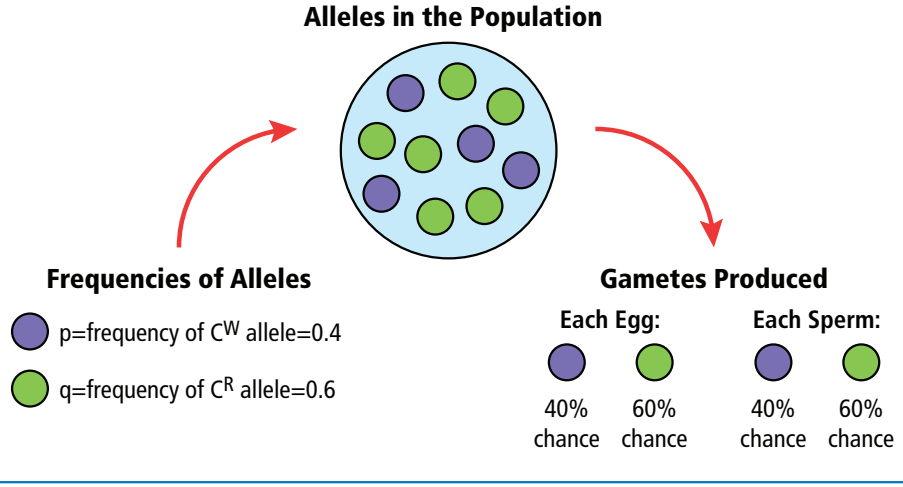
In early attempts to mathematically model the genetics of a population, scientists assumed the dominant allele would outcompete the recessive allele over time. However, research showed that allele frequencies remained stable over several generations, as long as the alleles segregated independently and the chromosomes recombined during meiosis. Only catastrophic changes in the population affected the allelic frequency. Independently, the British mathematician G.H. Hardy and the German physician W. Weinberg created a set of equations that describe the frequencies of alleles from one generation to the next under a set of "idealized conditions" (Figure 1). These classic equations, now called the Hardy-Weinberg principle of genetic equilibrium, have become the basis for population genetics.

Figure 1: Criteria to Meet Hardy-Weinberg Equilibrium.

- **Absence of new mutations**
- **Random mating between individuals**
- **Natural selection is not occurring**
- **Very large population size**
- **No migration into/out of population.**

In extremely large populations, we can imagine that all the alleles in a population are in one common gene pool instead of within individuals (Figure 2). Each allele exists as a gamete at a specific frequency within the gene pool (i.e. the dominant allele "p" occurs 40% of the time, and the recessive allele "q" occurs 60% of the time). The frequencies of the two alleles must equal 100%, which represents the entire gene pool. This leads to the first of the Hardy-Weinberg equations, where p + q = 1. In this example, the frequency of the p allele is 40%, or 0.4, and the frequency of the q allele is 60%, or 0.6. The frequencies of the two alleles add up to 1, the total population.

Figure 2: At the Population Level, Genes Are Randomly Selected.



To create the next generation, two gametes are chosen from the gene pool at random and combined to create three different genotypes – pp, pq and qq. The likelihood of each genotype can be calculated

using mathematical probability rules. The multiplication rule states that to determine the probability of two events occurring (i.e. the selection of two alleles), we multiply the probability of the first event by the probability of the second event: $(p + q)(p + q) = 1$. Expanding this equation generates the second Hardy-Weinberg equation: $p^2 + 2pq + q^2 = 1$. When this equation is applied to an ideal population, the frequency of homozygous dominant individuals is defined as p^2, the frequency of the heterozygotes is $2pq$, and the frequency of homozygous recessive individuals is q^2.

By having the alleles combine at random, away from evolutionary pressures and unexpected gene flow, gene frequencies remain constant over many generations. However, it is impossible to eliminate these evolutionary forces in most natural populations. In these situations, the Hardy-Weinberg theorem is useful since unexpected deviations can point to the occurrence of evolutionary significant events such as speciation.

STUDYING SINGLE ALLELE TRAITS IN HUMANS

Cystic fibrosis (CF) is a disease that results from a mutation in the CFTR gene, which codes for a channel that moves chloride ions across cell membranes. This ion channel is crucial for the proper transport of water within tissues. Mutations in this gene lead to accumulation of thick, viscous mucus in the lungs and other organs. This mucus acts as breeding ground for dangerous infections, meaning that individuals with CF must be careful to avoid contact with germs. Medical innovations have extended the average lifespan of individuals with CF to 40 years, and new treatments are being developed all the time.

In the United States, one out of every 2500 individuals of European descent is born with CF (cystic fibrosis is more rare in other ethnic groups). Genomic sequencing efforts have identified over 1,700 variants of the CF gene, all of which are recessive to the wild-type allele. Since we know how often individuals are born with cystic fibrosis, we can calculate how often each allele appears in the population.

In these calculations, we will define C as the dominant wild-type allele and c as the recessive disease allele. For simplicity, we will count all the variants as a single recessive allele. First, we convert the raw numbers to percentages. Since CF occurs in 1 of every 2500 births, we can divide 1 by 2500 to convert the frequency to a fraction.

$1/2500 = 0.0004$ (or 0.04%)

This value represents individuals with two copies of the recessive allele, or q^2. Knowing this, we can calculate the value for q.

$q^2 = 0.0004$ therefore $q = \sqrt{0.0004}$
$q = 0.02$, or the frequency of the c allele.

Knowing that $p + q = 1$, we can substitute our q value into the equation and solve for p.

$p + 0.02 = 1$ therefore $p = 1 - 0.02$
$p = 0.98$, or the frequency of the C allele

Next, we can use the allele frequencies to determine the frequency of the genotypes in the population.

Homozygous wild-type (CC) = $p^2 = 0.98^2 = 0.9604$
Heterozygous carriers (Cc) = $2pq = 2 * 0.98 * 0.02 = 0.0392$
Individuals with CF (cc) = $q^2 = 0.02^2 = 0.0004$

To check our work, we will add the frequencies of each genotype. To be in Hardy-Weinberg equilibrium, the sum must be equal to one.

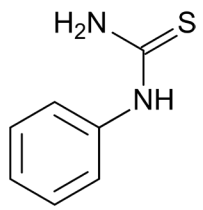
$p^2 + 2pq + q^2 = 1$
$0.9604 + 0.0392 + 0.0004 = 1$ ✔

In reality, the likelihood of a CF carrier in an American of Northern European decent are 1 in 29, which is about 4% of the population. This is very close to the value we calculated above, suggesting that the population is in equilibrium.

INVESTIGATING THE GENETIC OF BITTER TASTE

Individuals vary greatly in their sensitivity to the bitter compound Phenylthiocarbamide (PTC) (Figure 3). This fact was discovered in 1931 in a series of events that involved impressive scientific curiosity and questionable laboratory safety. A chemist named Arthur Fox was mixing a powdered chemical when he accidentally let a bit of the powder blow into the air. A nearby colleague exclaimed how bitter the powder tasted, but Fox (who was closer to the chemical) tasted nothing. Interested, both men took turns tasting the chemical. Fox continued to find the chemical tasteless while his college found it bitter. Next, Fox tested a large number of people. Again he found a mix of "tasters" and "non-tasters" and published his findings. This caught the interest of geneticist L.H. Snyder who tested the compound on families and hypothesized that the taster/non-taster state was genetically determined.

Figure 3: The structure of PTC.



Ability to taste PTC compound is now linked to the protein Taste Receptor 2 Member 38 that is encoded by the TAS2R38 gene. TAS2R38 has two alleles: the dominant allele (T), which confers the ability to taste PTC, and the recessive non-taster allele (t). A person inherits one copy of the gene from each of his/her parents. The combination of these different alleles dictates whether an individual is a "taster" or "non-taster". PTC-tasters have one of two possible genotypes; either they are homozygous dominant and have two copies of the taster allele (TT), or they are heterozygous and have one taster allele and one non-taster allele (Tt). "Nontasters" are homozygous recessive and have two copies of the non-taster allele (tt). Within the general population, about 70% of the people tested can taste PTC, whereas the other 30% cannot. The ability to taste PTC has no known selective advantage, making it an excellent gene to explore in class.

In this exercise, students will determine whether they are PTC tasters. They will use the Hardy-Weinberg equation to analyze the collected class data. Finally, students will use a computer spreadsheet to build a mathematical model that describes the behavior of the PTC gene in a hypothetical gene pool. After performing these exercises, students will understand how mathematics and model populations have enhanced the study of population genetics.

Experiment Overview and General Instructions

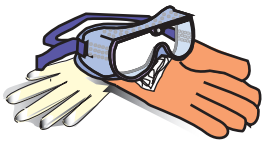
EXPERIMENT OBJECTIVE

In this exercise, students will determine whether they are PTC tasters. They will use the Hardy-Weinberg equation to analyze the collected class data. Finally, students will use a computer spreadsheet to build a mathematical model that describes the behavior of the PTC gene in a hypothetical gene pool. After performing these exercises, students will understand how mathematics and model populations have enhanced the study of population genetics.

WORKING HYPOTHESIS

If there is no selection for any allele in a large randomly-mating population, then the gene frequencies will remain constant over many generations. However, if there are outside forces such as selection for an allele, heterozygote advantage, and genetic drift working in a population, then the gene frequencies will change over time.

LABORATORY SAFETY GUIDELINES



1. Wear personal protective equipment while working in the laboratory.
2. Always wash hands thoroughly with soap and water after working in the laboratory.
3. If you are unsure of something, ASK YOUR INSTRUCTOR!

LABORATORY NOTEBOOKS

Scientists document everything that happens during an experiment, including experimental conditions, thoughts and observations while conducting the experiment, and, of course, any data collected. Today, you'll be documenting your experiment in a laboratory notebook or on a separate worksheet.

Before Starting the Experiment:

- Carefully read the introduction and the protocol. Use this information to form a hypothesis for this experiment.
- Predict the results of your experiment.

During the Experiment:

- Record your observations.

After the Experiment:

- Interpret the results - does your data support or contradict your hypothesis?
- If you repeated this experiment, what would you change? Revise your hypothesis to reflect this change.

Investigation I: Calculating the Frequency of PTC Alleles Within a Small Population

In this exercise, each student will determine whether or not they are PTC tasters. Using the class data and the Hardy-Weinberg equations, allelic frequencies are calculated.

Each student should receive the following materials:

- PTC Paper
- Control Taste Paper

PROCEDURE:

1. **TASTE** the Control strip of paper first. **RECORD** your thoughts on the taste.

_____

2. **TASTE** the PTC strip of paper. **RECORD** your thoughts on the taste.

_____

3. **COMPARE** the taste of the Control and the PTC paper.
   - Notice what the PTC paper tastes like compared to the Control paper: intensely bitter, somewhat bitter, or tasteless.
   - If you are a taster, the PTC paper strip will be bitter. Non-tasters will not notice a difference between the strips of paper.

4. **RECORD** the total number of tasters and the total number of non-tasters on the whiteboard and in your notes.
   a. Tasters ($p^2 + 2pq$): _____

   b. Non-tasters (q^2): _____

5. **CALCULATE** the percent of tasters and non-tasters in the class. Record your values in Table 1.

6. **CALCULATE** the allele frequency for your class and the North American population using the Hardy-Weinberg equations. See the sample calculation on page 4 for reference. Record your values in Table 1.

Table 1: Prevalence of PTC Alleles in Selected Populations.

	Phenotypes		Calculated Allele Frequencies	
	Taster (%)	Non-Taster (%)	p (Taster allele)	q (Non-taster allele)
Class Population				
North American Population	70	30		

Investigation I: Calculating the Frequency of PTC Alleles Within a Small Population, cont.

STUDY QUESTIONS:

1. Calculate the number of homozygous tasters (p^2), heterozygous tasters (2pq), and homozygous non-tasters (q^2) in your class.

2. Calculate the number of homozygous tasters (p^2), heterozygous tasters (2pq), and homozygous non-tasters (q^2) in North America, given a population of approximately 580 million people in 2016.

3. How do the allelic frequencies from your class compare to the allelic frequencies for North American populations? How do the frequencies of each genotype compare? Explain your answer.

EXTENSION ACTIVITY:

The original research paper looking at the genetics of PTC tasting is available online through Science Magazine (free with registration, available http://science.sciencemag.org/content/74/1910/151). Students should read the paper and calculate the allelic frequencies using Snyder's data.

Investigation II: Building a Mathematical Model of a Population in Hardy-Weinberg Equilibrium

A. EXPLORING SPREADSHEET FUNCTIONS

Scientists often use computers to create models of biological systems so that they can study complex phenomenon. For example, complex mathematical models allow meteorologists to forecast the weather. Researchers can change certain parameters within the weather system (i.e. temperature, sunshine, air quality, clouds, etc.) to determine their influence on the model. This allows for more accurate predictions of the weather. In the same way, population geneticists create models that examine the influence of environmental factors on a given population. In this investigation, you will create a simple population genetics model using a computer spreadsheet program. This will allow you to explore how allele frequencies in a population can change. This model was created using Microsoft® Excel, but any spreadsheet program should work.

Helpful Hints:

- Options for formatting cells can be found in the status bar or Format>Cells>Fill menu or by right clicking to open flyout menu and going to format cell menu.
- Be sure to save your work periodically in case the program closes unexpectedly.
- Be sure to clear your spreadsheet or to open a new workbook before starting the exercise.
- In addition to the normal operators (+, – ,*, /, =), you will find the following commands necessary to build your model.

Formula	Description
RAND()	Generates a random number between 0 and 1.
IF (test, value if true, value if false)	Evaluates a statement as true or false and produces an output. Example: =IF(G5>1, "A", "B") Output: If the value in the cell G5 is greater than 1, the output will be A. If the value in cell G5 is less than 1, the output will be B.
CONCATENATE (text 1, text 2, ...)	Combines text from two or more spreadsheet cells
SUM (number 1, number 2)	Adds selected values
$	Anchors a cell identity when placed before a row of column identifier.

If you are not comfortable using these functions in spreadsheet programs, be sure to try them out before moving on with the exercise. For example, the =RAND() function will be used to randomly select gametes in our population. How does this formula work? Let's explore the =RAND() function together.

1. Enter the command **=RAND()** in an empty cell in your spreadsheet and hit **ENTER**. A number between 0 and 1 should appear in the cell. Record the number in a cell or on a sheet of paper.

2. Click on the same cell to make it active again. The formula will appear in the formula bar at the top of the screen.

3. To generate new random numbers, press **F9** on a PC or **Command +** on a Mac. This refreshes the spreadsheet. After each refresh, record the number. Do you get the exact same number twice?

4. Enter any information into another cell and hit **ENTER**. You will notice that the =RAND() value changes. This is a normal behavior of the function and it will not affect the predictive power of your model.

Investigation II: Building a Mathematical Model of a Population in Hardy-Weinberg Equilibrium, cont.

B. USING A SPREADSHEET TO MODEL THE PTC GENE IN A POPULATION

1. Establish the "gene pool," i.e. the frequencies of each allele in the population.

    a. In this exercise, we will define p as the frequency of the dominant taster allele (T). Enter a label for "p" in cell A2. The frequency of "p" will be in cell B2.

    b. In this exercise, we will define q as the frequency of the recessive non-taster allele (N). Enter a label for "q" in cell A3. The frequency of "q" will be in cell B3.

    c. Enter a value for p in A2. Since we're modeling the PTC gene, let's start with 0.70. **Knowing the value of p, what equation do we use to solve for q?** Enter this equation into cell B3.

    d. Select all of the cells from A2 to B3 and color them blue (or a color of your choice). This area represents the gene pool.

2. Randomly generate offspring.

    a. Type "Egg" into cell C4 and "Sperm" into cell D4. The columns will represent the genotype of the gamete from each parent.

    b. In cell C5, enter **=IF(RAND()<=B2,"T","N")** and press **ENTER**. A letter will appear in the cell. **What does this equation mean, and what do the output letters represent?**

    c. Enter the same formula in cell D5.

    d. Type "Offspring" into cell E4. This column represents the combination of the two alleles randomly selected in cells C5 and D5. To combine the alleles and create your first offspring, enter **=CONCATENATE(C5, D5)** into cell E5.

    e. Press the **F9** or **Cmd +** key to refresh your spreadsheet. If you have entered the functions correctly, the values in the gamete cells (C5 and D5) should change randomly. The value in the offspring cell (E5) should be the combination of the two alleles.

    f. To create more offspring, click and drag to select the three completed cells (C5, D5, and E5), then drag straight down without releasing the mouse button. To create 30 offspring total, drag down to row 34 before releasing the mouse button. Then, go to Edit>Fill>Down to populate the cells with the correct formulas.

    g. If desired, color the Gametes orange and the Offspring green. This will help you keep the data organized.

3. Tabulate the genotypes of the offspring.

    a. Select cells F3, G3, and H3 and merge into one cell using the "Merge Cells" function. Enter text "Number of Each Genotype" and align the text to the center of the cell. These columns will be used to record the number of offspring of each genotype.

    b. Type "TT", "TN", and "NN" into cells F4, G4, and H4, respectively. These columns will track the number of offspring of each genotype.

Investigation II: Building a Mathematical Model of a Population in Hardy-Weinberg Equilibrium, cont.

c. The first column keeps track of all of the homozygous tasters (TT). Enter **=IF(E5="TT", 1, 0)** in cell F5 and press **ENTER**. This formula evaluates the contents of cell E5 and records the output in cell F5. **In your own words, what is this formula doing?**

d. The second column keeps track of all of the heterozygous tasters (TN, NT). However, the IF command will only register a "true" statement if the value matches EXACTLY, meaning that the IF command considers TN and NT to be different. To address this problem, we will use a nested function. This means that if the first logical test fails, the contents are reevaluated using a second logical test that is listed as the "value if false" in the IF statement. In cell "G5" enter the function: **=IF(E5="TN",1,(IF(E5="NT",1,0)))** and press **ENTER**. **In your own words, what is this formula doing?**

e. The third column keeps track of all of the homozygous non-tasters (NN). Enter **=IF(E5="NN", 1, 0)** in cell H5 and press **ENTER**. This formula evaluates the contents of cell E5 and records the output in cell H5.

f. Evaluate your equations. Press the **F9** or **Cmd +** key to refresh your spreadsheet. If you have entered the functions correctly, the value in the genotype cells (F5, G5, H5) should read "1" in the column that matches the genotype. Repeat several times until you are satisfied that the equations are working properly.

g. To evaluate the remaining offspring, click and drag to select the three completed cells (F5, G5, and H5), then drag straight down to row 34 before releasing the mouse button. Then, go to Edit>Fill>Down to populate the cells with the correct formulas.

h. Enter the label "Totals:" in cell E36.

i. Cell F36 will represent the total number of TT offspring produced by the simulation. This is calculated by adding all of the above rows using the SUM function. Repeat the SUM function in cells G36 and H36 to calculate the total number of heterozygous tasters (TN, NT) and homozygous non-tasters (NN).

j. Ensure that your spreadsheet is correctly tabulating the different genotypes. Count the number of ones in each genotype column. The number should match the sum in the totals row. Using the **F9** or **Cmd +** key, refresh your spreadsheet and reevaluate the totals. Repeat several times until you are satisfied that the totals update correctly each time a new population is generated.

4. Using the chart menu, create a histogram using the genotype totals.

a. Click and drag to select cell F36, G36, and H36.

b. Select the chart tab to activate the chart options, and select column from the options. This should create a histogram on your active sheet

c. Edit the labels for each bar in the histogram to match the appropriate genotype.

d. The visual representation of the different genotypes should change each time the spreadsheet is refreshed. Using the **F9** or **Cmd +** key, refresh your spreadsheet and reevaluate the histogram. Repeat several times until you are satisfied that the graph updates correctly each time a new population is generated.

Investigation II: Building a Mathematical Model of a Population in Hardy-Weinberg Equilibrium, cont.

5. Use the genotypes generated by the spreadsheet to calculate the frequencies of alleles for the next generation.

   a. Label cell E38 as "#T Allele" and cell E39 as "#N Allele". Label cell H38 as "p (freq. of T)" and cell H39 as "q (freq. of N)".

   b. Each offspring has two alleles that are in three possible combinations – TT, TN, NN. **Knowing this, create an equation in cell F38 that calculates the number of taster alleles in the population that we created. In your own words, how does this equation calculate the number of alleles?**

   c. **In cell F39, create an equation that calculates the number of non-taster alleles in the population that we created. Does it work the same way as your taster allele equation?**

   d. Using the values in cells F38 and F39, calculate the frequency of p and q. **What formula did you use to calculate the frequency? Is your system still in Hardy-Weinberg Equilibrium? Is the allele frequency in the next generation the same as the initial allele frequency?**

Investigation III: Using Your Mathematical Model to Explore Population Genetics

Once your spreadsheet is working properly, use it to explore questions in population genetics. Below are some suggestions. Be sure to research the topic and to formulate a hypothesis to be tested. Does the data support or reject your hypothesis? Explain your reasoning in your lab report.

1. **Behavior of allele frequencies when initial frequencies are changed.**

   In this simulation, we set the p value to 0.70 to simulate the behavior of the PTC gene in the North American population. What happens to the allele frequencies in the next generation if you change the initial frequencies? Perform an Internet search to determine the frequencies of some common human traits and see how they perform in your model.

2. **Behavior of allele frequencies in large populations.**

   In our initial spreadsheet, we created a population of thirty offspring. While this seems like a large number of individuals, it represents an extremely small population. Create a new spreadsheet with more rows and compare the initial allele frequencies to the allele frequencies in the next generation. Do they fluctuate more or less than before?

3. **Behavior of allele frequencies over multiple generations.**

   We have explored how allele frequencies can change in one generation, but how do they change over multiple generations? Try modeling this by starting a new generation using the p and q values calculated at the end of the existing generation. Each subsequent generation depends on the p and q values from the previous generation. How do gene frequencies change over time? Does the population size affect this?

4. **Behavior of allele frequencies in response to environmental changes**

   A genetic bottleneck occurs when a sudden, drastic change occurs in an ecosystem that devastates the population of an area. The remaining few animals will mate and produce healthy offspring, but the genetic diversity of the area will have been dramatically changed. The ecosystem may incur genetic drift until a larger population can be established. How could you use the spreadsheet to model a genetic bottleneck?

STUDY QUESTIONS:

1. In *Drosophila melanogaster*, normal wings (W) are dominant over curly wings (w). In a large population, it was shown that the curly wings (ww) are present 10% of the time. Using this information, calculate the frequency of the WW and Ww genotypes.

2. A gene has been identified that increases an individual's chances of developing cancer within their lifetime. The disease-causing allele, d, is recessive to the wild-type allele, D. Data from genetic testing of 1000 unrelated individuals revealed the following genotypes: 401 DD, 363 Dd, and 236 dd. Analyze the population using the Hardy Weinberg equations. Is this population in equilibrium? Explain your answer.

Instructor's Guide

NOTES TO THE INSTRUCTOR AND PRE-LAB PREPARATIONS

The "hands-on" laboratory experience is a very important component of science courses. Laboratory experiment activities allow students to identify assumptions, use critical and logical thinking, and consider alternative explanations, as well as help apply themes and concepts to biological processes.

EDVOTEK experiments have been designed to provide students the opportunity to learn very important concepts and techniques used by scientists in laboratories conducting biotechnology research. Some of the experimental procedures may have been modified or adapted to minimize equipment requirements and to emphasize safety in the classroom, but do not compromise the educational experience for the student. The experiments have been tested repeatedly to maximize a successful transition from the laboratory to the classroom setting. Furthermore, the experiments allow teachers and students the flexibility to further modify and adapt procedures for laboratory extensions or alternative inquiry-based investigations.

ORGANIZING AND IMPLEMENTING THE EXPERIMENT

Class size, length of laboratory sessions, and availability of equipment are factors, which must be considered in the planning and the implementation of this experiment with your students. These guidelines can be adapted to fit your specific set of circumstances.

To perform this experiment, students will need access to a computer and a spreadsheet program.

Investigations I and II can be completed in one class period or less. Investigation III can be completed in one class period, or it can be assigned as an independent exercise at home. If you do not find the answers to your questions in this section, a variety of resources are continuously being added to the EDVOTEK web site at **www.edvotek.com**

In addition, Technical Service is available from 8:00 am to 5:30 pm, Eastern Time Zone. Call for help from our knowledgeable technical staff at **1-800-EDVOTEK** (1-800-338-6835).

Pre-Lab Preparations

INVESTIGATION I:

- Distribute control and PTC paper.
- After tasting exercise is complete, be sure to record data to distribute to the class. If more than one section is performing the activity, it may be interesting to combine the data and calculate the allelic frequencies for the entire group.

INVESTIGATION II AND III:

Be sure that students have access to a computer spreadsheet program before beginning the exercise. This tutorial was created using Microsoft® Excel, but any spreadsheet program should work.

Please refer to the kit insert for the Answers to Study Questions