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

Progress Report 3

Following our discussion, I considered your feedback and have refined the scope and technical direction of my senior project. I agree that relying solely on string-searching algorithms would not fully demonstrate the expected level of algorithmic complexity. The core idea of a DNA sequence visualizer will stay the same (with the string searching and ORF finder), however, I will implement more functionalities. I decided on Smith-Waterman algorithm for local sequence alignment which will find regions of similarity between two DNA sequences by building a scoring matrix and determining an optimal alignment path.

I continued establishing the core data structures. The central class is the DnaSequence, which is responsible for storing the raw sequence string parsed from a FASTA file and, importantly, a list of Feature objects. For the polymorphism, I defined Feature as an abstract base class, with specific classes like Gene and AlignmentRegion planned to inherit from it. Mainly I worked on the FastaParser class. The initial implementation handles the basic structure of a standard FASTA file, reading the input line-by-line, distinguishing header lines beginning with a > from sequence data lines, and linking the sequence data into a single string for the DnaSequence object. This parser is currently functional for well-formatted files, but it is not yet complete as I need to implement error handling. Additionally, I continued searching for more relevant literature.

https://github.com/yYana33/senior-project