YOU: CDC INVESTIGATORS (TEAM 1)
WHERE: Puff the Magic Tapas Restaurant in Aurora, CO
WHY: Food poisoning outbreak

Introduction. A pathogen responsible for a foodborne outbreak at Puff the Magic Tapas restaurant in town has been sequenced with a Whole Genome Sequencing experiment to determine the culprit. About 30% of the customers from the night of Valentine's were either hospitalized or treated as outpatients for nausea, vomiting and stomach cramps associated with the illness. The investigators believe the illness was derived from eating one of the dishes: Show Me the Bacon Tart, Crazy about Cluckers Chicken Bake or the Spin Me Right Round a Spinach Salad. In this activity, you will build the reads generated by the sequencing and then you will help the team of bioinformaticians to assembly a short fragment of the genome and discover the pathogen to which the DNA belongs.

Material and Methods.

1. 16 pieces of 1X1 Lego bricks (blue, red, yellow and green), simulating the 4 bases in DNA
    - **A = BLUE**
    - **G = RED**
    - **C = YELLOW**
    - **T = GREEN**
2. 8 pieces of 1X1 Lego round plates simulating the 4 bases making an index (A, G, C and T)
3. 1 baseplate (32x32) simulating the flow cell
4. Reference genomes table. This database contains the short-genomes of six foodborne pathogens:
   Listeria monocytogenes, Staphylococcus aureus, Escherichia coli O157:H7, *Salmonella, Campylobacter jejuni, Clostridium botulinum*
5. Worksheet to record your results
6. A pen to write the sequences.

Directions. Follow the instructions for discovering your pathogen.

1. *Build the reads!* Using the Lego bricks provided (or the ppt "bricks"), build the following reads belonging to the unknown pathogen sequenced. Add the first appropriate Lego color block to the baseplate and then continue building up to complete each read. If using the ppt to build, you may choose to build the sequences left to right or from the bottom up. Just be sure to keep track of the 5' – 3' orientation as you build. I suggest building on one slide, grouping your build, and then cut and paste on the next slide so you will be able to arrange them later (Fig. 1). Copy and paste shapes if you run out.

Group 1	**Bases** 5'-XXXXX-3'
1	TCGCACG
2	ATTATCG
3	CACGTCC
4	ACGTCCA

5	TGAACCA
6	ATGTTGG
7	TGGTGAA
8	CCATATG

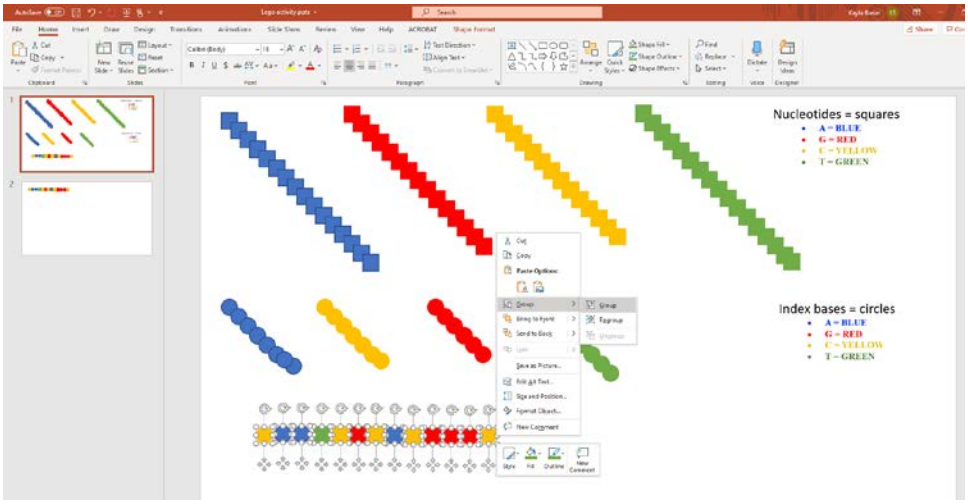


Figure 1: Example of building “legos” in PowerPoint and how to “group” your build.

2. *Attach the indexes.* Choose a couple of indexes (one i3A and one i3B) from the following table and add them at the ends of any reads in order to recognize the reads of your pathogen in a pool of different experiments. Write the indexes chosen in the “Formative assessment quiz”. These are not used in the identification of the pathogen. This is an exercise to understand how the final reads appear and how they would normally be categorized (Fig. 2).

i3A Index Name	i3A Bases	i3B Index Name	i3B Bases
A501	CAA	B701	GGC
A502	GGA	B702	TGT
A503	TCA	B703	TGG
A504	AGG	B704	TTG
A505	CTT	B705	ATT

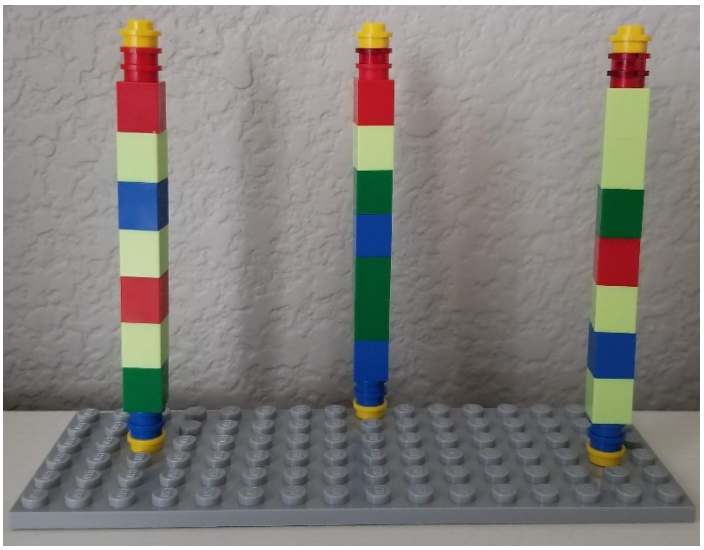


Figure 1: Example of 3 DNA sequencing reads. Contains the sequence and indexes.

3. *"Bioinformatic" activity.* Align the reads obtained in order to find the contig(s). Be sure you have the reads aligned the same way, 5' to 3' (Fig. 3). This is one of the computational problems that bioinformaticians have to resolve when they must assemble a genome, starting from the sequencer output. (Hint: in this exercise, the reads can be aligned only in one direction and the reads overlap for at least two bases).

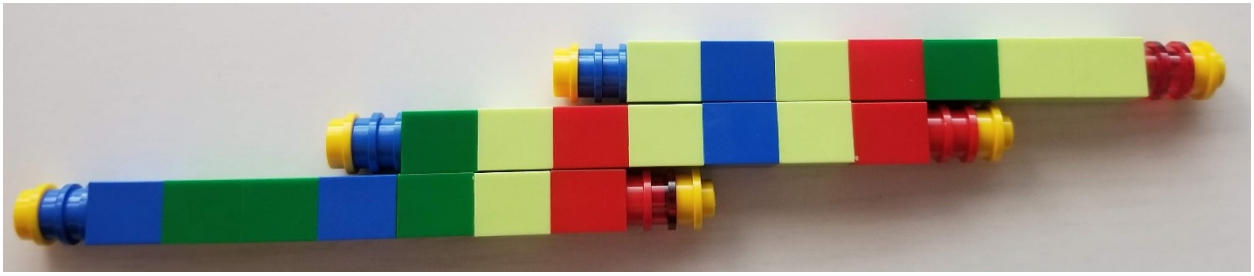


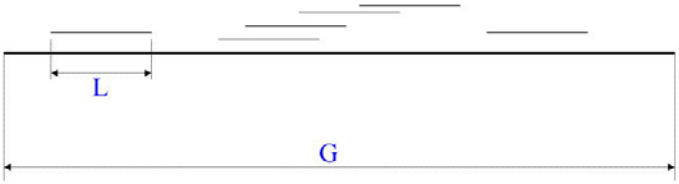
Figure 3: Example of what an alignment looks like. Notice the indexes are not utilized for the alignment.

4. *Draft the sequence.* Write the sequence in the table provided.
5. *Compare, align and find the match.* Someone in your group should now open up the "Genome Samples" document provided in the canvas course. Search for the sequence you aligned and determine the infective agent that caused Puff the Magic Tapas to close indefinitely. (You can also "eyeball" the sequences and find the full read.) There is only one possible solution. Report the results in your notebook.

Conclusion. Complete the activity by filling in the worksheet and answering all questions. Please pull apart the Legos into individual pieces and make sure that all the pieces go back into the bag/box when you are finished with this activity!

Lander Waterman Statistics

G = Genome Length
L = Fragment Length
N = Number of Fragments
T = Required Overlap
c = Coverage = LN/G
α = N/G
θ = T/L
σ = 1 - θ



*Copy this to your OneNote Notebook and complete there

CDC INVESTIGATORS (TEAM 1): FINAL REPORT

INVESTIGATORS
(Names):_____

INDEXES used:

i3A Index Name	**i3A Bases**	**i3B Index Name**	**i3B Bases**

SEQUENCE obtained:

PATHOGEN identification (which dish do you think they found it in? – just for fun):

What is a read?

A) It is a software used by bioinformatician
B) It is a code of NGS
C) In reality consists only in a sequence of 7 bases
D) Is the base unit for sequencing generated from genome fragmentation.

Which of these is INCORRECT?

A) Shotgun and next-generation sequencing (NGS) involve shredding the genome into smaller fragments, and sequence either full or part of the fragments
B) Reads are the genes of virus called “short fragmented tobacco virus”
C) Overlapping of sequences between reads are the basis of *de novo assembly*
D) A contig refers to overlapping sequence data (reads)

For an alien very short genome of 800 bp, and your new technology methods that produce reads of 7 bases, calculate the number of reads for a coverage of 10 (C=10) following the Lander/Waterman equation:

C = LN / G where G = the genome length; L = the read length; N = the number of reads

Result:_____

In this activity the coverage was _____