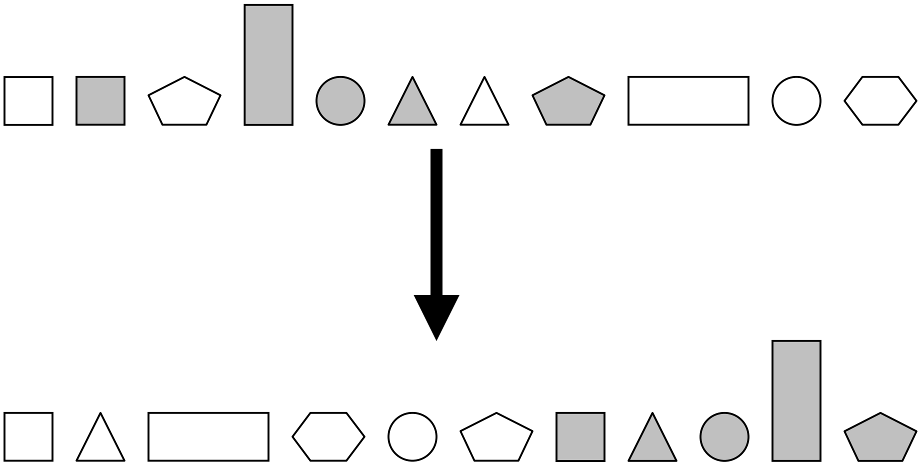
Genetic Inversion Rearrangement:

Are you a rotated mouse?



BMCI '05

Kathy Erickson
Great Barrington, MA
spence417@aol.com

Tom Fleetwood
Wilmington, DE
tfleetwood@charterschool.org

Patrick Flynn
Kansas City, KS
flynnp@turnerusd202.org

Brian VanGorden
Port Allegany, PA
bvangorden@verizon.net

TEACHER MATERIAL

Overview for Teacher

Topics explored in module: genetic mutations, development and application of an algorithm and optimization.

Target audience: 9th or 10th grade mathematics or biology class or the two combined

Length: 4 - 5 class periods of 45 minutes each

Standards Alignment: Topics in this module align with the following:

- NCTM Standards for School Mathematics.
  - Problem Solving
  - Reasoning and Proof
  - Communication
  - Connections
  - Representation

- National Science Education Standards
  - Biological Evolution
  - Molecular Basis of Heredity
  - Investigating and Analyzing Science Questions
  - Applying the results of experiments to scientific argument and explanations.

Goal: The student will understand the role of chromosome inversion mutations in evolution.

- Objective 1: The student will be able to define a chromosome inversion mutation.
- Objective 2: The student will be able to diagram an inversion event.
- Objective 3: The student will be able to list the 3 possible outcomes of an inversion event.
- Objective 4: The student will be able to explain why each outcome could arise.
- Objective 5: The student will be able to calculate the number of inversions necessary to reach a target sequence.
- Objective 6: The student will be able to calculate the most likely phylogeny of a set of organisms using genome inversion data.

Goal: The student will understand algorithmic thinking.

- Objective 1: The student will be able to develop an algorithm to solve inversion problems.
- Objective 2: The student will be able to apply a given algorithm to inversion problems.
- Objective 3: The student will be able to analyze and compare algorithms for efficiency.

Synopsis of unit: Students will investigate the basic concepts of DNA and mutations and play a game rearranging sequences by inverting subsequences. After becoming familiar with the process, they will be challenged to develop and write an algorithm that can be applied to any sequence. A second algorithm will be introduced and analyzed. After mastering the algorithm, students will apply the ideas to the concept of gene mutation.

Teacher Timeline

Day	Topics	Materials	Homework
1	• Reverso • Chromosomes • Genes • Types of gene mutations	• *Reverso* game or playing cards • Handout - *Reverso* instructions and Biology background questions • Handout - *Biology Background* • Computers with Internet connection for online tutorials	• Handout – Human evolutionary tree and questions • Inversion homework activity
2	• Possible mutation outcomes • Inversions and evolution • Algorithms – following and writing an algorithm	• Transparency - *Joint genome Institute: Targets and Associated Diseases* • *Origami algorithm* and paper cut into circle • *Creating an algorithm* cutouts • *Biology Background 2*	• *Write an algorithm* • *Optional Sequence Worksheet*
3	• Algorithms – analyzing and optimization; inversion algorithm • Upper and lower bounds	• Handout – *Optimization and a new algorithm*	• *New Algorithm Homework*
4	• Algorithms – analyzing and optimization; inversion algorithm • Upper and lower bounds • Inversions and evolution	• Handout – *Day 4 handout*	• *None*
5	• Phylogenetic Trees	• Handout – *Phylogenetic Tree Creations*	•

Day 1

Reverso is a computer game designed to prepare your students for the concept of inversion in gene mutations. The basic goal of the game is to have students transform one sequence of colors into another sequence using a series of subsequence inversions. The game is distributed free of charge, and you are welcome to make modifications and improvements. If possible, install the game on the computers the day before the class meets.

Reverso was written using the LOGO computer language. A free version of MSWLogo can be downloaded from http://www.softronix.com/logo.html. The game file can be downloaded from http://dimacs.rutgers.edu/bmci_web/reverso.htm. After installing MSWLogo, you must extract the *Reverso* game file.

It is very beneficial to have students play Reverso earlier in the year to build up the knowledge base.

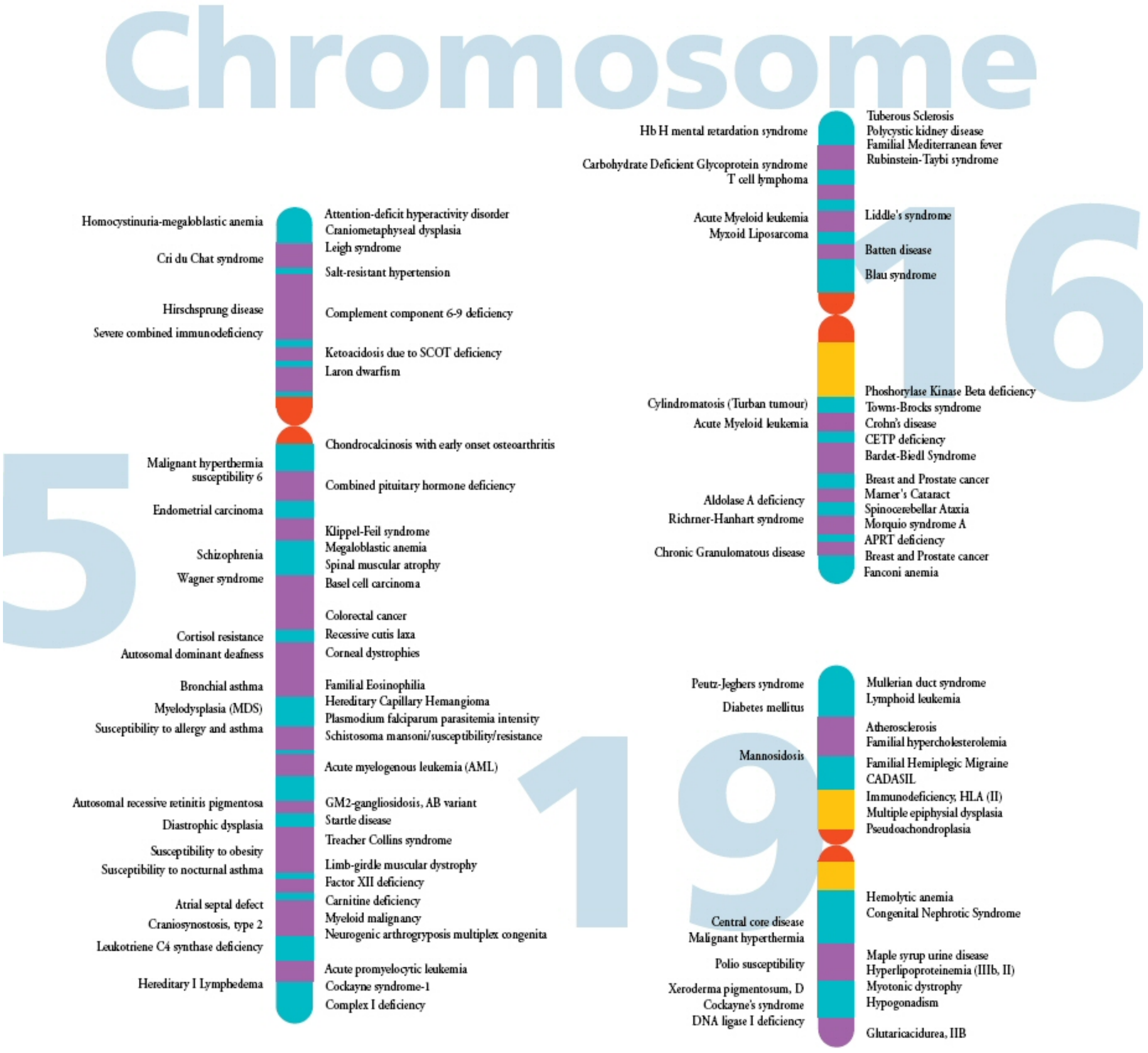
If computers are not available, an alternative form of the game can be played using playing cards. Lay out the Ace through 6 of one suit in a horizontal row, in random order. The goal is to use *inversions* to restore the sequence to the identity – Ace through six.

An *inversion* is the reversing of a sequence of adjacent cards. For example, inverting the sequence 4 5 2 within A 4 5 2 6 3 yields A 2 5 4 6 3.

Students will use written or internet resources to answer the five background Biology questions. They may wish to consult a Biology teacher or textbook, if necessary.

The background homework is designed to bring out students' previous knowledge. Discussion on day 2 should help to clarify misinformation and fill in gaps in understanding.

Sequencing Targets and Associated Diseases



JGI JOINT GENOME INSTITUTE

http://www.ornl.gov/sci/techresources/Human_Genome/graphics/chr51619.pdf

Day 2

Tracking chromosomal inversions is important biologically because it can help determine evolutionary phylogeny (the evolutionary development and history of a species). Mutations are the driving force of evolution and inversions are one form of these mutations. By determining the most parsimonious sequence of inversions to lead from the given sequence back to the identity sequence one can infer the minimum number of evolutionary steps between two organisms and the evolutionary order of a group of organisms with those inversions. For example, given the identity sequence 12345 and the species sequence 54321, it can be seen that a single inversion can separate one organism from the other. Therefore, it would be deduced the two species are closely related evolutionarily. On the other hand, 52314 would require three inversions to obtain the identity sequence and is therefore a more distant relative than the single inversion organism. One inversion sequence to obtain this is: 52314 – 13254 – 12354 – 12345.

Additionally, a chromosomal inversion can actually break in the middle of a gene and rotate, thus rendering that gene inoperable. For this module it is assumed that the chromosomal breaks do not break within a gene but rather shift genes in their entirety.

Omission Activity

This activity may be easier for students if you put the three strips onto the overhead. Have the students rotate individual pieces until the word OMISSION is formed.

I	W	O

S	S	I

O	N

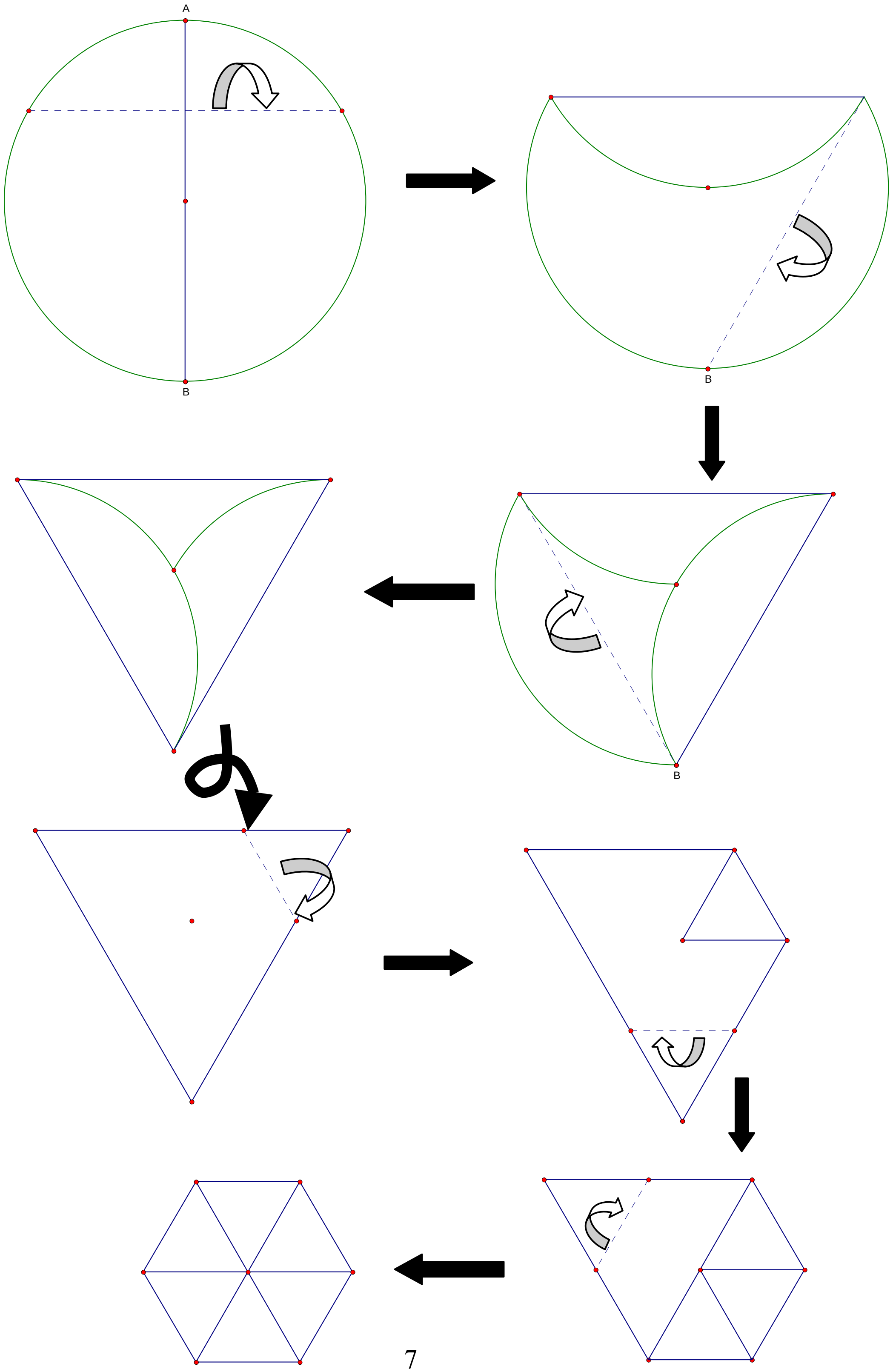
This was not too difficult because the word was only broken into three pieces and the places where the breaks occurred, known as **break points**, were already given. But, what if the break points were not known? Or better yet, could there be more than two break points?

By playing reverso, students should convince themselves that every sequence can be transformed into the identity sequence. One easy way is to invert the sequence that places the 1 in the first place, then invert the sequence that places the 2 in the second position, and so on. This method is easy to understand and shows that the upper bound is n-1 inversions for a string of length n. This is the algorithm students will write for homework. They will develop this upper bound on day 3.

Algorithms

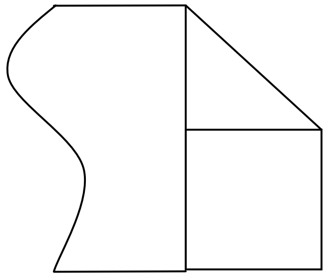
The study of mathematics and science involve a variety of algorithms. Suggest examples such as long division, adding fractions and following a lab procedure. Ask for other examples of algorithms they have used in school.

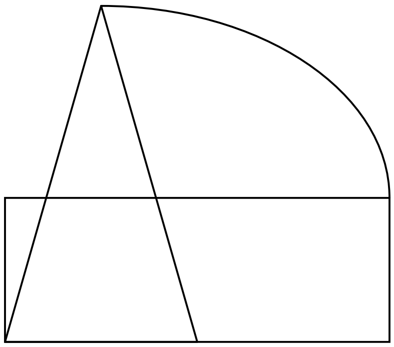
Origami Algorithm
Students practice following an algorithm with a shape that transforms from a circle to an equilateral triangle to its end shape of a regular hexagon.   You will definitely want to practice this before class. A set of diagrams is included as a guide.

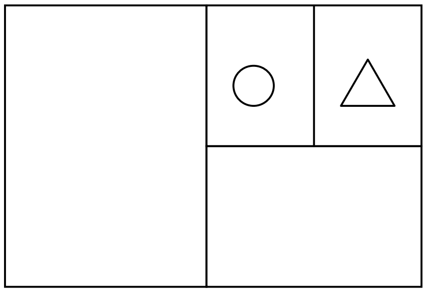


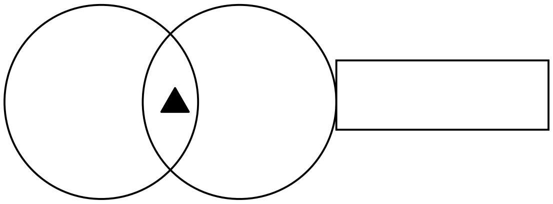
Creating an algorithm activity

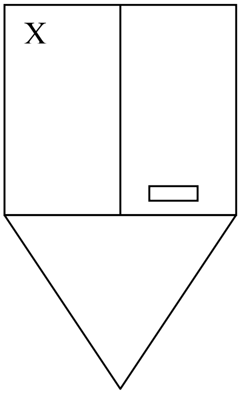
The students will now work in pairs to develop their own algorithms. Give one student in each pair one of the diagrams below. The student with the diagram tells the second student how to recreate the picture. The second student may not ask questions nor see the picture. The directions the first student gives constitute an algorithm. Have students switch roles with a different diagram. Useful discussion may arise when students with the same diagram share challenges and successes. This activity prepares the students to write an algorithm about inversions.

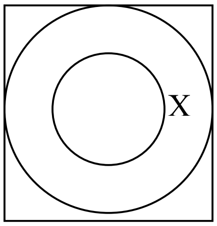












Day 3

The topic of optimization is a recurring theme in a mathematics curriculum. Application problems in Algebra and Calculus may involve minimizing the time to complete a task or maximizing the amount of money earned. This is a good place to give students a preview of topics they will study later.

One easy way is to invert the sequence that places the 1 in the first place, then invert the sequence that places the 2 in the second position, and so on. This method is easy to understand and shows that the upper bound is n-1 inversions for a string of length n.

The students will likely need a good portion of class time to practice and understand the new algorithm.

Day 4

A good deal of class time may be needed to review the homework.

The lower bound (also called the inversion number) gives the smallest possible number of inversions needed to start with one sequence and end with the other. This can be referred to as the evolutionary distance between the species, giving an approximation of the difference between the two.

|3 2 | 5 6 7 | 4 | 1 | 8
lower bound = 3

|3 2 | 5 | 7 | 4 | 6 | 8 | 1 |
lower bound = 4

| 4 5 | 2 1 | 6 | 3 | 8 7 | 9
lower bound = 3

|3 2 1 | 5 4 | 8 9 | 7 | 10 | 12 | 6 | 11 |
lower bound = 5

The specimen with the cabbage is turnip.

Genome Rearrangements In Man and Mouse

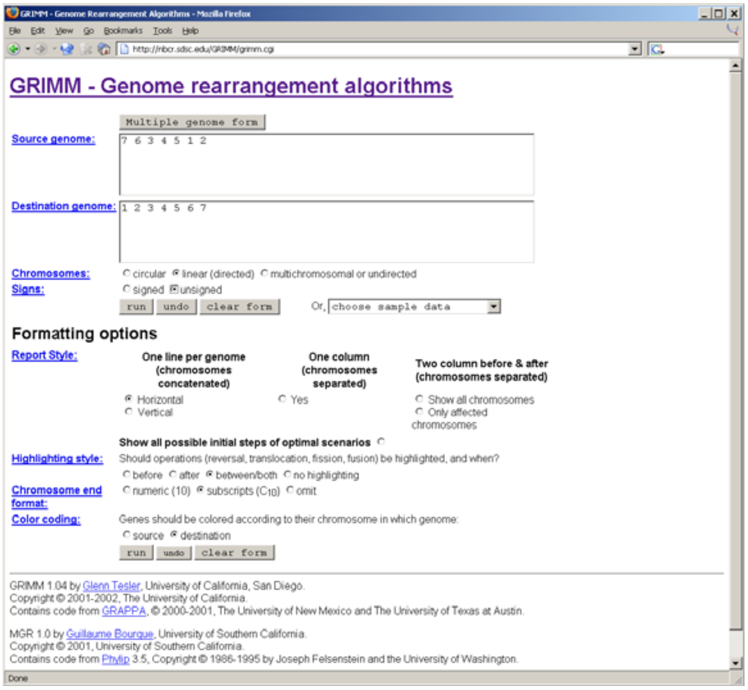
http://nbcr.sdsc.edu/GRIMM/grimm.cgi

GRIMM is a powerful tool for analyzing rearrangements in pairs of genomes. It can be used to analyze reversal scenarios between pairs of genomes. To get started, open a web browser and navigate to the link above. Since the program itself is very powerful, the interface may seem fairly cluttered. However, this lesson will only be using a few features for the purpose of this module.

Complete the following steps to determine a reversal scenario for the sequence 7 6 3 4 5 1 2.

1) Type the sequence 7 6 3 4 5 1 2 into the *Source Genome* box.

2) Type the sequence 1 2 3 4 5 6 7 into the *Destination Genome* box.
3) Click the *Linear (directed)* radio button.
4) Click the *unsigned* radio button.
5) Click *Run*.



GRIMM - Genome rearrangement algorithms

Source genome: 7 6 3 4 5 1 2

Destination genome: 1 2 3 4 5 6 7

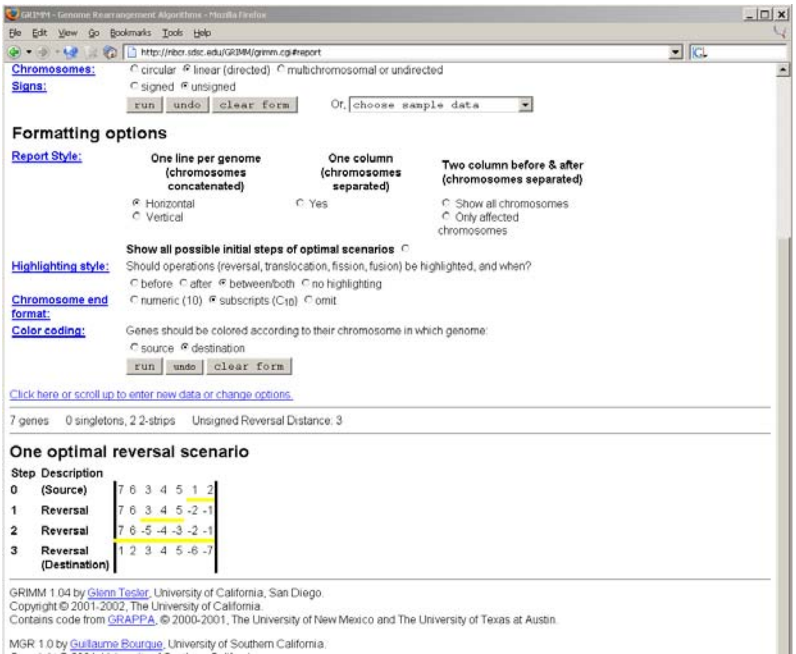
Chromosomes: ○ circular ● linear (directed) ○ multichromosomal or undirected
Signs: ○ signed ● unsigned
Or, choose sample data

Formatting options
Report Style: One line per genome (chromosomes concatenated): ● Horizontal ○ Vertical; One column (chromosomes separated): ○ Yes; Two column before & after (chromosomes separated): ○ Show all chromosomes ○ Only affected chromosomes
Show all possible initial steps of optimal scenarios ○
Highlighting style: Should operations (reversal, translocation, fission, fusion) be highlighted, and when? ○ before ○ after ● between/both ○ no highlighting
Chromosome end format: ○ numeric (10) ● subscripts (C_{10}) ○ omit
Color coding: Genes should be colored according to their chromosome in which genome: ○ source ● destination

GRIMM 1.04 by Glenn Tesler, University of California, San Diego.
Copyright © 2001-2002, The University of California.
Contains code from GRAPPA, © 2000-2001, The University of New Mexico and The University of Texas at Austin.

MGR 1.0 by Guillaume Bourque, University of Southern California.
Copyright © 2001, University of Southern California.
Contains code from Phylip 3.5, Copyright © 1986-1995 by Joseph Felsenstein and the University of Washington.

After running the program, the screen below should appear. The area o f interest is the matrix at the bottom. The rows in the matrix represent intermediate steps in the evolution from the source sequence to the destination sequence. The underlined strip of numbers represent the portion of the sequence that was inverted in each step. For example, in the first step of the scenario below, the strip 1 2 was inverted. Notice that after the inversion these values became –2 –1. The negative notation indicates that the orientation of each gene has been reversed. For this module, any changes in orientation of individual genes are not considered. Therefore, disregard the sign of any number in the program output. Note that the given scenario uses three reversals. This is only one optimal scenario. There may be others.



Chromosomes: ○ circular ● linear (directed) ○ multichromosomal or undirected
Signs: ○ signed ● unsigned
Or, choose sample data

Formatting options
Report Style: One line per genome (chromosomes concatenated): ● Horizontal ○ Vertical; One column (chromosomes separated): ○ Yes; Two column before & after (chromosomes separated): ○ Show all chromosomes ○ Only affected chromosomes
Show all possible initial steps of optimal scenarios ○
Highlighting style: Should operations (reversal, translocation, fission, fusion) be highlighted, and when? ○ before ○ after ● between/both ○ no highlighting
Chromosome end format: ○ numeric (10) ● subscripts (C_{10}) ○ omit
Color coding: Genes should be colored according to their chromosome in which genome: ○ source ● destination

Click here or scroll up to enter new data or change options.

```
7 genes     0 singletons, 2 2-strips     Unsigned Reversal Distance: 3

One optimal reversal scenario
Step Description
0    (Source)       7 6  3  4  5  1  2
1    Reversal       7 6  3  4  5 -2 -1
2    Reversal       7 6 -5 -4 -3 -2 -1
3    Reversal       1 2  3  4  5 -6 -7
     (Destination)
```

GRIMM 1.04 by Glenn Tesler, University of California, San Diego.
Copyright © 2001-2002, The University of California.
Contains code from GRAPPA, © 2000-2001, The University of New Mexico and The University of Texas at Austin.

MGR 1.0 by Guillaume Bourque, University of Southern California.

Day 5

The assessment is designed as a group activity. Groups should complete the question about the phylogenetic tree and you can choose a few of the theory questions for them to answer.

STUDENT MATERIAL

Day 1

REVERSO

The exciting game of color inversion!

Reverso was written using the LOGO computer language. A free version of MSWLogo and the game can be downloaded from http://www.softronix.com/logo.html. After installing MSWLogo, you must extract the *Reverso* game file.

Instructions: To run the game, open MSWLogo and choose "Load…" from the "File" menu. Select "reverso2.lgo" and click "Open". At this point you may wish to minimize the *Commander* screen. When the colored circles stop flashing, press enter to start the game. Input the number of colors you wish to use in the string. Any number of colors up to 13 may be used. A sequence of colors will be displayed representing the *identity sequence*. A second sequence of colors will be displayed which must be transformed into the identity sequence. you will input 2 numbers separated by a comma. These numbers represent the *breakpoints* between which the subsequence will be inverted. These breakpoints are numbered 1, 2, 3… where 1 is located before the first circle, and the final number is located after the last circle. For example, entering 2, 5 inverts the second, third and forth. While playing the game, you can scroll up or down by pressing "u" or "d" and then "enter" in the *Sequence Cuts* window. When a player wins the game a message is displayed. To play again, simply restore the *Commander* window and type "reverso" in the white area at the bottom, then press "enter".

Biology background

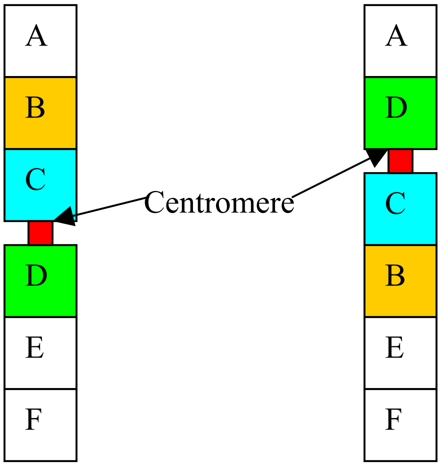
Answer the following by using one or more of the references listed below.

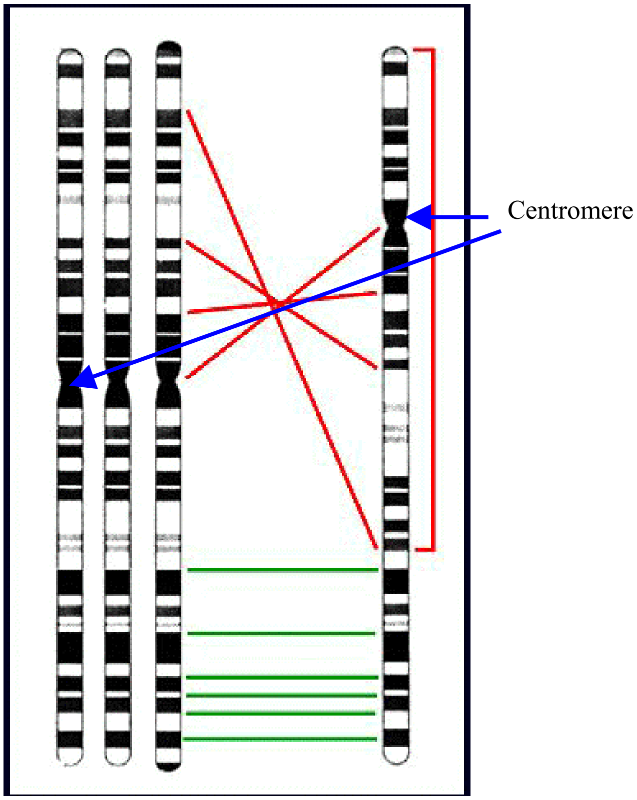
1. What is a gene? What does a gene do?
2. What is the term used when a gene mutates by rotating 180°?
3. What are the three possible outcomes of a gene mutation?
4. Why is a mutation in a gamete more significant than a mutation in any other type of cell?
5. Explain the relation between centromere location and evolution.

Biology background handout
http://www.dnai.org/c/index.html
http://cstaff.hinsdale86.org/~kgabric/DIMACS/DimacsIntro.htm

Biology Background Handout

Cells are controlled by the genetic information contained within its DNA. DNA is coiled and packed into a structure called a chromosome. A segment of the DNA/chromosome that codes for a particular protein is a gene. The chromosome has many genes lined up in a particular order. Another piece of the chromosome is called the centromere. Mutations can occur that change the order of the genes and the centromere along a chromosome. These "chromosomal mutations" include insertions, 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.



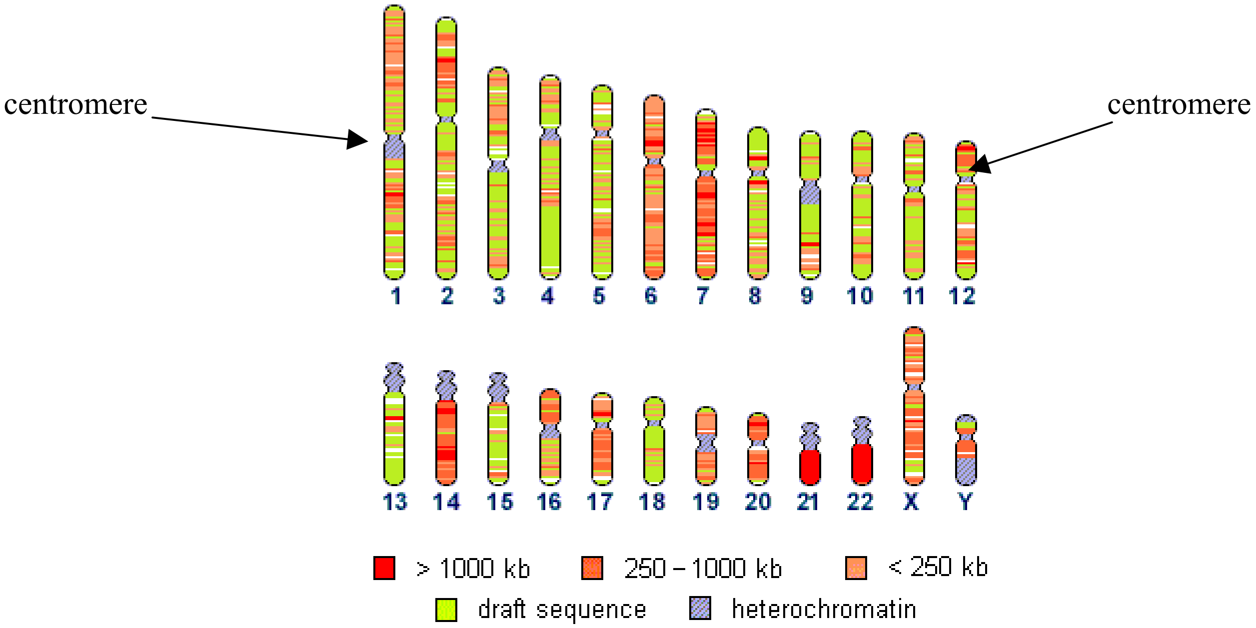


http://www.msu.edu/~lattasch/inversion.html

The pictures above show a section of the chromosome, including the centromere that has been inverted, thus moving the centromere toward the top of the chromosome. Any genes that were within this section of the chromosome have now been moved to a new position with the exception of the potential gene at the 'pivot point.' Illustrated a different way with numbers, let us start with the gene sequence 12345. If we invert the strip '234' of the chromosome then we have the new alignment 14325. If the centromere was within the 234 segment then it will move positions, otherwise no change in centromere location will be noted. Note that genes outside the inversion area remain unchanged in their global location on the chromosome but may now be next to a gene they were not adjacent to before. In the above example, gene 1 started next to gene 2. After the inversion of 234, gene 1 is now adjacent to gene 4, even though gene 1 was not involved in the inversion.

The change in location of the centromere can indicate that an inversion has occurred. If you know the original chromosome has the centromere in the center and, after DNA replication, a copy has the centromere in a different location you can deduce that an inversion event has occurred. Again, this can be seen in the picture above.

The picture below shows different human chromosomes – note the centromere location is quite different in any given chromosome. The fact that it is not in the center is suggestive that inversions occurred throughout evolution to shift its position, and therefore the relative gene positions, which led to the human species.



http://www.ncbi.nlm.nih.gov/Class/MLACourse/Original8Hour/Genetics/chromosome_ideograms.gif

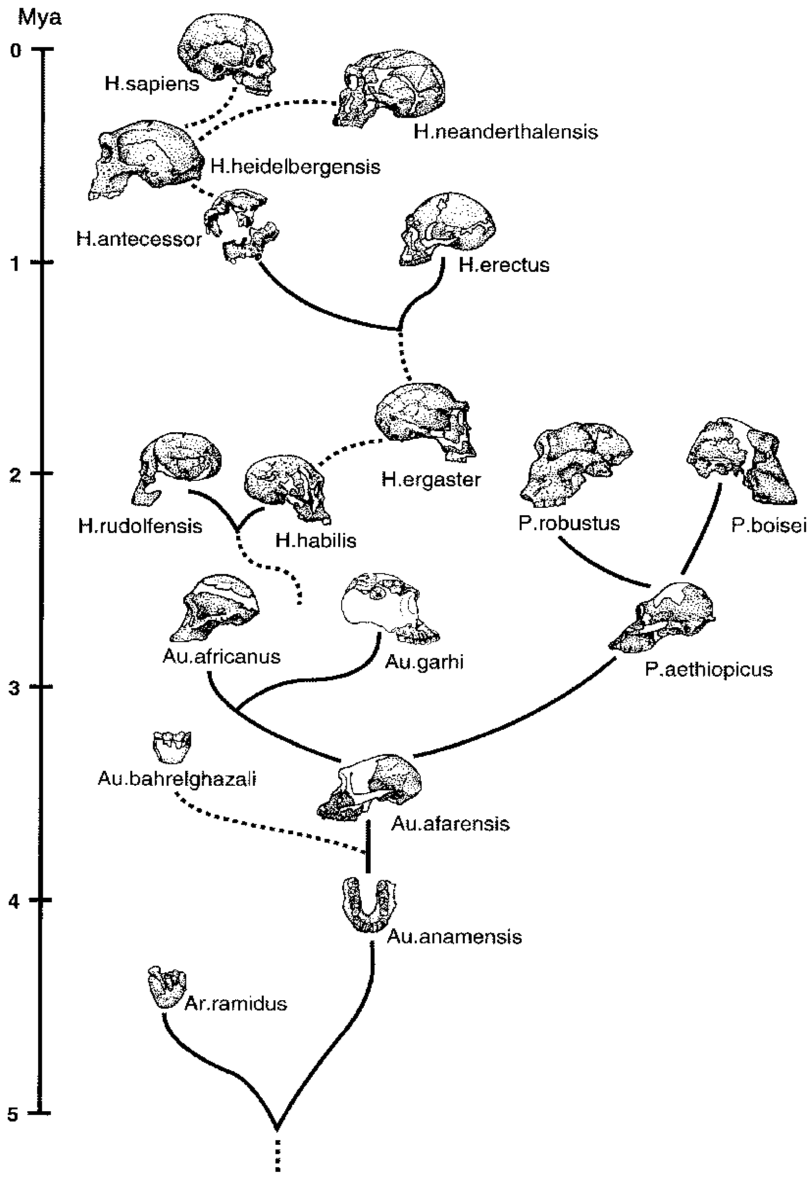
Chromosomal inversions can occur in any cell in the human body (or in any living creature) but the most important events occur in sex cells (sperm or egg cells). These changes exhibit their full effect on any offspring formed from these inverted sperm or egg chromosomes. In other words, an inversion that occurs in a skin cell only effects that skin cell but the other trillion cells in your body remain normal. However, if the inversion occurs in a sperm or egg cell, when the cells combine to form a single cell zygote every new cell created in that offspring will contain that inverted sequence.

Chromosomal inversion events can result in 3 basic outcomes for the offspring:

1. Advantageous - perhaps due to the activation of a gene that was previously inhibited due to its position on the chromosome or deactivation of a gene that was previously active.
   i.e. In an arctic environment, the inversion deactivates a gene that codes for black fur causing the animal to have white fur thus blending better with its environment.
2. Disadvantageous – same as above but either the environment was different so the color change was not preferred or the inactivated gene coded for a protein that is essential for the survival of the organism. Without the essential protein the animal will not survive. Therefore disadvantageous changes can result in an offspring that does not survive or an offspring with less chance to survive in that particular environment.
   i.e. hemophilia, a disease where an individual continues to bleed when injured, can be caused by genetic inversions on the x-chromosome
3. No effect – this could occur for 2 reasons
   a. The activation or deactivation of a gene was not important to the organism's survival.
   b. The inversion caused no change in the gene activity – the genes are simply in a new location but work exactly the same as before the inversion.

Refer to the included ***Joint Genome Institute: Sequencing Targets and Associated Diseases*** to see a list of diseases that have been mapped to mutations of chromosome 5, 16 and 19. All of these diseases are not necessarily caused by an inversion mutation but, nonetheless, the image is an excellent visual to demonstrate the importance of mutations and genetic study.

Background Homework

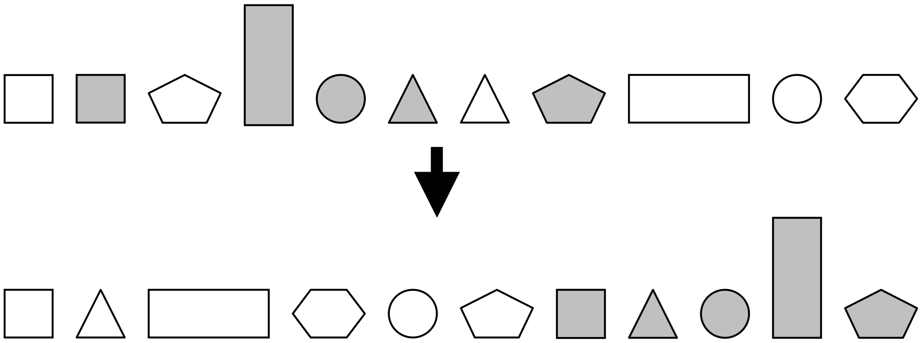


http://www.theosophy-nw.org/theosnw/evol/s3amib2.gif

1. **What do we call the type of diagram above?**
2. **Which two are more closely related, H. rudolfensis and H. erectus or P. robustus and P. boisei**
3. **Who is the direct ancestor of P. boisei? Au. Garhi?**
4. **What does it mean if a species has no descendants?**
5. **How do you think they made decisions regarding where to place species on this tree?**
6. **What other ways could they have compared these species to place the on the tree? What way do you think would give the most accurate answer?**

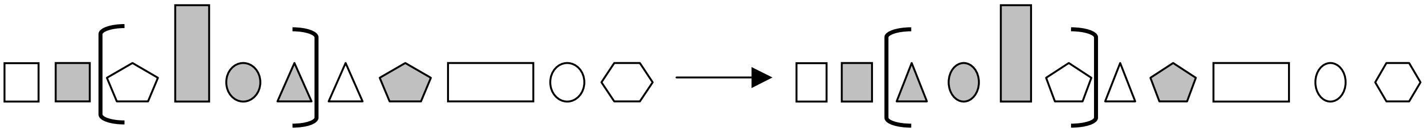
Inversion Homework Activity

Directions: Convert the top into the shown identity sequence using only inversions.



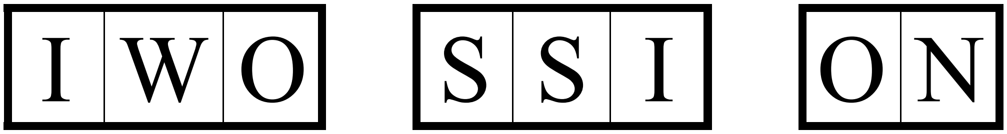
Identity sequence

For example, your first step <u>could</u> be….

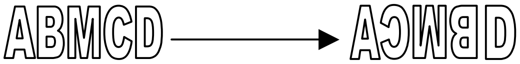


Day 2

The three strips of letters shown below can be reassembled to form a word. Strips can be inverted, like the chromosomes, but not reordered. Can you find the missing word?



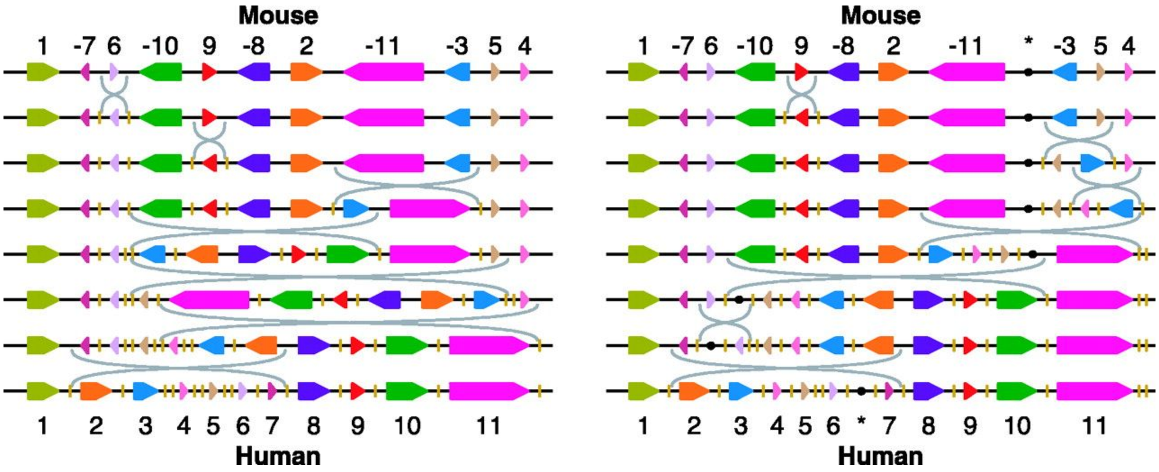
When genes are actually inverted in an organism, they not only change position but also read 'backwards' within the chromosome. This is best illustrated if one gene is inverted. For example, in the sequence ABMCD, if the M is removed, rotated 180° and reinserted in the same spot it would actually lead to ABWCD. Note the M is now 'upside down' to become the W. In a longer inversion it would look like the following where BMC is inverted.



In mathematical notation the change above would be represented as

1 2 3 4 5 → 1 -4 -3 -2 5

Note the negative signs denote a gene that is now positioned 'backward.' This can also be seen in the actual version of the mouse to human rearrangement sequence below.



http://www.pnas.org/content/vol100/issue13/images/large/pq1330369002.jpeg

Note the directional changes of certain genes denoted by the negative sign (i.e. -11)

This diagram shows two ways a mouse chromosome can be converted to a human chromosome in 7 steps by simply using inversions. Again, this module's version of the mouse to human conversion disregards these reversals of direction and simply concentrates on the sequence of numbers. In reality, a sequence of 1 2 3 4 -5 is not the same as 1 2 3 4 5. It will require one more inversion of the -5 to obtain the +5 direction. Taking both sequence and direction into account is an excellent extension exercise and tie-in to graph theory.

Did you have any questions on the background homework?

Explain how the inversion homework activity and the game Reverso both relate to the biology information.

Can every sequence be transformed into the identity sequence using only inversions? Develop a convincing argument why the answer is yes or show that the answer is no by giving a counterexample.

Definition An algorithm is a step-by-step set of rules used to solve a given problem in a finite number of steps. Following an algorithm is often more successful and less frustrating than using trial and error. Practice following an algorithm with the origami activity below.

Origami algorithm

1. If not already done, cut circle out of paper.
2. Draw a diameter of the circle and label the endpoints A and B.
3. Fold the paper so that point A touches the center of the circle.

4. Fold the paper along the chord from point B to the point where the previous fold meets the circumference (on either side).
5. Repeat step 4 from point B to the point where the previous fold meets the circumference on the other side.
6. Fold one vertex of the triangle to what was the center of the circle.
7. Repeat step 6 for the other two vertices of the triangle.

Writing an algorithm

Now that you have followed an algorithm, you will create one. One person from each pair will be given a diagram that other will be asked to reproduce. The person doing to drawing cannot look at the diagram and cannot ask questions. The person with the diagram will give a set of directions aloud. This set of directions is the algorithm.

Switch roles with a new diagram.

For homework tonight, write an algorithm to transform any sequence into its identity sequence.

If additional practice is needed, use Optional Sequence Worksheet.

Day 2 Homework: Write an algorithm to transform any sequence into its identity sequence.

Algorithm

In class follow-up

Feedback from peer

Two positive aspects of the algorithm:

Two suggestions for change:

Optional Sequence Worksheet

Sort the following sequences into the identity sequence 1 2 3 4 5 6

2	1	4	3	6	5		6	5	3	2	1	4		6	4	1	2	3	5
5	3	6	1	4	2		1	6	2	5	4	3		4	3	2	6	5	1

Day 3

Peer edit of algorithms

Switch algorithms with a classmate and comment on at least two things you like about the algorithm and give at least two suggestions.

To make the exploration of inversions easier, we will use sequences of numbers instead of shapes or colors. As an example, transform 4 3 1 6 5 2 into the identity sequence 1 2 3 4 5 6. How many inversions were needed? How does this math question actually represent a biology question?

One of the mathematical questions related to this is "What is the smallest number of inversions needed?" This is an example of an optimization question. One way to explore optimization is to examine upper and lower bounds, the worst case and best case scenarios. The upper bound for this problem is the maximum number of inversions necessary to transform the sequence to its identity.

What is the maximum number of inversions necessary to transform a sequence that has:

a. 4 numbers?
b. 5 numbers?
c. 15 numbers?
d. n numbers?

The last value you found is the upper bound for a sequence of length n. Finding a lower bound, however, is more challenging. Using a new algorithm for transforming the sequence will give some insight into the question.

A new algorithm

The new algorithm requires familiarity with the following vocabulary.

A **breakpoint** of a sequence of ***n*** numbers occurs at the following places:

- before the first number of a sequence, unless the first number is 1
- after the last number of a sequence, unless the last number is *n*
- between nonconsecutive numbers.

The sequence 3 4 8 7 6 1 2 5 has five breakpoints and the sequence 3 4 7 6 5 1 2 8 has four breakpoints. Draw them in below.

3 4 8 7 6 1 2 5 3 4 7 6 5 1 2 8

Breakpoints separate a sequence into subsequences called **strips**. A strip is labeled **increasing** if it contains numbers that increase by 1 when read from left to right (example: 3 4 5). A strip is labeled **decreasing** if it contains numbers that decrease by 1 (example: 7 6 5 4). A strip with exactly one number is considered decreasing.

For example, the sequence | 3 4 | 8 7 6 | 1 2 | 5 | contains two decreasing strips and two increasing strips. The sequence | 3 4 | 7 6 5 | 1 2 | 8 also has contains two decreasing strips and two increasing strips. Which are the two decreasing strips?

Improved Algorithm for transforming a sequence of *n* numbers into the identity sequence.

1) If there is no decreasing strip, create one by doing ***one*** of the following:
    a. If 1 is not in the first position, invert the first increasing strip OR
    b. If *n* is not in the last position, invert the last increasing strip OR
    c. Invert the strip between the first and second breakpoints after 1.

2) Find the lowest number that is in any decreasing strip. Call this number *x*.

3) If *x* is 1 and *x* is not in the first position, invert the subsequence from the first position through 1. (This will put 1 in the first position.) Otherwise, invert the subsequence that results in *x* and $x-1$ being adjacent. (This subsequence must have breakpoints at each end).

4) Repeat steps 1 through 3 until you have the identity sequence.

Apply the new algorithm to the transformation of each sequence below. Show each inversion and note which step of the algorithm is being applied.

3 4 5 1 2　　　　4 6 3 2 1 5

Use the new algorithm to transform each of the sequences below. Note the number of breakpoints at each step.

3 5 4 1 2 6　　　　6 3 5 2 1 4　　　　2 1 4 6 3 5 7　　　　2 4 6 7 5 3 1

Homework

For each of the following sequences, place a vertical slash at each breakpoint.

1) 1 4 3 2 5

2) 1 2 3 4 5

3) 4 3 5 2 1 6

4) 3 2 5 6 4 1

5) 7 8 6 5 4 3 2 1

Find the number of breakpoints for each sequence below.

6) 5 4 3 2 1

7) 1 2 5 4 3

8) 3 5 6 4 2 1

9) 1 3 2 6 4 5

10) 3 1 2 4 8 7 5 6

For each of the following sequences, place a vertical slash at each breakpoint and label each strip as either *increasing* or *decreasing*.

11) 3 4 6 5 12 8 9 11 7 1 2 10

12) 4 5 1 2 3 8 9 7 6

13) 1 2 3 20 14 18 19 7 8 9 13 5 4 6 11 12 10 17 16 15

Use the new algorithm given in class to transform the following sequences into the identity sequence. Note the total number of breakpoints at each step.

14) 2 3 4 1 5

15) 5 1 2 3 4 6

16) 1 3 2 5 7 4 6 8

17) 4 5 2 1 6 3 8 7 9

18) 1 2 3 6 7 8 4 5

Day 4:

Analysis of new algorithm

In the second part of step 3 of the algorithm, x and x-1 become adjacent. What would happen if they were adjacent before the inversion?

Explain why step 3 always decreases the number of break points by at least 1. In what case(s) does step 3 decrease the number of break points by more than 1? More than 2?

What is the largest number of break points that one step can eliminate?

The lower bound is the smallest number of steps necessary to invert the sequence. If b is the number of breakpoints, explain why the lower bound is equal to $\frac{b}{2}$ if n is even, or $\frac{b+1}{2}$ if b is odd.

Calculate the lower bound for the inversion distance in each sequence below.

3 2 5 6 7 4 1 8

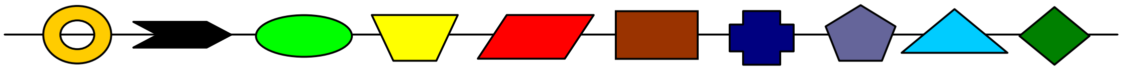
3 2 5 7 4 6 8 1

4 5 2 1 6 3 8 7 9

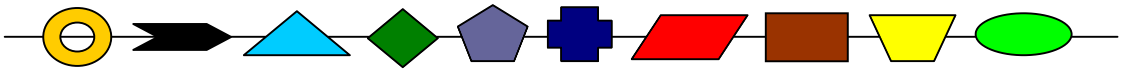
3 2 1 5 4 8 9 7 10 12 6 11

Genes from two different specimens are shown below. Determine the minimum number of inversions necessary to transform one into the other. If one is cabbage, what do you think the other is?

B. oleracea



B. campestris



data adapted from “Transforming Cabbage to Turnip” by S. Hannenhalli and P.A. Pevzner , Journal of the ACM, Vol 46, No. 1, January 1999

Phylogenetic Tree Creation

You are a biomathematic graduate student conducting research in the Amazon Rain Forest. 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 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 for your local Natural History Museum. The museum would also like to create a display of your findings using model creatures. Therefore, a drawing of the phylogenetic tree complete with pictures is needed. The order of genes is as follows:

New creature: 7 9 3 5 2 1 8 4 6

Frillneck lizard (*Chlamydosaurus kingii)*: 1 2 3 4 5 6 7 8 9

Theory

1) Explain why the identity sequence has no breakpoints.

2) Can a sequence have exactly one breakpoint? Explain.

3) Write a sequence that has exactly two breakpoints.

4) Write a sequence with four breakpoints that has no decreasing strips.

5) Assuming that the standard alphabet is the identity sequence, find the lower bound for the inversion distance of the following sequence.

a b c d e f j k l m p q r s u v z i h g n o t w y x

6) Is it possible that a mutation by inversion is a *good* thing for an organism? Explain.

7) Draw a flowchart for the algorithm given in the worksheet.

EXTENSION ACTIVITIES

Extension Activities

You may have noticed that any inversion transformation has been possible. This however is not always the case. In July 2004, a team of researchers from the Department of Genetics and Microbiology at the Universitat Autònoma de Barcelona (UAB) discovered that transposons, small DNA sequences that travel through the genomes, can silence the genes adjacent to them (published in the journal Proceedings of the National Academy of Sciences (PNAS).

"Transposons are repeated DNA sequences that move through the genomes. For a long time they have been considered as a useless part of genetic material, DNA leftovers. However, it is more and more clear that transposons can cause favorable changes for the adaptation and survival of the organism"

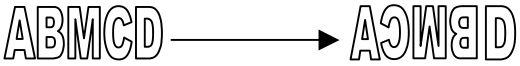
(http://news.biocompare.com/newsstory.asp?id=44536, 2004)

This is interesting, because transposons represent 45% of the genetic material in DNA. Almost half of the possible inversions would contain transposons, and their break points. If an inversion suddenly places a transponson next to a gene necessary for life (regulates heart rate/ breathing rate) the effects would be terrible for the possible offspring. Taking this in mind, pretend there is a transposon at the end of gene 3, and death results in the organism's descendants if gene 5 if turned off. Find the inversion that would transform into the original identity sequence with out ever placing gene 3 immediately to the left of gene 5.

14253

Extension 2

When genes are actually inverted in an organism they not only change position but read ‘backwards’ within the chromosome. This is best illustrated if one gene is inverted. For example, in the sequence ABMCD, if the M is removed, rotated 180° and reinserted in the same spot it would actually lead to ABWCD. Note the M is now ‘upside down’ and becomes the W. In a longer inversion it would look like the following where BMC is inverted.



In mathematical notation the change above would be represented as

1 2 3 4 5 → 1 -4 -3 -2 5

Note the negative signs denote a gene that is now positioned ‘backward.’ Therefore, in reality a sequence of 1 2 3 4 -5 is not the same as 1 2 3 4 5. It will require one more inversion of the -5 to obtain the +5 direction. Taking both sequence and direction into account is a much more interesting and realistic problem.

For example, to convert 2 -3 -4 5 -1 into the sequence 1 2 3 4 5, the following inversions would be performed

2 -3 -4 5 -1 = start
2 3 -4 5 -1
2 3 4 5 -1
-5 -4 -3 -2 -1
1 2 3 4 5 = identity

Note that you can not follow the same algorithm used in this module to complete these problems. In fact, an intriguing mathematical concept known as *graph theory* is used to determine the solution to this problem. But that is a story for another day. In the meanwhile, see if you can convert the following sequence:

-3 4 -1 2 5 = start
1 2 3 4 5 = identity

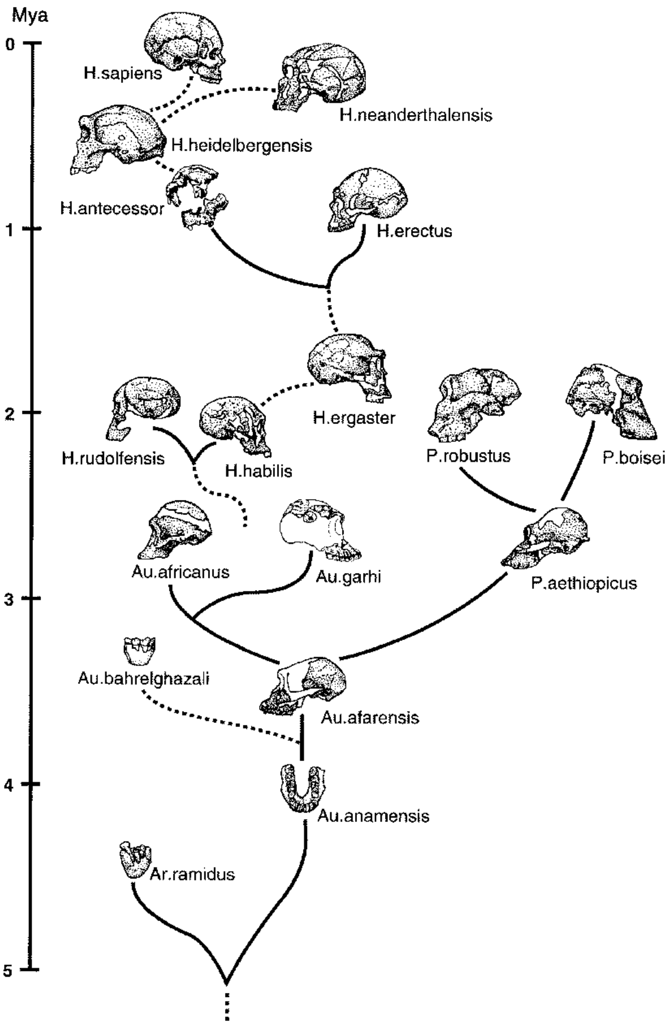
This concept can also be seen in the actual version of the mouse to human rearrangement sequence.

Mouse: 1 -7 6 -10 9 -8 2 -11 -3 5 4
Human: 1 2 3 4 5 6 7 8 9 10 11

This rearrangement can be done in 7 steps. Can you find the solution?

ANSWER KEYS

Answer Key to day 1 homework



http://www.theosophy-nw.org/theosnw/evol/s3amib2.gif

1. **What do we call the type of diagram above? *(Phylogenetic tree.)***
2. **Which two are more closely related, H. rudolfensis and H. erectus or P. robustus and P. boisei? *(P. robustus and P. boisei because they are direct descendants of a common ancestor.)***
3. **Who is the direct ancestor of P. boisei? *(P. aethiopicus)* Au. Garhi? *(the node is blank so we have not found that species yet)***
4. **What does it mean if a species has no descendants? *(that species is extant – still living – or extinct)***
5. **How do you think they made decisions regarding where to place species on this tree? *(answers may vary but will likely be based on the skull structure)***
6. **What other ways could they have compared these species to place the on the tree? What way do you think would give the most accurate answer? *(answers may vary but discussion should lead to the idea that we now use genetic information to compare species)***

Solutions to exercises.

1) 1 | 4 3 2 | 5

2) no breakpoints

3) | 4 3 | 5 | 2 1 | 6

4) | 3 2 | 5 6 | 4 | 1 |

5) | 7 8 6 5 4 3 2 1 |

6) | 5 4 3 2 1 |
2 breakpoints

7) 1 2 | 5 4 3 |
2 breakpoints

8) | 3 | 5 6 | 4 | 2 1 |
5 breakpoints

9) 1 | 3 2 | 6 | 4 5|
4 breakpoints

10) | 3 | 1 2 | 4 | 8 7 | 5 6 |
6 breakpoints

11) | 3 4 | 6 5 | 12 | 8 9 | 11 | 7 | 1 2 | 10 |
I D D I D D I D

12) | 4 5 | 1 2 3 | 8 9 | 7 6|
I I I I

13) 1 2 3 | 20 | 14 | 18 19 | 7 8 9 | 13 | 5 4 | 6 | 11 12 | 10 | 17 16 15 |
I D D I I D D D I D D

14) | $\underline{2\ 3\ 4 \mid 1}$ | 5
1 | $\underline{4\ 3\ 2}$ | 5
1 2 3 4 5

15) | 5 | $\underline{1\ 2\ 3\ 4}$ | 6
| $\underline{5\ 4\ 3\ 2\ 1}$ | 6
1 2 3 4 5 6

16) 1 | $\underline{3\ 2}$ | 5 | 7 | 4 | 6 | 8
1 2 3 | $\underline{5 \mid 7 \mid 4}$ | 6 | 8
1 2 3 4 | 7 | $\underline{5\ 6}$ | 8
1 2 3 4 | $\underline{7\ 6\ 5}$ | 8
1 2 3 4 5 6 7 8

17) | $\underline{4\ 5 \mid 2\ 1}$ | 6 | 3 | 8 7 | 9
1 2 | $\underline{5\ 4 \mid 6 \mid 3}$ | 8 7 | 9
1 2 3 | 6 | $\underline{4\ 5}$ | 8 7 | 9
1 2 3 | $\underline{6\ 5\ 4}$ | 8 7 | 9
1 2 3 4 5 6 | $\underline{8\ 7}$ | 9
1 2 3 4 5 6 7 8 9

18) 1 2 3 | 6 7 8 $\underline{\mid 4\ 5}$ |
1 2 3 | $\underline{6\ 7\ 8 \mid 5\ 4}$ |
1 2 3 4 5 | $\underline{8\ 7\ 6}$ |
1 2 3 4 5 6 7 8

Sample Phylogenetic Tree

(reference for frillneck lizard problem)

$1\ 3\ 2\ 5\ 7\ 4\ 6 \longrightarrow 1\ 2\ 3\ 4\ 5\ 6$

The inversion sequence

1 3 2 5 7 4 6 = *start sequence*

1 2 3 5 7 4 6

1 2 3 5 6 4 7

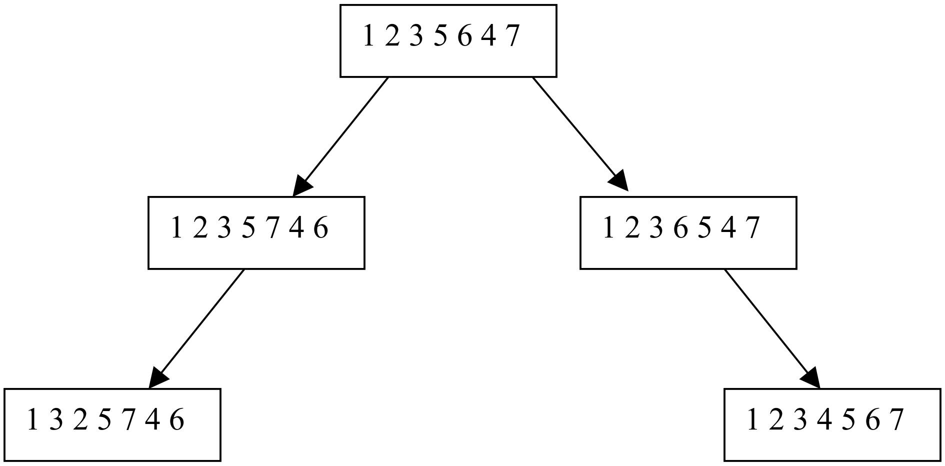
1 2 3 6 5 4 7

1 2 3 4 5 6 7 = *identity sequence*

The start sequence and identity sequence represent two currently living organisms. The question is… **How related are they?** As noted from the inversion sequence above, they are only separated by four inversions. From an evolutionary perspective there are two basic interpretations of this data.

1. One sequence arose from the other through the series of inversions
2. Both sequences (start and identity) shared a common ancestor which can be located within the intermediate inversion steps. This is the more likely scenario.

If a tree is created to represent scenario 2 it might look like this



Could you create another tree that is consistent with this data?
1) It begins with 1 and ends with the largest number and all internal values differ by one from their neighbors.

2) No.
Case #1: If there is a breakpoint at an end of a sequence, then there must be an internal pair of values that differ by more than one creating another breakpoint.

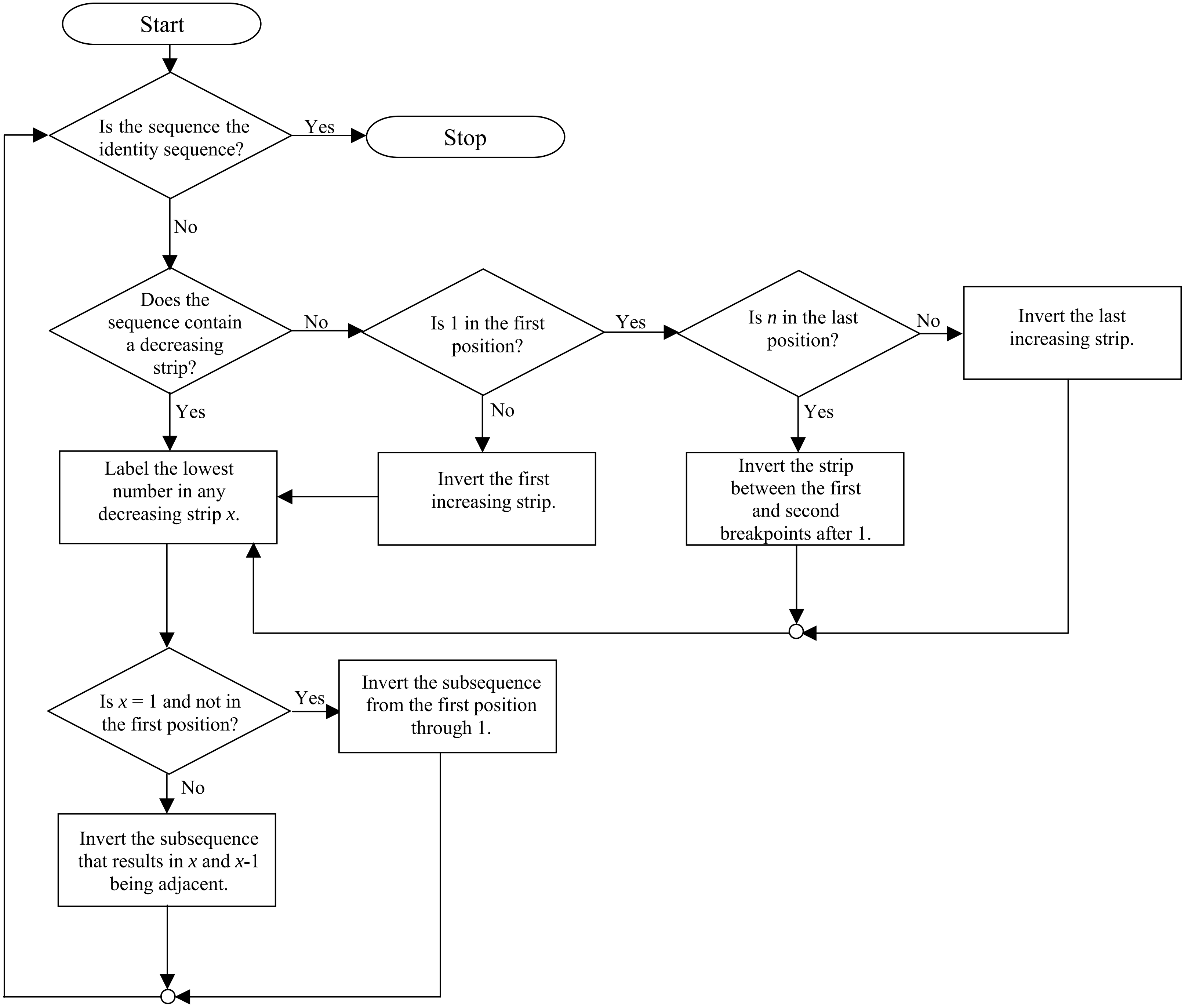
Case #2: If there is a breakpoint in the interior of a sequence, then there are two values that differ by more than one. This creates two strips, one to the left of the breakpoint and one to the right. If this were the only breakpoint, then the strip to the left would have to be increasing (in order to start at one), and the strip to the right would also have to be increasing (in order to end with the largest number.) This however is impossible since this would only occur in a sequence that is already in order. Therefore there must be another breakpoint somewhere.

3) 5 4 3 2 1 Actually, any decreasing sequence will work.

4) 1 2 7 8 5 6 3 4 Other answers are possible.

5) a b c d e f | j k l m | p q r s | u v | z | i h g | n o | t | w | y x |
lower bound = 5

6) Yes. This is called evolution. Numerous organisms exist today that have gone through many mutations.


Start
Is the sequence the identity sequence?
Yes
Stop
No
Does the sequence contain a decreasing strip?
No
Is 1 in the first position?
Yes
Is *n* in the last position?
No
Invert the last increasing strip.
Yes
No
Yes
Label the lowest number in any decreasing strip *x*.
Invert the first increasing strip.
Invert the strip between the first and second breakpoints after 1.
Is *x* = 1 and not in the first position?
Yes
Invert the subsequence from the first position through 1.
No
Invert the subsequence that results in *x* and *x*-1 being adjacent.