CSC-314: Final Project Rubric

Option A Projects

Note: ALL PROGRAMS SUBMITTED MUST BE YOUR OWN. IF ANY PORTIONS OF YOUR PROGRAM ARE FOUND TO BE PLAGIARIZED, THE ASSIGNMENT WILL NOT BE ACCEPTED. If you have any questions about the acceptability of your code, or whether the use of packages or libraries is allowed, you should contact me.

	Poor (C or below)	Acceptable (B range)	Excellent (A range)
Documentation (all programs)	Code documentation is minimal or not provided. Variable and function names are not descriptive and the code is formatted poorly, making the program difficult to read.	The majority of functions and major code segments are documented. Variable and function names are chosen appropriately and proper formatting (such as indentation) is used in the majority of the code. The algorithm is easy to follow but some aspects of the code are not.	All functions and major code segments are properly documented. Variable and function names are chosen appropriately and proper formatting (such as indentation) is used, making the program logic easy to follow.
Translation Program	The program does not work correctly (sequences are not translated correctly)	The program works correctly but is not user-friendly. For example, the user cannot specify the name of the file.	The program works correctly. DNA/RNA sequences are read from a specified file (it is assumed that the sequence is written in the 5' to 3' direction). The program translates all 6 reading frames, and open reading frames are highlighted. The program is user-friendly and handles both lower- and upper-case nucleotide characters.
Pairwise Alignment Program	The program does not work correctly (the score of the optimal alignment is not correct and the optimal alignment cannot be found)	The program works correctly but is not user-friendly. For example, the alignment is correct but is not displayed in the standard format.	The program works correctly and is not case-sensitive. The user can upload two sequences stored in a single file or stored in separate files. A dynamic programming matrix is created in order to find the score of the optimal alignment and a traceback procedure is used to find and display the optimal alignment.
Gene Prediction Program	The program does not work correctly, and either does not identify likely genes, or mistakenly identifies regions that do not have all of the desired properties.	The program has small mistakes that prevent correct identification of all potential genes. For example, the program requires that the Shine-Dalgarno sequence matches exactly, rather than allowing mismatches.	The program is user friendly, works correctly, and is not case-sensitive. The user uploads a file containing one or more sequences in FASTA format. For each sequence, the program outputs the number of predicted genes, and for each predicted gene, the location and length of the CDS.

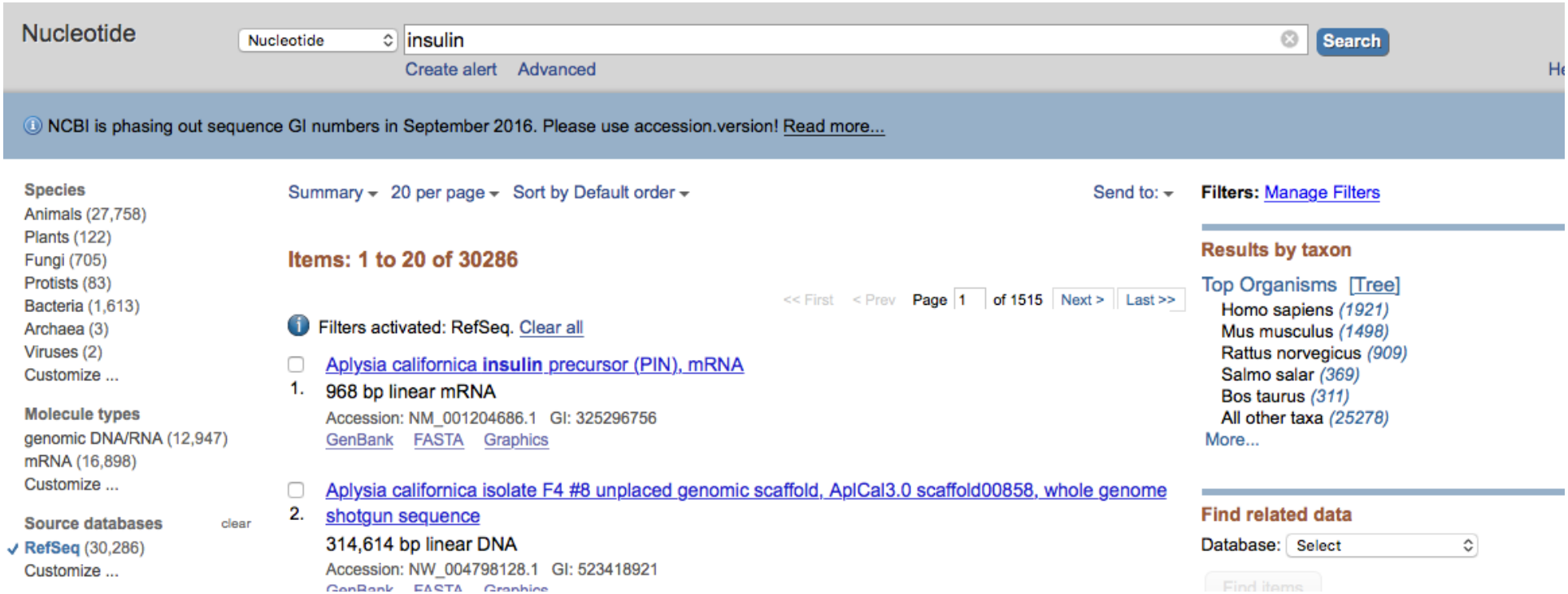
BLAST	The program does not work correctly.	The program works correctly but is not complete. For example, the organism and E-values are not correct.	The program works correctly, BLASTING the sequences using the appropriate parameters and outputting the organism, accession number, similarity, E-value, and alignments of the top 3 hits. No values in your program should be hardcoded (i.e., if I run the program with another sequence, I should get the correct results for that sequence). The sequences should be read from a single file or from separate files.

Option B Projects

	Poor (C or below)	Acceptable (B range)	Excellent (A range)
Questions	The majority of your questions have not been answered or are not correct	A small number of questions have not been answered or are not correct.	All or nearly all questions are answered correctly.
Write-up	Answers are not submitted in the form of a written report that forms a cohesive narrative.	The write-up is a cohesive narrative, but may include several spelling or grammatical mistakes.	The write-up is a cohesive narrative. Little or no spelling or grammatical mistakes are made.
Methods	No methods are included.	Almost all methods are included.	The write up describes all methods used to obtain your answers. These methods should be specific enough to be repeatable (i.e., so I can get the same answers as you; see below). You may choose to include a separate Methods section or integrate the methods with your answers (see below).
Screenshots	No screenshots are included	Almost all screenshots are included	Screen shots of all key results and methods are included, which includes but is not limited to: GenBank queries, GenBank or GenPept entries, OMIM entries, BLAST results, and GEO2R results.

Example question: How many RefSeq entries are there for the keyword "insulin"?

Example answer: There are 30,286 RefSeq entries for the keyword "insulin". This was determined by searching for "insulin" (without the quotes) from the GenBank website (https://www.ncbi.nlm.nih.gov/genbank/). From the side panel, selecting RefSeq under source databases yielded 30,286 results (see screenshot on next page).



Note for group projects:

1. For Option A projects, documentation must include the person responsible for writing the code (this will generally be at the function / module level)
2. For Option B projects, you must include a contribution section at the end which describes the contribution for each individual. In this section, it is customary to use initials (e.g., GD instead of Garrett Dancik). *Example:* GD retrieved sequences from GenBank, and performed the BLAST analysis. AFP analyzed the gene expression data using the Gene Expression Omnibus (GEO) and collected relevant information from the OMIM database.