## Teaching Notes on Bioinformatics FMN 2021

### By Alice Tarun, Ph.D.

Assistant Professor, Biology

Johnson Hall of Science, Rm 125

St. Lawrence University

23 Romoda Dr., Canton, NY 13617

Web: https://www.stlawu.edu/people/alice-tarun

e-mail: [atarun@stlawu.edu](mailto:atarun@stlawu.edu)

**Course Information**

Department: Biology

Level: **Lower Undergraduate**

Course type: **Lab**

Students: **Majors**

Number of Students: 1-100

**Module Information**

# Original Module Name:

# Exercise 1 – Similarity and Sequence Alignment

# Exercise 2 – Sequence Alignment to a Database of Sequences

[Adapted Module Name: Tapprich, W. (2019). [Sequence Similarity Resource Adaptation: Exploring Ebola Virus](http://dx.doi.org/10.25334/Q47X63). [Bring Bioinformatics to Your Biology Classroom](https://qubeshub.org/groups/niblse2019), QUBES Educational Resources. [doi:10.25334/Q47X63](http://dx.doi.org/10.25334/Q47X63).

PowerPoint slides of module screenshots were adapted from a presentation by Hudson Alpha (<http://hghelix.hudsonalpha.org/>) on COVID-19 teaching resources.

Viral Sequence Similarity – A Bioinformatics Activity Using COVID-19, is a derivative work of "[Sequence Similarity: An Inquiry-based and 'Under the Hood' Approach for Incorporating Molecular Sequence Alignment in Introductory Undergraduate Biology Courses](https://qubeshub.org/qubesresources/publications/56/5)" by Adam Kleinschmit, Benita Brink, Steven Roof, Carlos Christopher Goller, and Sabrina Robinson, used under [CC-BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode). Viral Sequence Similarity – A Bioinformatics Activity Using COVID-19, is licensed under [CC-BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode) by HudsonAlpha Institute for Biotechnology.

**Files associated**: Supplementary materials to Module 1 and 2 (Introductory video and PowerPoint slides to accompany Modules 1 and 2, Bioinformatics activity on sequence similarity, alignment, BLAST and sequence databases; and self-reported assessment by students who completed the exercise.)

**Modification Learning Goals**:

After completing Module 1, you should be able to:

1. Define similarity in a non-biological and biological sense.
2. Quantify the similarity between two sequences.
3. Explain how a substitution matrix is used to quantify similarity.
4. Calculate amino acid similarity scores using various matrices.

After completing Module 2, you should be able to:

1. Explain how similarity is used to perform a BLAST search.
2. Explain the BLAST search algorithm.
3. Evaluate the results of a BLAST search.

Other learning objectives

1. Relate the structure of amino acids to their effect in amino acid substitutions in protein structure.

2. Familiarize how to look up genetic information from NCBI

**Teaching Notes**

*(Think about what you would like to read about this activity if you came back to it in 2 years)*

Suggestions for this section (not all required, and extras always welcome):

* What did you change and why?

I used the viral sequence similarity activity by Tapprich (2019) but I used SARS-COVID 19 as a context instead of Ebola, because of the COVID pandemic and to tie it to the topics covered in the lecture.

* How did the activity go?

The two modules were taught in an online lab. The modules and the accompanying worksheets were posted in the lab LMS (Sakai). Together with the modules, I included a recorded video that walks the students to complete the modules using screenshots of the exercises.

* What went wrong and why?

I was not able to fully engage the students with a discussion on the topic because of the online format of instruction. The students completed the activity on their own. Most of them were able to complete it with no problem, but there were a few who were confused or resorted to copying their answers from the answer key available online.

* What was the prep like?
* How much time went into prep?

I prepared the video with screenshots of the activity. Lab instructors piloted the activity to ensure that there would not be any confusing parts in the two modules.

* Did you have to do any prep (i.e. grow cultures, grow seeds, order supplies) ahead of implementation? Not applicable
* Would you do this activity again?

Yes

* What would you change in the future?

In the future, I would like to adapt the modules to illustrate concepts such as identifying consensus sequences or making sequence comparisons to understand SARS-COVID 2 variant.

In terms of implementation, here are some observations and suggestions I have:

1. I will emphasize how each incidence of the query word in a local alignment is a separate probability. In the worksheet, when there were two sites with the MPY query word, it was not clear if the students had to add each incidence into one probability or treat them separately. The students also tend to “force” a pattern, so they will score a partial query word (MP).
2. I will change the Query word and/or sequences in the worksheet to prevent students merely copying the answer key that is available online. I will also re-word the questions in the worksheets to prevent students from copying answers from the answer key and not doing the work they were supposed to be doing.
3. Doing these modules in class would be ideal since it can prevent many issues- such as confusion or looking up the answers on the answer keys.
4. I may introduce sequence alignment with DNA or RNA sequences, and include a section on how consensus sequences can be derived from sequence alignments.

* What do you wish you’d known before you ran the activity?

There were very few complaints about the two modules. Module 1 was very helpful in reinforcing the lecture discussion on amino acid structure. It would be nice to have more information on how the BLOSUM 62 matrix is derived, particularly for the instructors. I wish I had anticipated that students would look up the answer key. I could have modified the questions in the worksheet.

* Is there anything else you would like to make note of?

I conducted a pre- and post- self-reflection survey to determine student’s understanding of concepts covered in the two modules. The results of the survey showed 89% increase in self-reported proficiency using a 5 point likert scale. From of a survey of 99 students, the responses indicate that almost all of the students who responded were unfamiliar with a sequence-scoring matrix, BLAST, NCBI, conserved sequences or FASTA. After conducting module 1 and module 2 exercises, the students reported higher levels of familiarity and the use of the sequence-scoring matrix (91% increase); BLAST (131% increase); NCBI (81% increase); conserved sequences (60% increase); and FASTA (98% increase). The results of the self-reported survey is shown in the attached table.

I also included a question in their lab exam that required them to calculate sequence similarity using a sequence-scoring matrix, and found that more that 80% of the students were able to correctly answer the question.

Table 1. Self-Reported Assessment of students who completed the modules

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Question | Ave Pre-assessmentAve Pre-assess | | Post-Assessment | | Improvement |
|  | Pre-test | St. Dev | Post-test | St. Dev | Post-Pre/Pre |
| 1 | 2.144329897 | 1.331534196 | 4.107143 | 0.92605 | 91.53502747 |
| 2 | 1.663265306 | 1.029601444 | 3.845238 | 1.063396 | 131.1860941 |
| 3 | 2.443298969 | 1.361995747 | 3.904762 | 0.995455 | 59.81514969 |
| 4 | 2.051546392 | 1.263044383 | 3.705882 | 1.060783 | 80.63848655 |
| 5 | 1.75257732 | 1.15794813 | 3.464286 | 1.017074 | 97.66806723 |
|  |  |  |  |  |  |
|  | 2.011003577 | 1.22882478 | 3.805462 | 1.012551 | 89.23199486 |
|  |  |  |  |  |  |
|  | n= 99 |  | n= 86 |  |  |

Questions:

1.       I can use a sequence-scoring matrix to quantitatively compare sequence similarity between two sequences.

2.       I can describe how the BLAST algorithm finds partial regions of similarly within two sequence records.

3.       I know at which level (nucleotide or protein) coding sequences exhibit the most conservation.

4.       I am confident in my ability to obtain sequence data housed in databases within NCBI.

5.       I can describe the FASTA file format.

Likert Scale

5= Strongly agree

4= Slightly agree

3= Neither agree nor disagree

2= Slightly disagree

1= Strongly disagree

* How does this activity fit in your overall course curriculum? This topic was taught as part of the physiology unit (respiration) of the course.
* In what ways, if any, did you modify your teaching practice with this activity?

This would ideally be presented in a face-to-face class so the students can work out the activity with the instructor and with their peers.

I implemented these two modules in an online lab. The modules and the accompanying worksheets were posted in the lab LMS (Sakai). Together with the modules, I included a recorded video that guided the students to complete the modules using screenshots of the exercises.