

**CMPSC 300
Bioinformatics
Fall 2019**

**Lab 6:
Global Alignment**

GitHub starter link

<https://classroom.github.com/a/I41q90Bw>

To use this link, please follow the steps below.

- Click on the link and accept the assignment.
- Once the importing task has completed, click on the created assignment link which will take you to your newly created GitHub repository for this lab.
- Clone this repository (bearing your name) and work on the practical locally.
- As you are working on your practical, you are to commit and push regularly. You can use the following commands to add a single file, you must be in the directory where the file is located (or add the path to the file in the command):

```
- git add -A  
- git commit -m ‘‘Your notes about commit here’’  
- git push
```

Alternatively, you can use the following commands to add multiple files from your repository:

```
- git commit <nameOfFile> -m ‘‘Your notes about commit here’’  
- git push
```

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.

HANDED OUT: 7th OCTOBER 2019

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. In your calculations document, please label each series of calculations according to the letter in the cell.

| | | | |
|----------|----|----------|----------|
| | | A | T |
| | 0 | -1 | -2 |
| A | -1 | (a) | (b) |
| T | -2 | (c) | (d) |
| G | -3 | (e) | (f) |

Part 1: Small Tasks using the 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 the two sequences shown in the Table 1 using the Needleman-Wunsch algorithm. The match, mismatch and gap scores are, 1, 0, -1, respectively. In your written work using the file writing/calc.md, please be sure to have all calculations listed on the relevant lines according to the letter of the cell.
2. **By software:** Use the included program (file: `src/needman-Wunsch_localAlignment.i.py`) to implement a global alignment of the sequences of files `input/s1.txt` and `input/s2.txt`. Answer the below [Questions-in-blue](#) for part 1.
 - (a) [How similar were the two sequences \(s1.txt and \(s2.txt\) which you applied to the included alignment program written in Python3?](#)
 - (b) [Are the two sequences closely related to each other?](#)
 - (c) [What proof do you have to suggest such a claim?](#)

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 similarity in viruses.

Influenza viruses have received a great deal of study, and the ability to compare many strains has led to significant advances in understanding what allows one virus to cause a 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 of influenza also occur sporadically, mostly in individuals heavily exposed to infected birds, such as poultry farmers. Interestingly, 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 in lung tissue of the *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 the virus was altered to be able to bind to sialic acid receptors in the upper respiratory tract could be a very dangerous strain indeed for humans.

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 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 strain or one of noticeable differences.

1. Find the listed file (below) that have been have been placed in the **data** directory of your repository.
 - Influenza_A_Chicken_Vietnam2005_avianH5N1_segment4.txt
 - Influenza_avianHong_Kong2007_avianH5N1_segment4.txt
 - Influenza_A_ChinaGD012006_humanH5N1isolate_segment4.txt
2. You are to go to the EMBOSS *Needle* webpage at www.ebi.ac.uk/Tools/emboss/ to perform a global alignment of the above three sequences. Note, there will be three alignments (of sequence pairs) to perform. For example for the sequences, *a*, *b*, and *c*, there will be comparisons of (*a*, *b*), (*a*, *c*) and (*b*, *c*). Please save your results as a screen shot or as a text file.
3. **Questions-in-blue:** Add the following responses to your report:
 - (a) How much similarity exists between each of the sequences to the others?
 - (b) Based on your results (which are too few to provide a comprehensive study), 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?
 - (c) Statistics: What were the numbers of *Lengths*, *Similarities*, *Gaps* and *Scores* for each of your alignment tasks?

Required Deliverables

- Part1's calculations of the alignment done by hand, File: `/writing/calc.md`.
- Questions in blue of Parts 1 and 2, File: `/writing/report.md`

Please see the instructor if you have questions about assignment or its submission.