Table of contents

The CODATA-RDA Research Data Science Applied workshops on Extreme sources of data, Bioinformatics and IoT/Big-Data Analytics | (smr 3178)

Wednesday 26 July 2017

Bioinformatics - Adriatico Guest House - Informatics Laboratory (09:00-18:00)

Abstract: DNA aligners (such as BLAST, Bowtie or BWA) are very fast tools that allow searching occurrences of (short) DNA sequences in one or more (big) genomes. The idea behind these tools is to pre-process the genome file and build an index; such an index permits to search a DNA sequence in time proportional to its length, rather than to the length of the genome. Indexing accelerates DNA alignment by millions of times, but it introduces a problem: the index could be several times bigger than the text, exceeding the computer's RAM size. This is particularly concerning in view of recent developments in DNA sequencing technologies: projects such as the 1000 Genomes Project are producing thousands of sequenced genomes, which should be indexed in order to quickly align DNA sequences on them. Not all hope is lost, however. Two genomes from the same species are 99.99% identical, so compression techniques can be exploited to greatly reduce the index size. In this lecture I will introduce a famous compression and indexing technique that is having a huge impact in bioinformatics: the Burrows-Wheeler transform (BWT). We will see - both in theory and practice - how BWT-based aligners can achieve extremely high search speeds while taking (up to) thousands of times less space than the input collection of genomes.

time	title	presenter
09:00	Aligning DNA sequences on compressed collections of genomes	PREZZA, Nicola
10:30	Coffee break	
11:00	Aligning DNA sequences on compressed collections of genomes	PREZZA, Nicola
12:30	Lunch break	
14:00	Aligning DNA sequences on compressed collections of genomes	PREZZA, Nicola
15:30	Coffee break	
16:00	Lab Session: Aligning DNA sequences on compressed collections of genomes	PREZZA, Nicola