4-11-2023

Mapping next generation sequence data with BWA (Burrows-Wheel Aligner) on Galaxy software

Rabeh Z. Omar
Georgia Southern University

Mapping next generation sequence data with BWA (Burrows-Wheel Aligner) on Galaxy® software

An honors Thesis submitted in partial fulfillment of the requirements for Honors in *Biology*

By
Rabeh Omar

Under the mentorship of *Aaron Schrey*

Advancement of next-generation sequencing technologies introduces a vast amount of data which has become a challenge for researchers to organize and sequence data sets. BWA (Burrows-Wheeler Aligner) is one of the widely used software for aligning and mapping sequencing data against a reference genome. In my thesis, I present a comprehensive guide for analyzing genome sequences using BWA. I discuss the various steps involved in the process, including gathering the data, preparing the reference genome, aligning the sequences, and processing the data to visualize the results.

INDEX WORDS: BWA, Genome sequencing, Genetics, Next-Generation sequencing, DNA mapping,

Thesis Mentor: *Dr. Aaron Schrey*

Honors Dean: Dr. Steven Engel

April 2023
Department of Biology
Honors College

Georgia Southern University

Next-Generation sequencing (NGS) is an innovate method of sequencing millions of DNA fragments with the purpose of comparing genomes, identifying pathogens, and other means of advancing in genetic research. NGS involves genomic alignment which is the mapping of a genome sequence alongside a genomic reference, a representative digital genomic database (Alter 2017). Mapping refers the process of constructing maps that determine gene locations on chromosomes. This haploid representation of the genome was derived from a collection of volunteer donors for the use in bioinformatic research. My research consists of inputting genomic sequences into BWA, a fast short reader aligner with an unspliced mapper on the Galaxy application, for the purpose of analyzing a genome through next generation sequencing (NSG). This software is known for its speed, accuracy, and versatility in analyzing various sets of genomic data. Analyzing files involves gathering genomic data, uploading a genomic reference, and aligning the data with that genomic reference. The reference genome in my research consisted of the sequences from *Apis Mellifera,* the Western Honeybee. BWA consists of 3 algorithms that are implanted for different DNA base pair (bp) lengths: standard BWA backtrack (maps up to a 100 bp sequence), BWA-SES (maps 70-1 Mbp), and BWA MEM (new-generation analyzer that maps 70-1 Mbp) (Arasappan, 2022). In my research, the standard BWA was used with an illumina analyzer to map short reads that range from 20-190 bp. BWA backtrack tolerates sequencing error rates below 2% while BWA-SW and BWA-MEM tolerate higher rates due to the longer bp alignments (2022). The first step in analyzing genome sequences using BWA is to gather the data. In my research, data was premade and obtained through my research mentor. It is imperative to assure that the data obtained underwent quality

control and filtering of the raw version to target your appropriate genome sequence. The filtering phase removes low-quality reads which may involve reads that are too short or too long. The second step is to index the reference genome which involves generating or obtaining a preset index file for the reference genome. This reference genome is used by BWA to align the sequencing reads. Considering that reference genomes can be reused, the indexing phase needs to be done only once for a given reference genome. The third step involves aligning the sequencing reads against the previously presetted or obtained reference genome using BWA. BWA a transform feature to index the reference genome and align the reads.

References Cited

Mohammed Alser, Hasan Hassan, Hongyi Xin, Oğuz Ergin, Onur Mutlu, and Can Alkan. 2017. GateKeeper: a new hardware architecture for accelerating pre-alignment in DNA short read mapping. Bioinformatics 33, 21 (2017), 3355–3363.[1]

Arasappan, D. (n.d.). *Mapping with BWA*. UT Austin Wikis. Retrieved April 10, 2023, from https://wikis.utexas.edu/display/bioiteam/Mapping+with+BWA [2]

Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60.