



## Pre-Lab

### Audience

This lab is appropriate for High School (grades 10 and above); AP Biology (completed or currently enrolled) is a suggested pre-requisite.

### Important terms

Here are some important terms you may already be familiar with or could look up and links to some resources to review before watching the presentation.

- DNA
- Gene
- Polymerase Chain Reaction
- Sanger DNA sequencing
- Taxonomy
- Bioinformatics

### Bioinformatics Lab Computer Setup

To follow along with this laboratory you will need:

1. Access to an up-to-date web browser such as Chrome or Firefox.
2. (Optional) We suggest getting a DNA Subway account through CyVerse. To get an account, go to <https://user.cyverse.org/>

### Resources

#### Polymerase Chain Reaction

3D animation: <https://dnalc.cshl.edu/resources/3d/19-polymerase-chain-reaction.html>

2D animation with annotation: <https://dnalc.cshl.edu/resources/animations/pcr.html>

#### What is a Gene

DNA From the Beginning (concept 15): <http://www.dnafb.org/15/>

DNA From the Beginning (concept 17): <http://www.dnafb.org/17/>

DNA From the Beginning (concept 23): <http://www.dnafb.org/23/>



## Sanger DNA Sequencing

DNA Sequencing animation: <https://dnalc.cshl.edu/resources/animations/cycseq.html>

## DNA Barcoding

DNA Barcoding animation: <https://dnalc.cshl.edu/resources/animations/dna-barcoding.html>

## Bioinformatics

What does a bioinformatician do (short interview): <https://dnalc.cshl.edu/view/2018-Bioinformatics-defined.html>

## Laboratory Questions

These questions can be completed using the **Companion packet: Using DNA Subway to Analyze DNA Barcoding Sequences (Pages 1-8)** and by watching the video.

1. What is the accuracy associated with a Phred score of 20? \_\_\_\_%
  - a. How many nucleotide errors would you expect in sequence with a Phred score of 20 yield for:  
...100 bases of DNA? \_\_\_\_\_  
...1000 bases of DNA? \_\_\_\_\_  
...1,000,000 bases of DNA? \_\_\_\_\_  
...3.2 billion bases of DNA (i.e. sequencing the entire human genome) ? \_\_\_\_\_
  - b. Based on the calculations you did above, why do you think bioinformaticians typically trust Phred scores of 30 or higher?
2. Based on the lecture / reading, what does "N" in a sequencing read-out mean?
  - a. Describe the electropherogram peaks that correspond to sequence read-outs of "N."
  - b. Describe the quality scores at those positions.
  - c. Compare the electropherogram peaks and quality scores at a sequence read-out of "N" to peaks that are positively identified as any other nucleotide (A, T, G, or C).
3. Should any sequence(s) in the mosquito data set not be used in further analyses? Why do you think so?