

INTISARI

MULTIPLE SEQUENCE ALIGNMENT SECARA PARALEL MENGGUNAKAN MPI (MESSAGE PASSING INTERFACE)

Fathurrochman Habibie
13/347559/PA/15304

Seiring dengan berkembangnya teknologi dibidang bioinformatika, biaya untuk melakukan *sequencing* DNA dari tahun ke tahun semakin murah. Hal ini menyebabkan pertumbuhan basis data genetik naik bahkan melampaui hukum *Moore's law*. Pertumbuhan basis data genetik yang sangat cepat merupakan salah satu hambatan utama dalam melakukan *sequence alignment*. ClustalW merupakan salah satu metode MSA yang sering digunakan oleh praktisi biomolekuler. Metode ini dibagi menjadi tiga tahap yaitu perhitungan *distance matrix*, pembuatan *guided tree* dan *progressive alignment*.

Pemrograman sekuensial membutuhkan waktu yang lama untuk memproses data yang besar. Selain itu, pemrograman sekuensial memiliki keterbatasan *memory*, sehingga dapat menyebabkan *stack* pada program. Salah satu cara untuk mempercepat kinerja pemrosesan adalah menggunakan pemrograman paralel. MPI merupakan salah satu teknologi komputasi paralel yang populer. Proses paralel dijalankan pada kluster MPI yang terdiri dari empat buah komputer *single-board*. Percobaan menggunakan data sekuens dari BALiBase versi 3. Dari hasil penelitian menunjukkan bahwa pada tahap perhitungan *distance matrix* dapat mencapai waktu 12,7 kali, sedangkan pada tahap *progressive alignment* dapat mencapai waktu 5,71 kali lebih cepat dibanding proses sekuensialnya.

Kata Kunci : Paralelisasi, Multiple Sequence Alignment, ClustalW, Message Passing Interface

ABSTRACT

MULTIPLE SEQUENCE ALIGNMENT IN PARALLEL USING MPI (MESSAGE PASSING INTERFACE)

Fathurrochman Habibie
13/347559/PA/15304

Due to the rapid growth of bioinformatics technology, the cost of sequencing DNA from year to year is cheaper. This led to the growth of genetic databases to rise even beyond Moore's law. The rapid growth of genetic databases is one of the problems with sequence alignment. ClustalW is a popular method used by biomolecular practitioners. It is divided into three steps. There are distance matrix calculations, guided tree reconstruction, and progressive alignment.

Sequential programming takes a long time to process large dataset. Furthermore, sequential programming has limited memory, so it can cause a stack on the program. One way to speed up the performance is to use parallel programming. MPI is one of the most popular parallel computing technologies. The parallel process is run on the MPI cluster consisting of four single-board computers. The experiments used sequence data from BALiBase version 3. The results obtained that the distance matrix calculation can reach 12.7 times faster than sequential process. While at the progressive alignment can reach time 5.71 times faster than sequential process.

Keyword : *Parallelization, Multiple Sequence Alignment, ClustalW, Message Passing Interface*