Multi-threading the generation of Burrows-Wheeler Alignment

H. Jo

Department of Information Technology, Chonbuk National University, Republic of Korea

Corresponding author: H. Jo
E-mail: heeseung@jbnu.ac.kr

Received March 23, 2016
Accepted April 11, 2016
Published May 23, 2016
DOI http://dx.doi.org/10.4238/gmr.15028650

ABSTRACT. Along with recent progress in next-generation sequencing technology, it has become easier to process larger amounts of genome sequencing data at a lower cost. The most time-consuming step of next-generation sequencing data analysis involves the mapping of read data into a reference genome. Although the Burrows-Wheeler Alignment (BWA) tool is one of the most widely used open-source software tools for aligning read sequences, it still has a limitation in that it does not fully support a multi-thread mechanism during the alignment generation step. In this article, we propose a BWA-MT tool based on BWA that supports multi-thread mechanisms for processing alignment generation. To evaluate BWA-MT, we used an evaluation system equipped with 24 cores and 128 GB of memory. As workloads, we used the hg19 human genome reference sequence and sequences of various read sizes from the 1 to 40 M spots. In our evaluation, BWA-MT showed a maximum of 3.66-times better performance, and generated the same Sequence Alignment/Map result file as that of BWA. Although the ability to speed up the procedure might be dependent on computing resources, we confirmed that BWA-MT is a highly effective and fast alignment tool.

Key words: Genome sequencing; Next-generation sequencing; Burrow-Wheeler Aligner tool; Multi-thread