

Indice

Introduzione	5
1 BWT, eBWT, PFP	9
1.1 Definizioni preliminari	9
1.2 Trasformata di Burrows-Wheeler	12
1.3 Estensione della trasformata di Burrows-Wheeler	15
1.4 Costruzione della BWT con PFP	17
1.4.1 Costruzione del dizionario e del parse	18
1.4.2 Costruzione della BWT(P)	19
1.4.3 Costruzione della BWT(T)	21
1.5 PFP come struttura dati	24
1.5.1 Strutture dati del PFP-CST	25
1.5.2 Implementazione delle primitive	27
2 Misure di dissimilarità	31
2.1 Approcci alignment-based e alignment-free	31
2.2 Misura basata sull'alternanza di colori nella eBWT	32
2.3 Misura basata sulle run della eBWT	35
2.4 Misure basate sulla Burrows-Wheeler similarity distribution	36
2.5 Normalizzazione delle distanze	38
3 Costruzione della eBWT e del DA con PFP	41
3.1 Costruzione del dizionario e del parse	41
3.2 Generalized Conjugate Array	43
3.3 SAIS-for-eBWT	43
3.4 Costruzione della eBWT(P)	49
3.5 Costruzione del DA di eBWT(P)	50
3.6 Inverted list	51
3.7 Strutture dati ausiliarie	51
3.8 Costruzione della eBWT(M)	53
3.9 Costruzione del DA di eBWT(M)	55
3.10 Costruzione del vettore di indici I	55

4	Risultati sperimentali	57
4.1	Progettazione	57
4.2	Dataset	58
4.3	Alberi di filogenesi	58
	Conclusioni e futuri lavori	75