NGS uses a wide array of tools to obtain results based on the genome sequence. The most widely used tools are BLAST, HHMMPMER, and MUMMER. BLAST (Basic Local Alignment Search Tool) is a multi-sequence alignment tool developed by NH (National Institute of Health). It is used find similar regions in different sequences and then compare their similarities. MUMMER (Maximum Unique Matches) is a rapid alignment system used for rapidly aligning entire genomes. It can also align incomplete genomes and can easily handle 1000’s of contigs from a shotgun sequencing project. HHMMPMER (Hidden Markov Modeler) is used for searching sequence databases for homologs of protein sequences, and for making protein sequence alignments. It implements methods using probabilistic models called profile hidden Markov models (HMMs)

Two FASTA files related to the same nucleotide sequence were input into both BLAST and MUMMER and the results were parsed into tables. Then, the coverage of all hit contigs and nodes from both programs was found.

Using the data gathered from both BLAST and MUMMER, the frequency of the amount covered for each contig was plotted. From Fig 6, it can be inferred that MUMMER hit more accurately for contigs.

The same process was done with the Nodes. From Fig 7, it can be inferred that BLAST hit more accurately with nodes. However, there are more BLAST results with lower coverage.

Comparisons

Once the results were found using both the BLAST and MUMMER search tools, we created a program to see which sequencing tool had the most hits per contig. The total number of contigs in the database file is 160,749 and the total number of nodes in the query file is 552,305. BLAST returned a total of 123,070 hits and MUMMER returned a total of 121,829 hits. From the results, MUMMER hit more accurately than BLAST while BLAST hit more contigs than MUMMER.

Conclusion and Future Work

The future of next-generation sequencing can be broken down into a variety of categories such as personalized medicine, biofuels, climate change, and other life science fields.

- Personalized Medicine is a medical model that proposes the customization of medical decision to tailor an individual
- Bio Fuels present a source of alternative energy. Microbial biofuels use algae to synthesize the fuel. In order to optimize the process, an understanding of the gene-function relationship of algae would prove helpful.
- Climate change is the active study of past and future theoretical models which uses the past climate data to make future projections.

In conclusion, we hope to contribute the knowledge we have gained to contribute to fields such as these.

References