|  |  |  |  |
| --- | --- | --- | --- |
| Weekly Project Report | | | |
| Project Title: | DNA Sequence Comparison Project | | |
| Developer: | Blaze Milner | | |
| Date: | 2/22/2019 | | |
| Period Covering: | 2/13/19 | To | 2/20/19 |
| Progress Highlights | | | |
| This week my program successfully compared an 8,000 long DNA sequence 100 bases at a time and compared each one against NCBI’s database. This process took roughly 12 hours. After meeting with Dr. de Banzie on the 20th, I discovered that my program does not actually need to partition the sequence but simply compare the entire sequence as a whole against NCBI. | | | |
| Dates of Specific Progress | | | |
| Date: | Progress: | | |
| 2/19 | My program was modified to print the starting and ending values of the matches. After this, I ran my program using a sequence Dr. de Banzie had given me. The sequence was roughly 8,000 nucleotides long that I partitioned at 100 bases. This entire process began a little before 5:00PM and did not finish until 9:30 AM the next day. After reviewing the data I collected, there were over 4,000 rows of information in my CSV file. I was discouraged by the amount of time this took but was reassured since this time was significantly less than the traditional 16-week timeline Dr. de Banzie’s students took. | | |
| 2/20 | I met with Dr. de Banzie and presented my program. After discussing the large amount of data collected and the time the program took, Dr. de Banzie clarified that my program did not actually need to partition the sequence at a variable length defined by the user. He only needed this feature if we took a different approach in collecting the matching sequence values, since biopython is able to collect the values then this feature does not need to be implemented. After removing the partitioning element, my program was significantly faster and completed in less than 5 minutes. I sent my program output to Dr. de Banzie who is currently reviewing it to see if the program is fulfilling all that he needs it to. | | |
| Activities to be Started Next Week | | | |
| Next week I hope to modify my program to be able to read a fasta file. Dr. de Banzie stores his DNA sequences in these types of files and it would be significantly easier for the user to upload the file rather than entering the sequence from the keyboard. | | | |