



UNIVERSITÀ DEGLI STUDI DI MILANO - BICOCCA

Scuola di Scienze

Dipartimento di Informatica, Sistemistica e Comunicazione

Corso di Laurea Magistrale in Informatica

Algoritmi per la trasformata di Burrows-Wheeler Posizionale con compressione run-length, RLPBWT

Relatore: *Prof.ssa Raffaella Rizzi*

Correlatore:

Tesi di Laurea Magistrale di:

Davide Cozzi

Matricola 829827

Anno Accademico 2021-2022

Abstract

Indice

1	Introduzione	3
2	Preliminari	4
2.1	Motivazioni Biologiche	4
2.2	Trasformata di Burrows-Wheeler	4
2.2.1	Trasformata di Burrows-Wheeler run-length	4
2.3	Trasformata di Burrows-Wheeler posizionale	4
2.3.1	Implementazione originale	4
2.3.2	Varianti	4
2.4	Variant Calling Format	4
3	Metodo	5
3.1	Introduzione agli strumenti usati	5
3.2	Costruzione della RLPBWT	5
3.2.1	RLPBWT con divergence array	5
3.2.2	RLPBWT senza divergence array	5
3.2.3	Implementazione dell'uv-trick	5
3.2.4	Algoritmo per match massimali	5
4	Risultati	6
4.1	Ambiente di benchmark	6
4.2	Analisi temporale	6
4.2.1	Confronto con implementazione originale e varianti	6
4.3	Analisi spaziale	6
4.3.1	Confronto con implementazione originale e varianti	6
5	Conclusioni	7
5.1	Sviluppi futuri	7
	Bibliografia e sitografia	7

Capitolo 1

Introduzione

Capitolo 2

Preliminari

2.1 Motivazioni Biologiche

2.2 Trasformata di Burrows-Wheeler

2.2.1 Trasformata di Burrows-Wheeler run-length

MONI

2.3 Trasformata di Burrows-Wheeler posizionale

2.3.1 Implementazione originale

Gli algoritmi di Durbin

2.3.2 Varianti

PBWT multiallelica

PBWT con struttura LEAP

PBWT dinamica

PBWT bidirezionale

2.4 Variant Calling Format

Capitolo 3

Metodo

3.1 Introduzione agli strumenti usati

3.2 Costruzione della RLPBWT

3.2.1 RLPBWT con divergence array

3.2.2 RLPBWT senza divergence array

3.2.3 Implementazione dell'uv-trick

3.2.4 Algoritmo per match massimali

Caso con divergence array

Caso senza divergence array

Capitolo 4

Risultati

4.1 Ambiente di benchmark

4.2 Analisi temporale

4.2.1 Confronto con implementazione originale e varianti

4.3 Analisi spaziale

4.3.1 Confronto con implementazione originale e varianti

Capitolo 5

Conclusioni

5.1 Sviluppi futuri