TEXAS TECH UNIVERSITY
Edward E. Whitacre Jr.
College of Engineering
Computer Science

Data-Intensive Scalable Computing Laboratory (DISCL)

Technical Report

Department of Computer Science

Texas Tech University

Parallel Algorithms Research on Burrows-Wheeler Transformation for Short Read Alignment

Jiang Zhou, Frank Conlon, Shengping Yang, Yong Chen

jiang.zhou@ttu.edu, frank.conlon@ttu.edu, shengping.yang@ttuhsc.edu, yong.chen@ttu.edu

June 2017

Technical Report № TTU/DISCL-2017-01

http://discl.cs.ttu.edu

Parallel Algorithms Research on Burrows-Wheeler Transformation for Short Read Alignment

Jiang Zhou, Frank Conlon, Shengping Yang*, Yong Chen

Department of Computer Science, Texas Tech University, Lubbock, TX
*Texas Tech University Health Sciences Center, Lubbock, TX

jiang.zhou@ttu.edu, frank.conlon@ttu.edu, shengping.yang@ttuhsc.edu, yong.chen@ttu.edu

Abstract

Mapping a patient's genome is an important and time-consuming part of today's cutting edge cancer research. Because cancer is a mutation of normal healthy cells, mapping the genome can provide a lot of information about what causes cancers and could lead us to potential cures. The trouble with mapping the human genome though is that a single human DNA strand contains massive amounts of information. Because of the large amounts of information, brute force methods of DNA analysis are ineffective and intractable. In order to process all of the information in DNA clever optimizations are required to make the data more manageable. Three of the most commonly used ways to do this are hashing, sorting, and the method that will be discussed here: Burrows-Wheeler transformations. Even with these optimizations, the volume of data still requires a large amount of time to process. This time can be significantly improved by parallelizing the algorithms. In this report we will provide a short background of the problem, a description of the Burrows-Wheeler algorithm, and discuss how parallelization can be used to improve analysis speed.

Keywords: Burrows-Wheeler algorithm, parallel, cancer