Counting DNA Bases with Microsoft Excel

Grade: 9th – 10th, AP Biology

Rational or Purpose: This activity aims to incorporate the use of Microsoft Excel in biology. Students will use the "countif" function to count the number of DNA bases in a single-stranded DNA. When the students have the quantity of each base, they can use this information to rank the melting point of several DNA sequence.

Materials:
Computer with Microsoft Excel
Student worksheet

Lesson Duration: 50 minutes

TEKS Objectives:
§112.43. Biology. (c) (2a-d) (6a)

AP Biology
II. Heredity and Evolution
        B. Molecular genetics: DNA Structure

Background Information: At the beginning stage, we may use simple methods to count the amount of each DNA bases in a DNA sequence. However, scientists usually deal with a large numbers of DNA bases in a sequence. To avoid mistakes, scientists must rely on computers to do the job for us. The quantity of each base in a sequence is very important for scientists to determine the optimum temperature to carry out an experiment. If the temperature is too high, the heat energy may destroy the structure of DNA; or, if the temperature is too low, it may not enough to initiate a chemical reaction. Therefore, knowing the base quantities becomes very important.

Activity
Students will learn to use "countif" function to count the number of bases in Excel. This is a useful skill for the students not only on science, but in daily life. Students are encouraged to explore the application of the function in regular daily life.

Procedure

1. If there are enough computers, students should do this activity individually. If students are sharing computers, there should be no more than 3 students per group.

2. They will launch Microsoft Excel and follow the procedure on the student worksheets. Each cell should only contain <u>one</u> base.

3. Enter the sequence vertically. (Teachers can also download the attached Excel document as a template.)

4. Students should double check the sequences they have entered, to ensure correct data entry.

For AP Biology teachers: This exercise would be more useful after they learn polymerase chain reaction (PCR).

Counting DNA Bases with Microsoft Excel

Look at the following DNA sequence:

CGTAAA

How many adenine bases are there? _____

How many thymine bases are there? _____

How many cytosine bases are there? _____

How many guanine bases are there? _____

Let's take a closer of the following DNA sequence:

CGTAACGCCGCTAGCGCTAGCGTTTGACGCTGTCA

How much time do you think you would have to spend if you needed to count the total number of each base?

_____

How about this DNA sequence:

CTCGGGATGGATCTCTGAATCTTTAGCTCGTCGATGCTGCTAGGGTCT…

How much time do you think you would need to count this sequence, if the are 100 more base? How accurate do you think it will be if you do this by hand?

_____

_____

Although it is possible to finish all of this counting by hand, it will take a long time to count all the sequences. At the same time, as the sequence becomes longer, the probability of making mistakes increases as well. Therefore, scientists need to find a way to do this task in the most efficient and accurate manner.

In Microsoft Excel, we can apply a "`countif`" function to count the number of bases on a spreadsheet.

See the diagram below:
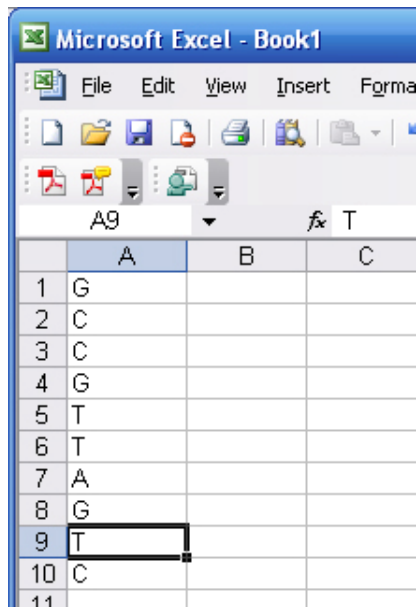
Now, we will try to use the "`countif`" together. The format of the function is:

=countif(range,criteria)

The "`range`" is the range of cells on the spreadsheet. If we have to count the number of G (guanine) between A1 and A10 in the above diagram, we will enter the following in A11 (or in any cell):

=countif(A1:A10,"G")

The criterion is what we are looking for. Since "G" is a non-numerical value, we have to put quotations around the criterion.

Now, enter the following DNA sequence in a column on a spreadsheet.

CGTGC AAGCT CGATG CAAAT CGAGT GCAGT CGATG

Environmental Science Institute
The University of Texas at Austin

Name: _____   Class Section: _____

Date: _____

Apply the "countif" function and find out the quantity of each base in the sequence.

How many adenine bases are in the sequence?   _____

How about thymine?   _____

How about cytosine?   _____

How about guanine?   _____

The purpose of teaching you how to use MS Excel to count the quantity of bases is that the components of the DNA often affect the melting temperature. If there are more cytosine-guanine base pairs, the melting temperature generally tends to be higher; if there are more cytosine-thymine base pairs, the melting temperature would be lower. You may wonder… why? Consider the following example:
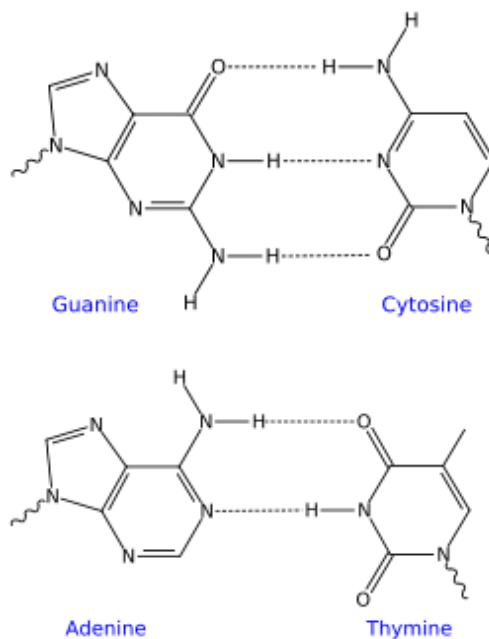

Figure 1. Base pairs.
Source: Wikipedia.com

As you can see, there are 3 hydrogen-bonds in cytosine-guanine base pair, but only 2 in adenine-thymine base pair. If there are more hydrogen-bonds (or any bonds), it means more energy is required to overcome the bond in order to separate two strands.

Although the complimentary strand is not given, we can still figure out the base pair by the given strand because adenine only pairs with thymine, and cytosine only pairs with guanine. Therefore comparing the quantity of "adenine + thymine" and "cytosine + guanine" you will be able to rank the melting temperature of DNA sequences.

Environmental Science Institute
The University of Texas at Austin

Name: _____   Class Section: _____
                                              Date: _____

Exercise

There are 5 sequences in the following. Ranking their melting temperature from 1 to 5; 1 is the lowest temperature and 5 is the highest temperature.

GCTGC GTCAA ATGTA GCCGT TAGCT AGCTT   _____

CGTAG GGCTA CGGCG ATGCA CGATC CGTTA   _____

TATAG TGTTC ATGCC CGGCA CCGTA CGATG   _____

CGTAG CGATT CTGAC CGATC CGGTT CCAGT   _____

CGTAG CGATG CCGTT AATTC CTGAT CGATG   _____