

Fast processing RNA-Seq on multicore processor

ABSTRACT

RNA Sequencing (RNA-Seq) is the sequencing and analysis of transcriptomes. The main purpose of RNA-Seq analysis is to find out the presence and quantity of RNA in an experimental sample under a specific condition. Essentially, RNA raw sequence data was massive. It can be as big as hundreds of Gigabytes (GB). This massive data always makes the processing time become longer and take several days. A multicore processor can speed up a program by separating the tasks and running the tasks' errands concurrently. Hence, a multicore processor will be a suitable choice to overcome this problem. Therefore, this study aims to use an Intel multicore processor to improve the RNA-Seq speed and analyze RNA-Seq analysis's performance with a multiprocessor. This study only processed RNA-Seq from quality control analysis until sorted the BAM (Binary Alignment/Map) file content. Three different sizes of RNA paired end has been used to make the comparison. The final experiment results showed that the implementation of RNA-Seq on an Intel multicore processor could achieve a higher speedup. The total processing time of RNA-Seq with the largest size of RNA raw sequence data (66.3 Megabytes) decreased from 317.638 seconds to 211.916 seconds. The reduced processing time was 105 seconds and near to 2 minutes. Furthermore, for the smallest RNA raw sequence data size, the total processing time decreased from 212.380 seconds to 163.961 seconds which reduced 48 seconds.

Keyword: Bioinformatics; High-performance computing; Multicore processors; RNA Sequencing