

DETEKSI GENOMIC REPEATS MENGGUNAKAN ALGORITMA KNUTH-MORRIS-PRATT PADA R HIGH-PERFORMANCE COMPUTING PACKAGE

Oleh

Ahmad Banyu Rachman — banyurachman@outlook.com

1303465

ABSTRAK

Proses pencarian *pattern* pada *string* untuk menemukan perulangan pasang basa pada sebuah sekuens *Deoxyribonucleic Acid* (DNA) membutuhkan waktu yang cukup lama. Penelitian ini membuat sebuah model komputasi paralel untuk melakukan sebuah pencarian *pattern* pada *string* dengan memodifikasi dan mengimplementasikan algoritma Knuth-Morris-Pratt pada R *High-Performance Computing package* bernama pbdMPI untuk data sekuens DNA manusia yang diunduh dari lama resmi Ensembl. Hasil penelitian ini menunjukkan adanya percepatan yang sangat signifikan antara penggunaan *stand alone* dan penggunaan *parallel computing* dengan 2 *cores*, 4 *cores*, 8 *cores*. Penelitian ini juga membuktikan bahwa jumlah *iterator* yang semakin besar tidak beriringan dengan waktu komputasi yang lebih baik pula.

Kata Kunci: *DNA, human genom, genomic repeats, string matching, knuth-morris-pratt, high-performance computing*

**GENOMIC REPEATS DETECTION BY USING KNUTH-MORRIS-PRATT
ALGORITHM ON R HIGH-PERFORMANCE PACKAGE**

Arranged by

Ahmad Banyu Rachman — banyurachman@outlook.com

1303465

ABSTRACT

Pattern searching in string process to find repeating base pairs in Deoxyribonucleic Acid (DNA) sequences take a long time. This research builds a parallel computing model to search pattern in string by modifies and implements Knuth-Morris-Pratt algorithm on R High-Performance Computing Package pbdMPI for human DNA sequences from Ensemble site. The result shows a significant acceleration between stand alone and parallel computing by 2 cores, 4 cores and 8 cores. This research proves that the higher value of iterator is not equal with better computation time.

Keywords: DNA, human genom, genomic repeats, string matching, knuth-morris-pratt, high-performance computing