CMPSC 300
Introduction to Bioinformatics
Fall 2017

Lab 6: Alignment
Save this lab assignment to: labs/lab6

Objectives

- To understand the value of aligning genes and recognize the practical applications of this technique.

- To gain familiarity with the use of web-based alignment tools to explore sequence similarity and understand how to modify their parameters.

Reading Assignment

Chapter 3 in Exploring Bioinformatics textbook.

Part 1: Needleman-Wunsch Algorithm

In this part of the lab you are invited to further practice using a particular global alignment technique called the *Needleman-Wunsch* algorithm. In this lab, you are to create a report document that provides the answers or solutions to the following tasks:

1. **By Hand**: Compute the alignment of two sequences using the Needleman-Wunsch algorithm to generate an optimal global alignments for the sequences **ACGTACT** and **ACTACGT** by hand. **Please show your work as demonstrated in class to complete this task by hand.** For this, you are expected to create your n by n matrix in as shown in Table 1 where you label each entry by a letter: $\{(a),(b),(c), \ldots (g)\}$ to be used to illustrate your three calculations, per entry, in your submitted document. *Note: Gap = -1, Match = 1, and Mismatch = 0.* To complete this task, show the sequences as they align optimally.

2. **By software**: Use the included python program (Needman-Wunsch.py) to implement a Needleman-Wunsch global alignment over the sequences **CTAG** and **CGCTAATC**. Hint: Use this program to determine 10 altogether. Please show your work as demonstrated in class by using screen shots of program output that helped you reach your conclusions.

3. When do you expect to happen if you aligned **CAG** with **TTTCAGCAGTTT**? Use the Needleman-Wunsch algorithm (using the program or with pencil and paper) to generate all optimal global alignments for these sequences. Please show your work as demonstrated in class (screen shots or typed in calculations similar to task 1, above). State your initial reasoning and then discuss how your reasoning met (*or failed to meet*) your expectations correct?

4. Note: in each case above, you can confirm your work by using the included python program. To run the program, you will have to edit each of the two input sequence files to update them with your sequences.

Table 1: Label your matrix in this fashion and use the lettering-system to indicate the calculations of your own implementation of the NeedleMan-Wunsch algorithm for alignment.

		A	**T**
	0	-1	-2
A	-1	(a)	(b)
T	-2	(d)	(c)
G	-3	(g)	(d)

Part 2: Investigation of Online Alignment Tools for Influenza Virus

In the second part of the lab you are invited to explore online global and local alignment tools to investigate virulence.

Influenza viruses have received a great deal of study, and the ability to compare many strains has lead to significant advances in understanding what allows one strain to cause more severe disease than another. The H5N1 "bird flu" virus makes an interesting case in point. The virus causes severe influenza in birds and has become established in populations of domestic chickens and turkeys. Human cases occur sporadically, mostly in individuals heavily exposed to infected birds, such as poultry farmers, and H5N1 flu is severe for humans as well. Once a human case occurs, however, spread to another human is exceedingly rare, even among family members in close contact with the infected individual.

A 2006 article by van Riel *et al.* demonstrated that the avian H5N1 virus binds to a form of sialic acid receptor that in humans is only found far down in the lungs and lower respiratory system. Human viruses, in contrast, bind to a form of the receptor common in the upper respiratory tract. Thus, it is difficult for H5N1 to infect humans because our respiratory defenses normally prevent the virus from reaching the lungs. However, a mutant strain in which HA was altered to be able to bind to sialic acid receptors in the upper respiratory tract could be a very dangerous strain indeed.

So far, no such H5N1 strains that infect humans efficiently have been observed. However, we might ask whether the strains that do make it into humans tend to have altered HA genes - if so, that would suggest that either adaptive mutations could be occurring within the human host or that the viruses that cause human infections are subpopulations that are already better adapted. There are many avian H5N1 sequences available and a number of sequences of the H5N1 viruses isolated from infected humans, so we can use sequence alignment to see whether these have essentially the same HA or noticeable differences.

1. Find the listed file (below) that have been have been uploaded to the lab6 directory in the shared repository.

   - Influenza_A_Chicken_Vietnam2005(avianH5N1)segment4.txt,
   - Influenza_avianHong_Kong2007(avianH5N1)segment4.txt,
   - Influenza_A_ChinaGD012006(humanH5N1isolate)segment4.txt

2. Perform pairwise comparisons of the three strains using at least **two** online tools (two are suggested below, for your convenience).

   - First, you are invited to go to EMBOSS webpage at `www.ebi.ac.uk/Tools/emboss/` and select a global alignment technique to perform pairwise alignment of the three sequences. Save your results as a screen shot.

   - Then, you are then invited to go to NCBI and use a tool called BLAST to perform a similar alignment experiment. BLAST is a local alignment technique, where sequence alignment is done by finding one sequence within another. The influenza virus M2 gene, for example, is another key player in the biology of the virus: once the virus enters the cell, M2 is involved in the release of the virus genome subunits so they can travel to the nucleus and direct viral replication. Suppose we have sequenced segment 7 from the 2009 H1N1 pandemic virus but are uncertain what part of it represents the actual M2 coding region. To find out, we could align the well-characterized M2 coding sequence from the Brisbane strain with the full segment 7 sequence from the newly sequenced virus.

3. Add the following responses to your report:

   - Based on your results (which of course are limited - it would be necessary to do many more comparisons in reality), do you believe there is evidence that human adaptation is occurring in H5N1 viruses that might merit concern about human-to-human transmission in the near future?

   - Discuss similarities and differences between the alignment techniques (Emboss versus Blast at NCBI) you have selected.

   - Please save screen shots from each tool to place into your report.

Required Deliverables

All of the deliverables specified below should be placed into a new folder named 'lab06' in your Bitbucket repository (`cs300f2017-bbill`) and shared with the instructor by correctly using appropriate Git commands, such as `git add -a`, `git commit -m ''your message''` and `git push` to send your documents to the Bitbucket's server. When you have finished, please ensure that you have sent your files correctly to the Bitbucket Web site by checking the `source` files. This will show you your recently pushed files on their web site. Please ask questions, if necessary.

1. A report including your hand-calculations and responses to the alignment problems in Part 1.

2. Your report is to contain your reflections from Part 2 concerning the differences between the online alignment tools.

3. Make sure to submit a LibreOffice file, not a text file, with proper formatting and **your name included at the top of the document**.

You should see the instructor if you have questions about assignment submission.

Handed out on: 4 Oct. 2017