Abstract

Parallel DNA Read Alignment Using the Amazon Cloud

By Kayleigh McGinley | Supervised by Jeremy Jones

Trinity College Dublin | School of Computer Science & Statistics

Master in Computer Science

Short-read alignment is the process of searching for short sequences of DNA within a species' entire set of genes. Many short-read alignment programs exist, such as BWA[1], SOAP2[2] and Bowtie[3]. However, these programs all have one thing in common; they use a single reference genome. The use of a single reference genome can lead to inherent biases and lower accuracy. BWBBLE was created to map short reads to a collection of genomes (a reference multi-genome) with high accuracy. It handles genetic variants thus avoiding the inherent bias to one specific genome[4].

One major concern with BWBBLE is that it is up to 100 times slower than other read aligners due to the larger amount of data it processes. The aim of this project is to introduce a new version of BWBBLE, called AWS-BWBBLE that uses Amazon Web Services to distribute the work amongst a number of virtual machines. A small distributed network was successfully created in AWS using Elastic Compute Cloud instances (VMs) and Elastic File Storage. The parallelization is achieved by instructing each VM in the network to process a different subset of the reads file. This straightforward approach was a complete success as proved by the linear speed-up of the program using up to four VMs.

[1] LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology, 10.
[2] LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England), 25.
[3] LI, R., YU, C., LI, Y., LAM, T.-W., YIU, S.-M., KRISTIANSEN, K. & WANG, J. 2009. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics (Oxford, England), 25.
[4] HUANG, L. & POPIC, V. 2013. Short read alignment with populations of genomes. Bioinformatics (Oxford, England), 29.