

**CMPSC 300  
Bioinformatics  
Spring 2021**

**Activity 10:  
A Comparison of Open Reading Frame Tools  
Submit deliverables through the Google Form.**

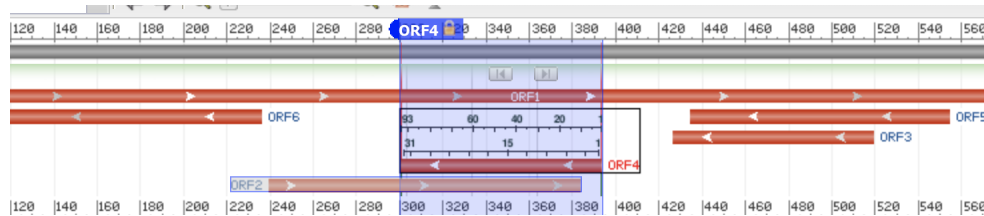


Figure 1: **Open Reading Frame finder from NCBI:** To discover genes in DNA is a complicated task. There are often many different algorithms which must be applied to perform different types of tests. Can you determine the difference that all these tests make?

## Objectives

To learn how to use online tools in conjunction with stand alone programs. To determine differences in results from these two methods of analysis. To use BlastP to learn more about the ORF results.

## Reading Assignment

Chapter 10 in the *Exploring Bioinformatics* textbook.

## Introduction

We have been studying gene prediction technologies. We noted that there are often many types of tests involved with correctly predicting whether a gene is present in DNA, as noted in Figure 1. In this activity, you are to compare the prediction qualities of a stand-alone Python program with the ORF finder tool which is available from NCBI at the below link.

<https://www.ncbi.nlm.nih.gov/orffinder/>

## What To Do

### Survey Questions

You will be responding to questions-in-blue in the below steps. Please place your answers in the Google Doc which can be found at the link:

HANDED OUT: 23<sup>th</sup> APRIL 2021

<https://forms.gle/9HNFCp6AmxQCnXfi9>

## Steps

1. Use your BioPython Docker container to complete all tasks with your Python code.
2. Locate the Python program (`src/orf_finder.py`) and a FASTA data file (`src/fleckOrchid.fasta.txt`) in today's lesson directory of the `classDocs` repository.
3. Run the Python `orf_finder.py` code using FASTA data file (`src/fleckOrchid.fasta.txt`) as a parameter. [What does the stand-alone program do?](#)
4. Using your favorite code editor, open the python code and locate line 10 (`threshold = 100`). Change this value to 50, save your program and run it again. [What does changing the threshold value appear to do?](#)
5. Try changing this threshold parameter to other values so that your program only returns five (5) ORFs from the FASTA file. [What threshold value did you use to find only five \(5\) potential ORF's?](#)
6. [What are the sequences that the program is returning?](#)
7. Set `threshold = 100` and save. Note: if your program becomes corrupt, you have a backup of the program and data file in `src/backUp/`. You could **copy** this file over the corrupt file to regain working code.
8. Go to NCBI's ORF finder at <https://www.ncbi.nlm.nih.gov/orffinder/>
9. Use your favorite editor to open the FASTA data file. Copy and paste the DNA code from into the ORG\_finder tool at NCBI and use the **SUBMIT** button to determine the open reading frames.
10. As you wait for your results to appear; [Briefly compare and contrast the facility and utility in the using script-based tools, as opposed to using web-based tools.](#)
11. Using NCBI's tool; [How many open reading frames did you find using NCBI's ORF tool?](#)
12. [Check and compare the results of the python tool and the online tool. Do any of the outputs match?](#)
13. [Which two sequences were found by both tools?](#)
14. [Why are there any differences between the tool and the web tool in terms of results?](#)
15. Choose an ORF (such as ORF1, beginning with `MFTT...`). Copy and paste this associated protein sequence into Blast's protein search tool at link: <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>;
16. [What is the name of a protein you found by using BlastP? What function can you argue that it has?](#)

17. Name two or three organisms which have similar proteins (or at least turn-up in your results).
18. In terms of E-values for the results you used to determine the protein's function, are you convinced that this protein is a match in the organisms you mentioned?
19. What major differences do you find in the quality of the results from the stand-alone and online tools?

### Required Deliverables

- Complete Google Doc survey containing the questions-in-blue above. See above link.

### Grading

The grade that you receive for this lab assignment will be based on the following:

- This activity is a check-mark grade.

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