

Yana Slavcheva
COS 4091A
Fall 2025
Project: DNA Sequence Visualizer and Analyzer
Prof. Zachary Hutchinson

Progress Report 4

This week, I implemented error handling for the FASTA file parser. It now properly checks not only for file opening failures but also validates the FASTA format by ensuring the presence of header lines and sequence data, and it verifies that all sequence characters are valid DNA bases. I had issues with file paths in Visual Studio, which I resolved by using the full file path directly in the code.

I then began implementing the Boyer-Moore algorithm, because it is the simplest of the three algorithms I plan to develop and also due to its efficiency with genomic data. The current algorithm successfully scans the loaded DNA sequence and outputs a list of every position where a given pattern is found. For demonstration, I made the program automatically search for common sequences such as the "ATG" start codon and the "TAA," "TAG," and "TGA" stop codons, and display the exact index and total count of each occurrence. At the end of the run, the program also allows the user to search for a specific DNA pattern of their choice. The user's input is capitalized and validated for valid nucleotides before the search is executed after which the exact positions and total count of matches are displayed again. The next step will be to refine the algorithm's error handling, which works relatively well at the moment but does not yet cover all possible edge cases.