

DNA Barcoding Basics

DNA Subway Blue Line Pt. I

Steps for today's session

- Introduction to Bioinformatics
- Get background on DNA Barcoding
- Examine DNA sequence quality

What is Bioinformatics?



In biology, **Bioinformatics** is an interdisciplinary field that develops and improves upon methods for storing, retrieving, organizing and analyzing **biological data**. A major activity in bioinformatics is to develop software tools to generate useful biological knowledge.

- <http://en.wikipedia.org/wiki/Bioinformatics> - retrieved April 23rd 2013

Bioinformatics is about data

Often, we we are speaking about data in biology, we are talking about DNA Sequence

Bioinformatics is about data

Often, when we are speaking about data in biology, we are talking about DNA Sequence (there are lots of other data, we just won't be talking about that today)

DNA Sequence



Quick Tour

- NCBI Homepage <https://www.ncbi.nlm.nih.gov/>
- Human Genome: <https://www.ncbi.nlm.nih.gov/projects/genome/guide/human/index.shtml>
- Corona Virus Genome: <https://mra.asm.org/content/9/11/e00169-20>

Lab: Mosquito Identification

Can we tell the difference between larvae
that look (nearly) identical?

Aedes adult



By Muhammad Mahdi Karim - Own work, GFDL 1.2, <https://commons.wikimedia.org/w/index.php?curid=11185617>

Anopheles adult



By Jim Gathany - (PHIL), ID #5814. <https://commons.wikimedia.org/w/index.php?curid=799284>

Culex adult



By Muhammad Mahdi Karim - Own work, GFDL 1.2, <https://commons.wikimedia.org/w/index.php?curid=7673048>

Aedes larva



Photograph by Michele M. Cutwa, University of Florida.

Anopheles larva



Culex larva



Photograph by Michelle Cutwa-Francis, University of Florida.

Why does this matter?

Aedes:

- Chikungunya
- Dengue fever
- Lymphatic filariasis
- Rift Valley fever
- Yellow fever
- Zika

Anopheles:

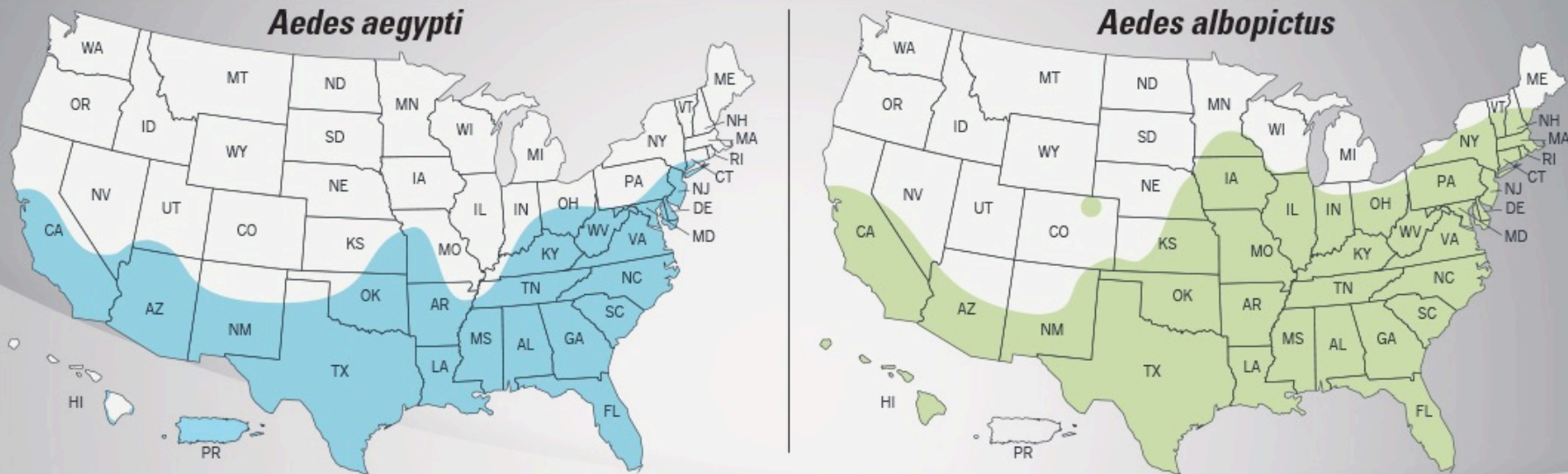
- Malaria
- Lymphatic filariasis

Culex:

- Japanese encephalitis
- Lymphatic filariasis
- West Nile fever

Why does this matter?

Estimated range of *Aedes aegypti* and *Aedes albopictus* in the United States, 2016*



***Aedes aegypti* mosquitoes are more likely to spread viruses like Zika, dengue, chikungunya than other types of mosquitoes such as *Aedes albopictus* mosquitoes.**

- These maps show CDC's best estimate of the potential range of *Aedes aegypti* and *Aedes albopictus* in the United States.
- These maps include areas where mosquitoes are or have been previously found.
- Shaded areas on the maps do not necessarily mean that there are infected mosquitoes in that area.

**Maps have been updated from a variety of sources. These maps represent CDC's best estimate of the potential range of *Aedes aegypti* and *Aedes albopictus* in the United States. Maps are not meant to represent risk for spread of disease.*

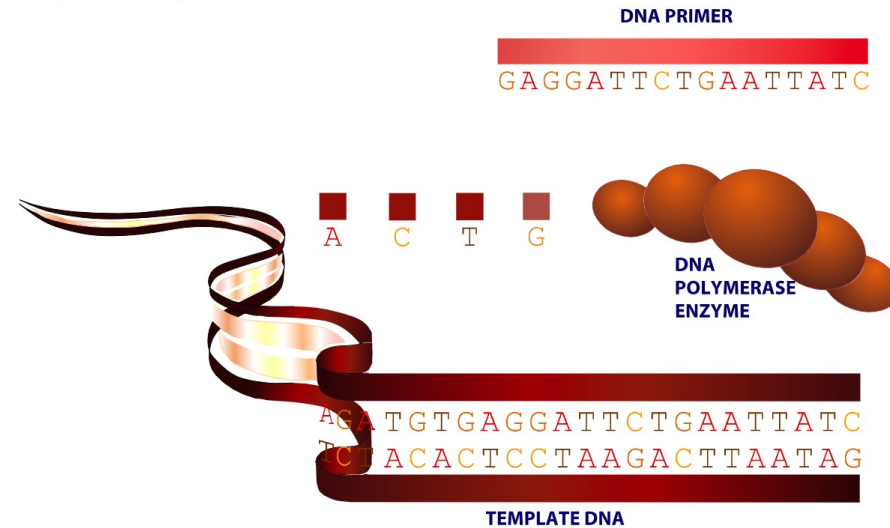
SOURCE: Zika: Vector Surveillance and Control. www.cdc.gov/zika/vector/index.html

Lab: DNA Sequencing Background

DNA Sequencing

Cycle Sequencing

To this mix, we also add a second type of nucleotide; one that has a slightly different chemical formula. These "dideoxynucleotides (ddNTP)" can be recognized by a DNA sequencer.



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

Some Anopheles DNA...

Anopheles gambiae isolate 10016 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence

GenBank: MK592083.1

[GenBank](#) [Graphics](#)

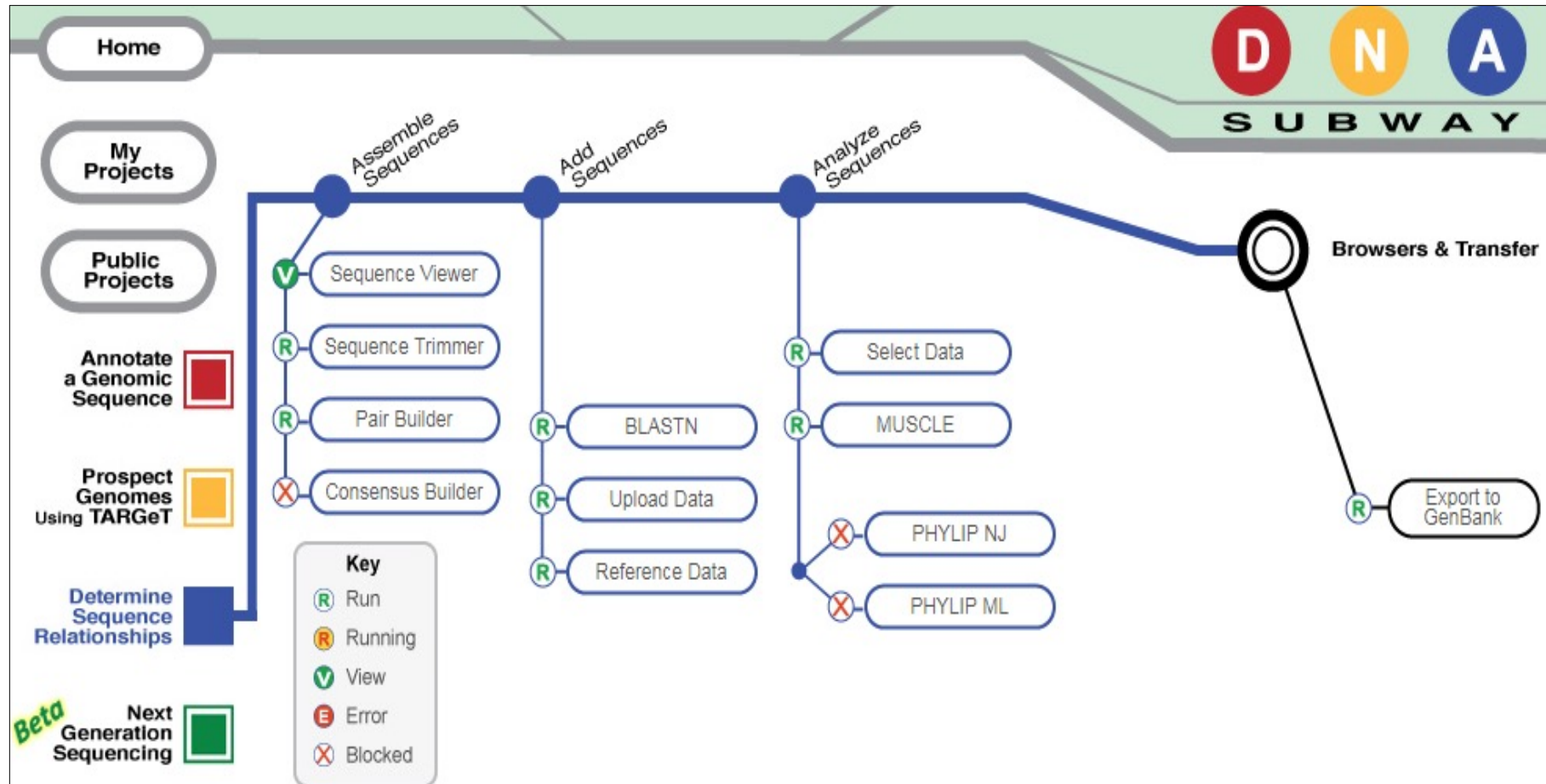
>MK592083.1 Anopheles gambiae isolate 10016 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence

```
ACACATGAACATTGATAAGTTGAACGCATATGGCGCATCGGACGTTTAATCCCGACCGATGCACACATTC
TTGAGTGCCTACTAATTACCAAAGTCTCATTTAGTTAACTACAGTGGCCGTCCGCGAAGGTGCCCGGGTC
ATCCGACGCACTGGGCGGTTCGCTGTGCATAATGACGTGCTTGGTCCCCGTCTGCGGGTCCTCGGGCGTTG
AAAGTGGACACTCTCGAGCGTATGTTGGATGCGTTTCGTGTTGGTGGTGTGTTGATGCGTAGGGCTTGTGG
TGTGTGTCAAGCCGCATGGTTCGAACTAATGCTACGTCGTTCCCGATGGCCACCGGCAGTCTACTCTCCA
GGCTAAAGTCGGCTCGTCTAGGGATTCGGAAAGCTAAGTCGCTGTAACATCATGTGGGCCCATACACGGCG
TTGCGCTACCACGCTAAGTTAGCCCTACATACACAAGCATCAACCCACGGCACGGGCGTAGCTGTAATAC
TTACGTCTCGGTTATACCACGTAGGCCTCAAGTGATGTGTGACTACCCCC
```

Lab: Creating a DNA Subway Project

(follow along in the packet)

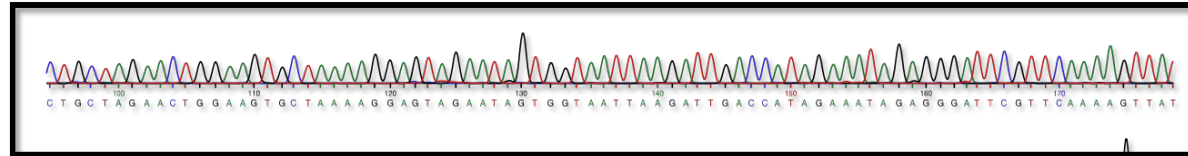
Working on DNA Subway Blue Line



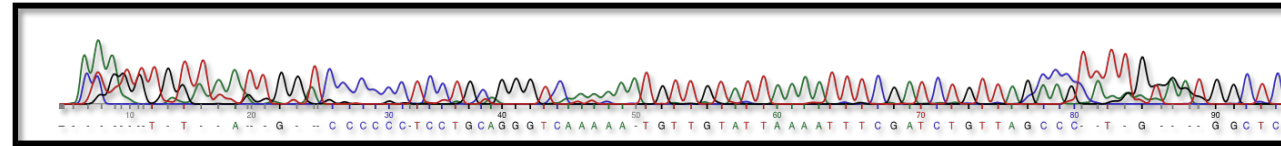
Key Concept: Data Quality

Some sequence examples...

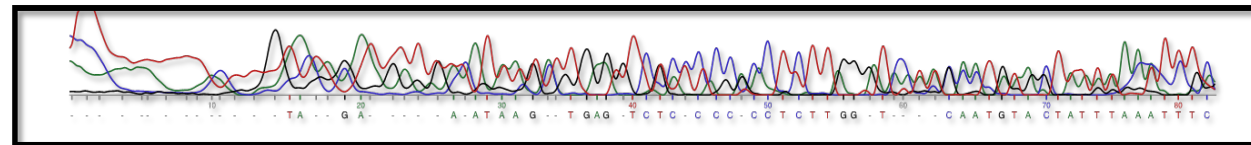
High Quality Sequence



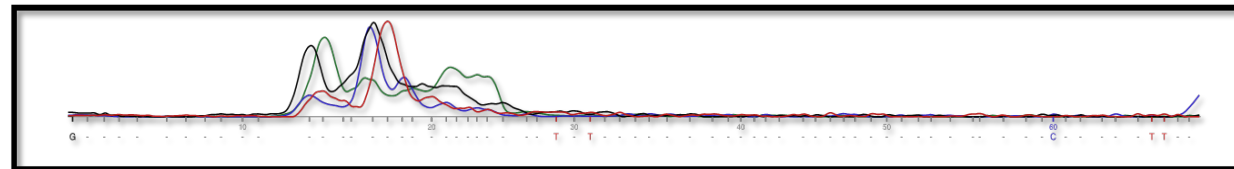
Acceptable Quality Sequence



Low Quality Sequence (multiple base calls per position)



Low Quality Sequence (no base calls)



Phred scores...

Phred Score	Error (bases miscalled)	Accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

Next time:
Comparing sequences with BLAST