BioMath Materials for High School Students

Margaret (Midge) Cozzens

Rutgers University Center for Discrete Mathematics and Theoretical Computer Science (DIMACS)

Research Professor and Education Projects Director

midgec@dimacs.rutgers.edu or midge6930@comcast.net

609-654-2104

Rutgers mailing address: DIMACS, 419 CoRE, Busch Campus, Rutgers University, 96 Frelinghuysen Drive, Piscataway, NJ 08854-8018

Abstract

This paper describes two companion projects designed to bring the interface between mathematics and biology into high school classrooms through the use of one week modules on a variety of biology topics. Samples from two of the modules are included after the general description of the projects and a list of the twenty completed modules.

BioMath Materials for High School Students

1. BioMath Material Development

Biology as a discipline is a small component of school learning from elementary school through high school and is most often the course taken in college to satisfy a science requirement. On the other hand, mathematics has always had the luxury and responsibility of being recognized as a major fundamental part of all school learning, indeed one of the three R's. Early on, the relationships between mathematics and the physical sciences have been appreciated and often have been used as a reason to study mathematics and its applications to the physical sciences. However, the interplay between mathematics and the biological sciences was understood by only a few.

All this has changed! Increasingly, many biological phenomena are being viewed as involving the processing of information which ultimately involves using the mathematical sciences. Modern mathematical and information sciences have played an important role in many major biological accomplishments, for example sequencing the human genome, and are of fundamental importance in the rapidly-expanding and evolving concept of "digital biology". Use of the mathematical sciences increasingly appears in papers and books in all areas of biology, while at the same time, biological ideas inspire new concepts and methods in the mathematical and information sciences. More and more, undergraduate and graduate students are being exposed to this interplay between the mathematical and biological sciences.

The National Science Foundation in 2006, and again in 2010, recognizing the need to facilitate the introduction of mathematics into biology classes and biology into mathematics classes, has provided grants to the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) at Rutgers University, in partnership with the Consortium for Mathematics and its Applications (COMAP), and Colorado State University (CSU), called the *Bio-Math Connection* (*BMC) and Interdisciplinary Mathematics and Biology (IMB).* BMC/IMB is a

pioneering project linking biology and mathematics in the high schools. It provides an opportunity for high school teachers, writers, researchers, and others to get in on the ground floor of developing innovative classroom materials. The principal goal of the BMC program is to provide teachers with curricular materials that highlight the interconnections between the mathematical and biological sciences. In addition, a secondary goal is to provide help for teachers in using these new materials and understanding the interface between the two disciplines. Thus, the main product of this project is a set of twenty high quality bio-math modules, and one and two semester books of some/all of the modules, including teacher support materials, that can be flexibly adapted for use in a variety of courses at a variety of grade levels in either biology or mathematics classes, or both. The books are intended for a senior level non-calculus based course, which will satisfy a semester's worth of a state requirement for a fourth year of mathematics or science Eighteen of the modules on topics of computational molecular biology, mathematical epidemiology, and ecology/population biology, and others, are complete.

To write materials that develop the interconnections between the biological and mathematical sciences in high school requires first answering a number of important questions. What mathematics and what biology does a student need to know to advance to college and/or work at the interface of biology and mathematics? How do teachers add more mathematics to a very full biology curriculum, and how can biological applications be added to a very full mathematics curriculum? How do mathematics teachers learn enough biology to incorporate biological applications in their courses, and how do biology teachers learn enough mathematics to incorporate more mathematics in their biology courses? Does the sequencing of biology and mathematics courses at the high school level work so that what is learned in one course can be used in the other course? For example, if Algebra 1 and Biology are taught in different years, is that a problem? Discussions of the answers to these questions preceded and informed the writing of the modules and this discussion continues today. Teachers are involved in the production, testing, and dissemination of the BMC/IMB modules. They are critically important in determining what teachers need to be able to teach the modules.

Not only have exciting, challenging materials at the interface of biology and mathematics been developed for high school mathematics and biology classes, but much has been learned along the way. First and foremost, we have learned how important a solid multi-pronged on-going evaluation program can be for the whole BMC/IMB project. Second, we have learned that

teachers are eager for new materials to use in their classrooms, and eager to participate in ground breaking workshops and activities. We learned that providing teachers with support materials is essential and that we can provide support systems online. Third, and perhaps most surprising to all of us working with BMC/IMB is how extraordinarily responsive students are to learning mathematics and biology together. One might have expected this to be true in AP courses in biology or mathematics, or even in 11th or 12th grade classes, but it is true in 8th grade, 9th grade, and 10th grade classes at all levels in all parts of the country in urban, rural and suburban schools. For example, in an alternative high school participating in the 2007-08 field-test program, the teachers reported that student interest and enthusiasm was unusually high. They noted that the students were "very proud to be a part of this special group" using the BMC module: "During the week of teaching this module, attendance was unusually high in our classes. They wanted to be here (in school) and in our class. Discussion was very good, and we did not have to force them to do the work. There was a friendly competitiveness in how to do the algorithm and who got it right. They were focused in the classroom, which is very atypical. Not a lot of redirecting had to happen. It was fun week to be teaching."

The following is a list of the titles of the modules:

Computational Biology – *Spider Silk, Genetic Inversions, Evolution by Substitution, Microarrays – Array of Hope, Searching and Sequencing*

Epidemiology – *Imperfect Testing, Competition in Disease Evolution, Modeling Disease Outbreaks, Genetic Epidemiology*

Ecology – *Food Webs, Home Range Analysis, Ecological Footprints, Drawing Lines – Voronoi Diagrams, Habitat Formation – Help I'm Surrounded by Squirrels, Evolutionarily Stable Strategies*

Miscellaneous topics – *Tomography, Quorum Sensing, Biostatistics in Practice, The Neuroscience of Pain*

All teachers are welcome to try any one of these modules. Please contact Midge Cozzens at midge6930@comcast.net or Sol Garfunkel sol@comap.com if you are interested in reviewing a module. The website http://www.dimacs@rutgers.edu/IMB provides descriptions of the modules.

2. Excerpts from two modules –*Imperfect Testing* and *Genetic Inversions*

Each module has introductory material similar to the following:

Imperfect Testing

by Nancy Crisler, Tasha Fingerlin, Tom Fleetwood, and Landy Godbold

Overview

Welcome to 'Imperfect Testing'! We hope you and your students enjoy investigating the many important topics presented.

The results of a mammogram, like those of many tests, are not always correct. A false positive test result may create unnecessary anxiety, while a false negative test result may lead to a false sense of security. In this unit, students examine the case of an adult female who learns her mammography test is positive. They then use real data to calculate the probabilities of receiving true (or false) test results and discuss the possible implications of a positive test result, given the properties of the test. These properties, which include sensitivity and specificity, can be used to help determine the rates of incorrect test results. Students also investigate the importance of disease prevalence.

Topics

Biology: imperfect testing, genetic testing, genetic variation, pharmacogenetics, ethical choices, decision making based on data interpretation, taking perspectives, gold standards

Mathematics: probability, conditional probability, Bayes' Rule

Prerequisites

Biology: basic understanding of DNA structure and function; what a gene is and that an allele is a version of a gene

Mathematics: ratios and proportions; calculating percentage

Imperfect Testing

What do the results of an imperfect medical test actually mean?

How does one measure the effectiveness of a particular medical test?

How does one compare various available tests?

How does this information affect public policy or personal decision-making?

The results of a mammogram, like those of many tests, are not always correct. A false positive test result may create unnecessary anxiety, while a false negative test result may result in a false sense of security. In this unit, you examine the case of an adult female that learns her mammography test is positive. You use real data to calculate the probabilities of receiving true (or false) test results and discuss the possible implications of a positive test result, given the properties of the test. These properties, which include sensitivity and specificity, help determine the rates of incorrect test results. The importance of disease prevalence is also investigated.

Lesson 1 begins with a detailed biological discussion of cancer, diagnostic techniques, and a a description of mammograms as a tool to test for breast cancer. It then begins a case study that forms the foundation for the rest of the module.

Case Study: The Doctor's Office

Elizabeth Johnson is a 46-year-old female following up on a routine yearly physical exam. Read the following interactions between Elizabeth and her doctor and examine the mammogram results to answer the questions.

Elizabeth: Hi, Dr. Smith. How are you today?

Dr. Smith: I'm doing well, Elizabeth. How are the children?

Elizabeth: They are great. Kira just turned 13, Celeste is expecting with my very first grandchild, and Jennie is getting ready to go to college next year! I'm not sure how we are going to afford that, but you know how important education is.

Dr. Smith: Wow, they grow up so fast. Tell them I said hello.

Elizabeth: How were my test results?

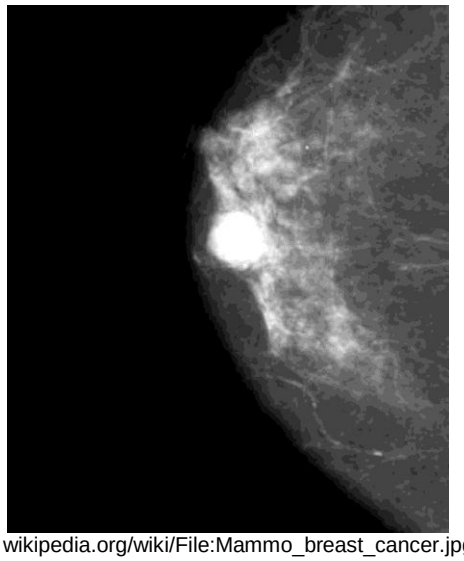
Dr. Smith: Let's take a look.

Questions for Discussion

1. What type of test is this?

2. How is this test performed?

3. Is Elizabeth's test normal or abnormal? How do you know?



wikipedia.org/wiki/File:Mammo_breast_cancer.jpg
Photo is in the public domain because it is work of the US Govenment.

Case Study: The Diagnosis

Dr. Smith: Do you see that bright white spot?

Elizabeth: Yes, is that bad?

Dr. Smith: Unfortunately, this is an abnormal **mammogram**.

Elizabeth: Oh, my, does that mean I have cancer?

Questions for Discussion

4. What does an abnormal mammogram indicate? Does Elizabeth have **cancer**?

5. What steps should be taken to confirm your answer to question 4?

Note: The module provides a picture of a normal mammogram for comparison purposes.

After this discussion (The teacher notes have hints to guide the discussion) the students participate in the following activity:

ACTIVITY

Cancer: A Dicey Situation

Objective: To simulate the steps in the process for cells to become cancerous.
Players: Groups of four students.

Materials:

- Pencil or pen
- Dice (two per group)
- Cancer FAQ
- Cancer: A Dicey Situation worksheet
- Red and black licorice sticks

Note: This activity is designed to familiarize you with the steps in cancer development. It is *not* intended to simulate the actual probability of getting cancer.

Part I

1. Each group selects two students to roll the dice. The two students rolls one die each; one to represent a proto-oncogene and the other a tumor-suppressor gene. The two other students record the results of each roll.
2. Each roller has 10 tries to roll a 1 (indicating a gene mutation). The recorders will keep track of the rolls.
    - Once a roller has rolled a 1, the gene has mutated and "waits" to see if the other gene mutates (rolls a 1).
    - If after 10 rolls both rollers do not obtain a 1, neither gene mutates.
    - If time permits, perform another set of 10 rolls.

3. Once both genes have mutated (both rollers have rolled a 1), or after 2 sets of rolls, stop and answer all of the questions in Part I (questions a-d) on the activity worksheet.

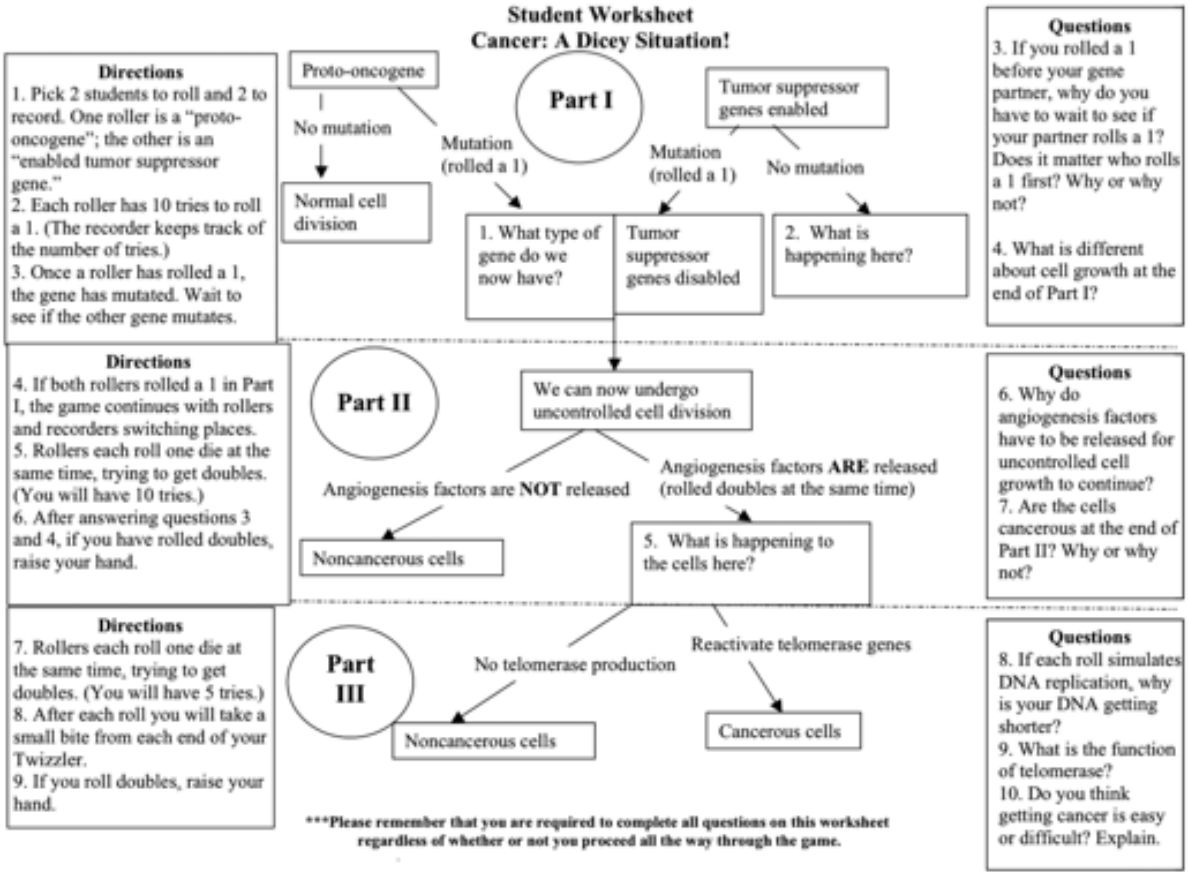
Part II

4. Continue the game with the recorders becoming rollers and the rollers becoming recorders.
5. Rollers now simultaneously roll the dice and if rolling the same number (doubles) angiogenesis factors are released. The rollers will have 10 tries to roll doubles.
    - If after 10 rolls doubles are not simultaneously obtained, the cells are cancer-free.
    - If time permits, perform another set of 10 rolls.
6. Once angiogenesis factors are released (rolled doubles), or after 2 sets of rolls, stop and answer all of the questions in Part II (questions e-g) on the activity worksheet.

- Raise your hand to have teacher check responses a-g.

Part III

7. Continue the game and pick two rollers and two recorders. Rollers will have five tries to simultaneously roll doubles.
8. Each roll simulates DNA replication and whether or not telomerase cells are reactivated.
    - Each time you do not roll doubles, eat a small piece from each end of the red licorice and place it in front of you. If after five rolls doubles are not simultaneously obtained, your red licorice is very small and the game is ended. Cells are cancer-free.
    - Each time doubles are rolled, one student puts a full-length black licorice in the center of the playing area.
    - After 5 rolls, stop and answer the final questions in Part III (h-k) on the following activity worksheet.



Lesson 2 begins after this activity and introduces the fundamentals of probability and probability trees, leading into using probability trees to answer questions in the case study in lessons 3.

Lesson 3 includes two actual data sets and asks the students to construct probability trees form these data sets. For example the following data on mammography screening for breast cancer were collected as part of a large study of mammograms.

Students are instructed to: Use these data to make a table, and then draw and fill in the two types of probability trees you've learned about.

Data Set 1

The total number of women who had a positive mammogram was 11,523.

Of the women with a positive mammogram, 831 had breast cancer.

Of the women with a positive mammogram, 10,692 did *not* have breast cancer.

The total number of women who had a negative mammogram was 88,477.

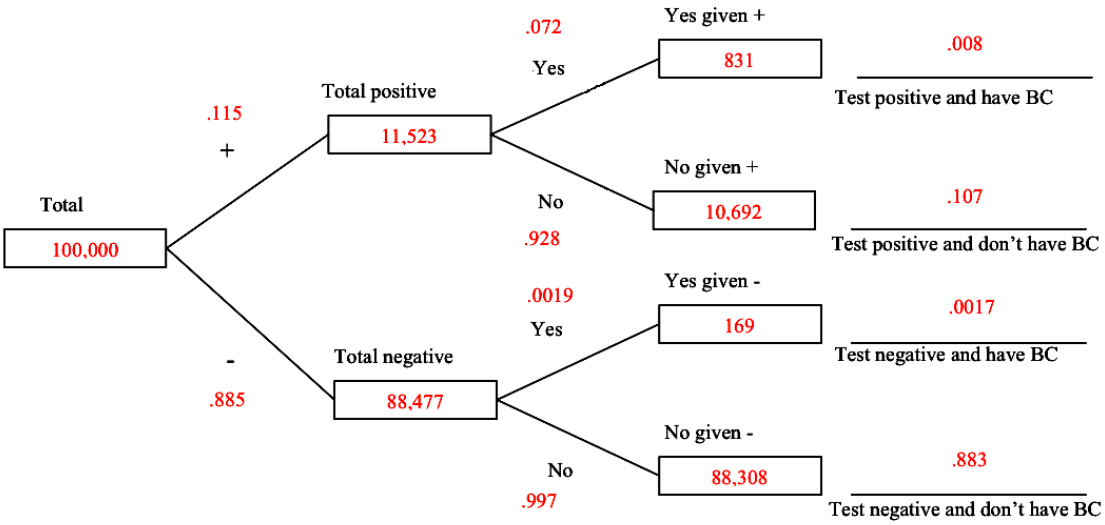
Of the women with a negative mammogram, 169 had breast cancer.

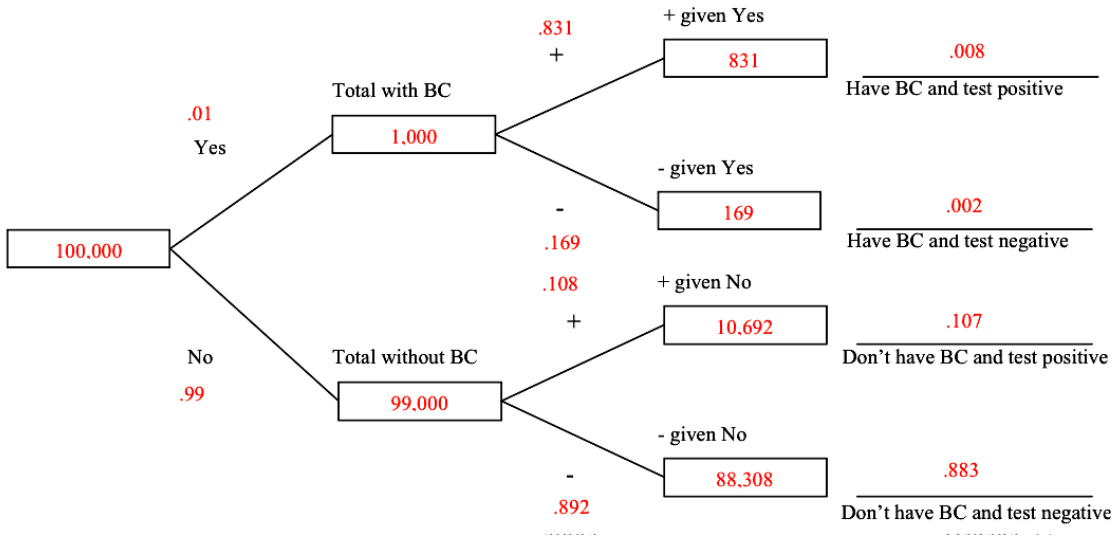
Of the women with a negative mammogram, 88,308 did *not* have breast cancer.

Adapted from real data reported in: Fenton, J.J., Taplin, S.H., Carney, P.A., Abraham, L., Sickles, E.A., D'Orsi, C., Berns, E.A., Cutter, G., Hendrick, R.E., Barlow, W.E., Elmore, J.G., "Influence of computer-aided detection on performance of screening mammography," *New England Journal of Medicine* Apr. 5, 2007, 356(14):1399–409.

Breast Cancer

		Yes	No	Total
Mammogram Test	Positive			
	Negative			
	Total			





Note: this tree is provided in the teacher version with answers.

Discussion Questions

Use your breast cancer screening data probability trees (or trees provided by your teacher) to answer the following questions:

1. What is the probability that a woman with a negative mammogram actually had breast cancer? Do you think this is low or high?
2. What is the probability that a woman with cancer had a negative mammogram? Do you think this is low or high?

3. What does it mean to have a negative mammogram?
4. What is the probability that a woman with cancer had a positive mammogram? Do you think this is low or high?
5. What is the probability that a woman with a positive mammogram actually had breast cancer? Do you think this is low or high?
6. What does it mean to have a positive mammogram?
7. If I am a patient with a positive test, which number do I care about most? Why?
8. If I think I have a disease, and am thinking about going to get tested, which number(s) do I care about most? Why?
9. What one value would you like to see improved in this mammogram?

Lesson 4 looks at the predictive value of tests, and includes a discussion of Bayes Rule, prevalence, predictive value, specificity and sensitivity. It also includes a Taste Test Activity and an Imperfect Search Activity, and even calculator and computer programs.

Now let's look at another module on a vastly different topic: *Genetic Inversions.* The student version begins quite differently from Imperfect Testing. We provide just the first few pages.

Genetic Inversion

by Tom Fleetwood, Paul Kehle, and Celeste Young

Molecular Basis of Heredity

- In all organisms, the instructions for specifying the characteristics of the organism are carried in DNA, a large polymer formed from subunits of four kinds (A, G, C, and T). The chemical and structural properties of DNA explain how the genetic information that underlies heredity is both encoded in genes (as a string of molecular "letters") and replicated (by a templating mechanism). Each DNA molecule in a cell forms a single chromosome.

- Most of the cells in a human contain two copies of each of 22 different chromosomes. In addition, there is a pair of chromosomes that determines sex: a female contains two X chromosomes and a male contains one X and one Y chromosome. Transmission of genetic information to offspring occurs through egg and sperm cells that contain only one representative from each chromosome pair. An egg and a sperm unite to form a new individual. The fact that the human body is formed from cells that contain two copies of each chromosome—and therefore two copies of each gene—explains many features of human heredity, such as how variations that are hidden in one generation can be expressed in the next.
- Changes in DNA (mutations) occur spontaneously at low rates. Some of these changes make no difference to the organism, whereas others can change cells and organisms. Only mutations in germ cells can create the variation that changes an organism's offspring.

You are a Biomath graduate student conducting research in the Amazon Rainforest. One day you are lucky enough to stumble upon a previously undiscovered creature. To name this creature (perhaps after yourself!) it is important to determine its place on the evolutionary tree of life. After sequencing a section of chromosome 6 you use a computer to look for matches to any known species. The computer finds that the genes from chromosome 6 correspond to the genes in the frillneck lizard (a currently living species) but they are not in the same order. Your task now is to create a phylogenetic tree to show the relationship between the newly discovered creature and the frillneck lizard for your local natural history museum. The museum wants a display of your findings using model creatures complete with a drawing of the most likely phylogenetic tree. The order of genes will be given to you later.

Figure 1. Frillneck Lizard. Public Domain photography by Miklos Schiberna

You might first be thinking, "What is a phylogenetic tree?" and then be trying to figure out how you would create one. It is not a probability tree just in case you did the module *Imperfect Testing* first. In this unit you learn about sequencing sections of chromosomes and analyzing inversions of these sequences. You will discover how to create a phylogenetic tree.

In this module, we will work on the problem of determining how closely related two different **species** of animals are. Because the genes located on the chromosomes of animals can be represented mathematically as **sequences** of numbers, we will focus on the related question of how closely related two numerical sequences are. Consider the three pairs of sequences shown below.

132456	321564	346521
123456	123456	123456

Many people might say that the pair of sequences of numbers in the left column above are much more similar than the pair of sequences in the right column above. How can we measure how closely related two given sequences are? What is the significance of this question for biology and for mathematics? These are the questions we explore in this module.

One way to determine how closely related two sequences are is to use the method of subsequence inversions. A **subsequence** is a sequence of two or more adjacent items that are usually a smaller portion of a larger sequence; however, sometimes the entire sequence might be considered a subsequence of itself. An **inversion** is the reversing of a subsequence within a sequence, or the reversing of the entire sequence.

We count the number of subsequence inversions required to change the **initial sequence** into the **target sequence**.

> Example 1: Start with the initial sequence 1 3 2 4 5 6. Invert the subsequence 3 2 to 2 3 to reach the target sequence 1 2 3 4 5 6. Just one inversion!

> Example 2: Now start with 3 2 1 5 6 4. We could invert the subsequence 3 2 1 to produce the new sequence 1 2 3 5 6 4. Next, we could invert the subsequence 5 6 to produce the new sequence 1 2 3 6 5 4. Finally, we could invert the subsequence 6 5 4 to produce the identity (or normal) sequence 1 2 3 4 5 6. We used three inversions to change 3 2 1 5 6 4 into 1 2 3 4 5 6.

ACTIVITY A

Playing Card Inversions

Objective: Use the minimum numbers of subsequence inversions to change the initial sequence of cards to the identity (normal) sequence.

Players: Groups of two students

Materials:

- Numbered index cards (from 1 to 10) or playing cards of one suit (from Ace to 10)
- Score Table

Round 1. Use only the cards from 1 to 6 (or Ace to 6). Shuffle the cards and have each player choose one card without looking. The player choosing the higher number (Player A) shuffles the cards and randomly places them in a sequence. The other player (Player B) must use subsequence inversions to change the random arrangement of cards into the identity (or normal) ordering of 1 2 3 4 5 6. Keep track of the inversions in the score table. The number of subsequence inversions used by Player B is Player B's score for the first round. Then Player B shuffles the cards and gives Player A a random sequence of the six cards. Player A changes the sequence into the identity ordering by using sequential inversions. The number of inversions used is Player A's score. After each player has had three sequences, add up the scores. The lowest score wins.

Round 2. Now play a few more times but with sequences of different lengths (choose lengths between 3 and 10).

Draw a score table as shown below to keep track of the lengths of your sequences and the numbers of inversions you use. Pay attention to the least and greatest numbers of inversions different pairs of sequences need to transform one sequence of the pair into the other sequence of the pair.

Length	Initial Sequence	Inversions	Target Sequence	# of Inversions

SCORE: _____

Genetic Inversions provides another activity that is actually an applet and then reminds the student of biology background for the module on genomes, cells, DNA, chromosomes, proteins, genes, centromere, and mutations.

Mutations can occur that change the order of the genes and the centromere along a chromosome. These "chromosomal mutations" include insertions, duplications, deletions, translocations and inversions. This module is only interested in inversions. An inversion occurs when a single gene or a group of genes detach from the DNA strand, rotate 180°, and reattach to the strand. The remainder of the module taks about inversions and evolution, algorithms , especially an inversion algorithm, and phylogenetic trees. Numerous activities, handouts, transparencies, and a final case study are included.

Lesson 5 studies Pharmacogenetics and relates it to the case study of Elizabeth and utilizes probability trees to help determine the effectiveness of various treatments. Lesson 6 follows up with a discussion of Ethics and Decision-Making, which concludes the case study.

Two types of assessments are included at the end of the module, one a short answer test and another project related, and a Glossary of terms.