# BIMM 104: Genome Informatics

## Course Instructor: Dr. Barry J. Grant (<u>bjgrant@umich.edu</u>) Class Website: <u>https://bioboot.github.io/bimm104\_w18/</u>

### Unit title:

Alignment method foundations of genomic analysis (2 Lectures)

## Lecture number(s): 4 & 5

#### Pre-class material:

Screen-cast:	Alignment fundamentals: Why, how and where?
Reading assignment:	Sean Eddy's <u>"What is dynamic programming?</u> "
	Nature Biotechnology 22(7) 2004, 909-910.

#### Unit learning goals:

At the end of this course students will:

- 1. Understand that sequence alignment is the most fundamental operation underlying genome informatics and indeed much of bioinformatics.
- 2. Be able to describe in general terms how dynamic programming works for pairwise sequence alignment.
- 3. Appreciate the differences between global and local alignment along with their major application areas.
- 4. Understand why heuristic approaches become necessary for large database searches and many genomic applications.
- 5. Appreciate that even when optimal solutions can be obtained they are not necessarily unique or reflective of the biologically correct alignment.

## Terminology:

Gaps	Ortholo
Global alignment	Percer
Heuristic	PSI-BL
HMMER	Scoring
Homologue	Seque
Indels	Seque
Insertions	Short r
Local alignment	Substit
Mutations (Substitutions)	
	Gaps Global alignment Heuristic HMMER Homologue Indels Insertions Local alignment Mutations (Substitutions)

Ortholog/Paralog Percent identity PSI-BLAST Scoring scheme Sequence identity Sequence similarity Short read aligners Substitution matrix

#### **Optional learning goals**:

Understand that sequence alignment methods are central to homolog detection, sequence comparison, database searching, genomic assembly, genomic re-sequencing and the mapping of reads to reference genomes, gene annotation, protein structure prediction and many other bioinformatics application areas.

# Potential student misconceptions:

- 1. Alignment tools are all the same in terms of performance, output and underlying assumptions.
- 2. There is one unique highest scoring alignment answer independent of the chosen scoring scheme.
- 3. Alignment output should not be inspected manually (*i.e.* trust the black box).

## Homework

This unit's homework consists of both (1) an online <u>knowledge assessment quiz</u> and (2) a Needleman-Wunsch dynamic programming assessment exercise. Both components contribute 50% to this unit's grade.

For the later we have two sample sequences, and we'd like to use the Needleman-Wunsch algorithm discussed in class to align them.

Sequence 1: ATTGC Sequence 2: AGTTC

Using a match score of +2, a mismatch score of -1, and a gap score of -2. Draw and fill in an alignment matrix and translate your completed matrix values into an alignment. What is the optimal score for this alignment? Is there one unique alignment with this score? Please submit your answer as a picture or word document to our BlackBoard site.

Step	Fulfilled Assessment Criteria	Points	
1	Setup labeled alignment matrix	10	
2	Include initial column and row for GAPs	10	
3	All alignment matrix elements filled in	10	
4	Evidence for correct use of scoring scheme	10	
5	Direction arrows drawn between all cells	10	
6	Evidence of multiple arrows to a given cell if appropriate	10	D
7	Correct optimal score position in matrix used	10	С
8	Correct optimal score obtained for given scoring scheme	10	В
9	Traceback path(s) clearly highlighted	10	А
10	Correct alignment(s) yielding optimal score listed	10	A+

# Scoring Rubric

(100 Total points)