

**MODELING OF  
DNA SEQUENCE WITH FILTERING-RANKING  
ALIGNMENT (F-R-A) TECHNIQUE USING PETRI NET**

**MASTURA MUHAMAD**

**DEGREE OF MASTER OF SCIENCE  
UNIVERSITY MALAYSIA TERENGGANU  
MALAYSIA**

**2010**

1100083433

Perpustakaan Sultanah Nur Zahirah  
Universiti Malaysia Terengganu (UMT)



tesis

TA 347 .N46 M3 2010



1100083433

Modeling of DNA sequence with filtering-ranking- alignment  
(F-R-A) technique using petri net / Mastura Muhamad.

PERPUSTAKAAN SULTANAH NUR ZAHIRAH  
UNIVERSITI MALAYSIA TERENGGANU (UMT)  
21000 KUALA TERENGGANU

1100083433

1100083433		

Lihat sebelah

HAK MILIK  
PERPUSTAKAAN SULTANAH NUR ZAHIRAH UMT

01010

**MODELING OF  
DNA SEQUENCE WITH FILTERING-RANKING-ALIGNMENT  
(F-R-A) TECHNIQUE USING PETRI NET**

**MASTURA MUHAMAD**

November 2010

**MASTURA MUHAMAD**

Thesis Submitted in Fulfillment of the Requirement for the  
Degree of Master of Science in the Faculty of Science and  
Technology  
University Malaysia Terengganu  
Malaysia

November 2010

828880011

## **ABSTRACT**

Abstract of thesis presented to the Senate of University Malaysia Terengganu in fulfillment of the requirement for the Degree of Master of Science.

### **MODELING OF DNA SEQUENCE WITH FILTERING-RANKING-ALIGNMENT (F-R-A) USING PETRI NET**

**MASTURA MUHAMAD**

**November 2010**

**Chairperson : Prof. Md Yazid Mohd Saman, Ph. D.**

**Member : Mustafa Man**

**Faculty : Science and Technology**

The growing advancement in bioinformatics requires for better framework and algorithms to deal with the complex and extensive biological data. DNA sequence comparison is one of the most important algorithms considered in computational biology. Nevertheless, the complexity of bioinformatics leads to problems in expanding the field. Existing systems and algorithms are quite hard to be understood. It is difficult to predict what happen if certain action is taken in order to modify or improve the existing system.

There is a need for an efficient method to understand existing systems and algorithms in order to improve or expand the system. Modeling develops a level of understanding regarding the interaction within the parts of a system

and the system as a whole. Providing a graphical representation helps in identifying the different elements of a process and understanding the interrelationships among the various steps. Graphical representation of system resources and process acts as a trial platform for any changes planned for the real system.

Development of DNA sequence comparison is an active research activity in computational biology application. Filtering-Ranking-Alignment (F-R-A) model is a new model that improves the speed of large scale DNA sequence similarity search and at the same time preserves the best possible alignment result [11]. It has been developed due to the importance of comparison of genome sequences (DNA, mRNA and protein) in the area of computational biology.

Focusing on matching DNA string sequence, F-R-A model involves several algorithms includes Aho Corasick Algorithm. Aho Corasick Algorithm is an automaton-based algorithm that uses a keyword tree to process multi-pattern string matching. It is an efficient algorithm to locate all occurrences of any of finite number of keyword in string of text [61]. Aho Corasick Algorithm is very beneficial in fast sequence searching. Well understanding of the model may result in improvement of the searching technique.

Modeling the F-R-A model and Aho Corasick Algorithm has been done using Petri net. Petri net interprets the model and algorithm into places, transitions, arcs and token to simulate the dynamic activities of the techniques. As a graphical tool, it can be used as a visual-communication aid. Platform Independent Petri net Editor 2 (PIPE 2) which provides graphical modeling editor complete with analysis has been chosen as tool.

### RASYDA MUHAMAD

It can be concluded that modeling tool may simplify the understanding of complex processes in bioinformatics. Modeling the F-R-A model and Aho Corasick Algorithm has been described using Petri net. It is a graphical and mathematical modeling tool. With a simple structure, Petri net is very effortless to employ and understand.

## **ABSTRAK**

Abstrak tesis yang dikemukakan kepada Senat Universiti Malaysia Terengganu sebagai memenuhi keperluan untuk Ijazah Master Sains

### **PEMODELAN JUJUKAN DNA DENGAN TEKNIK *FILTERING-RANKING-ALIGNMENT (F-R-A)* MENGGUNAKAN PETRI NET**

**MASTURA MUHAMAD**

**November 2010**

**Pengerusi : Prof. Md Yazid Mohd Saman, Ph. D.**  
**Ahli : Mustafa Man**  
**Fakulti : Sains dan Teknologi**

Kemajuan Bioinformatik memerlukan rangka kerja dan algoritma yang lebih baik untuk mengendalikan data biologi yang luas dan kompleks. Perbandingan jujukan DNA adalah salah satu algoritma yang dianggap paling penting dalam pengkomputeran biologi. Namun demikian, kompleksiti bioinformatik menimbulkan masalah dalam mengembangkan bidang. Sistem dan algoritma yang sedia ada cukup sukar untuk difahami. Sulit untuk menjangka apa yang terjadi jika tindakan tertentu diambil dengan tujuan untuk mengubah atau meningkatkan sistem yang ada.

Kaedah yang berkesan adalah perlu untuk memahami sistem dan algoritma sedia ada untuk meningkatkan atau mengembangkan sistem. Pemodelan

meningkatkan tahap pemahaman tentang interaksi bahagian-bahagian dari sistem dan sistem secara keseluruhan. Gambaran grafik membantu dalam mengenalpasti unsur-unsur yang berbeza dari sebuah proses dan memahami kaitan antara pelbagai langkah. Perwakilan grafik bagi sumber dan proses sistem bertindak sebagai platform percubaan untuk perubahan yang dirancang untuk sistem sebenar.

Pembangunan perbandingan urutan DNA adalah kegiatan penyelidikan yang aktif dalam aplikasi biologi perkomputeran. Model *Filtering-Ranking-Alignment (F-R-A)* merupakan model baru yang meningkatkan kelajuan carian urutan skala besar kesamaan DNA dan sekaligus mengekalkan keputusan pelurusan terbaik [11]. Ia telah dibangunkan kerana pentingnya perbandingan urutan genom (DNA, mRNA dan protein) dalam bidang biologi perkomputeran.

Fokus pada persamaan urutan jujukan DNA, model *F-R-A* melibatkan beberapa algoritma termasuk Algoritma *Aho Corasick*. Algoritma *Aho Corasick* adalah mesin algoritma yang berpusat yang menggunakan pohon kata kunci untuk proses pepadanan jujukan multi-pola. Ini merupakan algoritma yang efisien untuk mencari semua kemungkinan dari setiap jumlah kata kunci dalam teks jujukan[61]. Algoritma *Aho Corasick* sangat bermanfaat dalam pencarian urutan yang pantas. Pemahaman yang bagus tentang model boleh meningkatkan teknik pencarian.



Pemodelan model *F-R-A* dan algoritma *Aho Corasick* telah dilakukan dengan menggunakan *Petri net*. *Petri net* menafsirkan model dan algoritma kepada *place*, *transition*, *arc* dan *token* untuk mensimulasikan kegiatan dinamik bagi teknik tersebut. Sebagai alat grafik, ia boleh digunakan sebagai bantuan komunikasi visual. Platform *Independent Petri net Editor 2 (PIPE 2)* yang menyediakan editor grafik lengkap dengan analisis telah dipilih sebagai alat.

Boleh disimpulkan bahawa alat pemodelan boleh memudahkan pemahaman proses yang kompleks dalam bioinformatik. Pemodelan model *F-R-A* dan algoritma *Aho Corasick* telah diterangkan menggunakan *Petri net*. Ia merupakan alatan pemodelan grafik dan matematik. Dengan struktur yang mudah, *Petri net* sangat mudah digunakan dan difahami.