



UNIVERSITAS INDONESIA

**PERANCANGAN STRUKTUR
MODEL *HIDDEN* MARKOV UNTUK PREDIKSI
EKSON DNA GEN *Plasmodium falciparum***

DISERTASI

SUHARTATI AGOES

84 02 03 0024

**PROGRAM STUDI TEKNIK ELEKTRO
PASCASARJANA DEPARTEMEN TEKNIK ELEKTRO
FAKULTAS TEKNIK UNIVERSITAS INDONESIA**

2008

LEMBAR PERSETUJUAN

Disertasi dengan judul

PERANCANGAN STRUKTUR MODEL *HIDDEN MARKOV* UNTUK PREDIKSI EKSON DNA *Plasmodium falciparum*

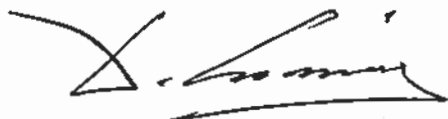
Disusun oleh : Suhartati Agoes
Nomor Mahasiswa : 8402030024
Program Studi : Teknik Elektro

Disusun untuk melengkapi persyaratan program Doktor Pascasarjana
Departemen Teknik Elektro Universitas Indonesia guna memperoleh gelar
Doktor

Penelitian ini disetujui untuk diajukan dalam Sidang Promosi

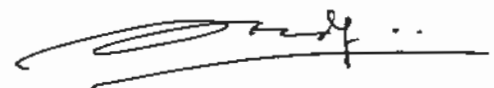
Depok, ²⁴/₃ 2008
Menyetujui,

Ko Promotor



Prof. Dr. Ir. Dadang Gunawan, M Eng
NIP. 131 475 421

Promotor



Prof. Dr. Ir. Sardy S. M Eng Sc
NIP. 130 517 309

ABSTRAK

Pemrosesan sinyal genom (*Genomic Signal Processing*) seperti Deoxyribonucleid Acid (DNA) dan protein dapat dilakukan untuk memprediksi ekson atau *coding region* suatu gen. Metoda yang paling banyak digunakan adalah *Hidden Markov Model* (HMM) yang kaya akan struktur matematik dan berpotensi untuk mengetahui lebih banyak tentang sumber sinyal tanpa harus tersedia sumber tersebut.

Pada penelitian ini dirancang struktur model yang menggunakan metoda HMM dengan struktur dasar model sesuai struktur ekson pada *coding sequence* (CDS) sehingga dapat memprediksi ekson DNA *Plasmodium falciparum*. Jumlah *state* model pada struktur dasar adalah 5, 7 dan 9 sedangkan untuk struktur pengembangan ditentukan secara acak yaitu 20, 30, 50 dan 100 *state*. Proses *training* HMM menggunakan algoritma Viterbi dan proses *testing* HMM menggunakan kedua algoritma yaitu Viterbi dan Baum-Welch sedangkan kinerja model menggunakan parameter *Correlation Coefficient* (CC). Sekuen yang digunakan adalah 152 sekuen DNA *Plasmodium falciparum* dengan panjang minimum 684 pb dan maksimum 10095 pb.

Hasil simulasi pada umumnya menghasilkan nilai CC rata-rata lebih baik dengan menggunakan algoritma Viterbi dibandingkan dengan algoritma Baum-Welch. Pada struktur dasar model 9 *state* menghasilkan nilai CC paling baik dibandingkan dengan struktur dasar model lainnya yaitu 0,7289 dengan menggunakan algoritma Viterbi dan 0,7166 dengan menggunakan algoritma Baum-Welch. Sedangkan untuk pengembangan model diperoleh nilai CC rata-rata paling baik untuk Model 2 dengan 100 *state* yaitu 0,7827 dengan menggunakan algoritma Baum-Welch dan 0,7820 dengan menggunakan algoritma Viterbi. Waktu proses *testing* HMM rata-rata seluruh model hampir dua kali lebih lama dengan algoritma Baum-Welch dibandingkan dengan algoritma Viterbi.

ABSTRACT

Genomic signal processing like as Deoxyribonucleid Acid (DNA) and protein can be done for exon prediction or coding region of the gene. The most used method is Hidden Markov Model (HMM) which has various mathematical structures and potentially capable of learning a great deal of signal source without having to have the source available.

The model structure designed in this research is using the HMM method with based structure model in accordance with exon structure in the coding sequence (CDS) in order to predict of DNA *Plasmodium falciparum*. The state number of model in the basic structure are 5, 7 and 9 states, meanwhile the expansions structure was randomly defined having 20, 30, 50 and 100 states. The HMM training process are using the Viterbi algorithm and the HMM testing are using both algorithms, Viterbi and Baum-Welch, meanwhile the performance indicator of the model are using the Correlation Coefficient (CC). It is using 152 sequences of DNA *Plasmodium falciparum* with the minimum length of 684 base-pair (bp) and maximum length of 10095 bp.

In general, the simulation results produced the best average of CC value by using Viterbi algorithm rather then Baum-Welch. In the basic structure of 9 states model produced the best CC value compared with the other basic structure models with 0.7289 using the Viterbi Algorithm and 0.7166 using Baum-Welch. Meanwhile, for the model expansion, the best CC average value is for Model 2 with state number 100 with 0.7827 using the Baum-Welch algorithm and 0.7820 using Viterbi. The average processing time of the HMM tests for all models using the Baum-Welch algorithm are almost two times slower than using Viterbi algorithm.

PERNYATAAN KEASLIAN PENELITIAN

Bersama ini,

Nama : Suhartati Agoes

Nomor Mahasiswa : 8402030024

Mahasiswa Program Pascasarjana Departemen Teknik Elektro Universitas Indonesia,

MENYATAKAN

Bahwa disertasi yang saya buat adalah karya saya sendiri dan bukan merupakan duplikasi sebagian atau seluruhnya dari karya orang lain, kecuali yang disebutkan sumbernya sebagai acuan.

Depok, Maret 2008

Suhartati Agoes

DAFTAR ISI

| | Halaman |
|---|---------|
| LEMBAR PERSETUJUAN | i |
| ABSTRAK | ii |
| ABSTRACT | iii |
| PERNYATAAN KEASLIAN PENELITIAN | iv |
| DAFTAR ISI | v |
| DAFTAR TABEL | viii |
| DAFTAR GAMBAR | ix |
| DAFTAR SINGKATAN | xi |
| DAFTAR ISTILAH | xii |
| | |
| BAB 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 2 |
| 1.2 Tujuan Penelitian | 13 |
| 1.3 Batasan Masalah | 13 |
| 1.4 Metodologi Penelitian | 14 |
| 1.5 Kontribusi Penelitian | 15 |
| 1.6 Sistematika Penulisan dan Penelitian | 16 |
| | |
| BAB 2 GENOM: STRUKTUR DAN ORGANISASI | 18 |
| 2.1 Deoxyribonucleic Acid (DNA) | 19 |
| 2.2 Ribonucleic Acid (RNA) | 20 |
| 2.3 Ekspresi Gen | 21 |
| 2.4 Ekson dan Intron | 23 |
| 2.5 Reading Frame | 28 |
| 2.6 Organisasi Genom <i>Plasmodium falciparum</i> | 29 |
| | |
| BAB 3 MODEL HIDDEN MARKOV | 32 |
| 3.1 Rantai Markov | 32 |
| 3.2 Elemen Model Hidden Markov | 33 |

| | | |
|---|---|-----------|
| 3.3 | Permasalahan Dasar HMM | 34 |
| 3.3.1 | Evaluasi | 34 |
| 3.3.2 | Decoding | 35 |
| 3.3.3 | Training | 35 |
| 3.4 | Solusi untuk setiap permasalahan dasar HMM | 35 |
| 3.4.1 | Solusi Evaluasi | 35 |
| 3.4.2 | Solusi Decoding | 41 |
| 3.4.3 | Solusi Training | 44 |
| BAB 4 PERANCANGAN HMM DAN SIMULASI | | 47 |
| 4.1 | Perancangan Struktur Model | 47 |
| 4.1.1 | Struktur Dasar Model 5 State | 47 |
| 4.1.2 | Struktur Model 8 State | 48 |
| 4.1.3 | Struktur Model 14 State | 48 |
| 4.1.4 | Pengembangan Struktur Dasar Model | 51 |
| 4.2 | Perancangan Simulasi | 54 |
| 4.2.1 | Input Sekuen DNA | 54 |
| 4.2.2 | Transisi State dan Emisi State | 58 |
| 4.2.3 | HMM Training | 60 |
| 4.2.4 | HMM Testing | 61 |
| 4.2.5 | Keluaran Simulasi | 62 |
| 4.2.6 | Spesifikasi Perangkat Simulasi | 63 |
| 4.2.7 | Diagram Alir Simulasi | 63 |
| BAB 5 PENGUJIAN DAN ANALISIS | | 66 |
| 5.1 | Pengujian | 66 |
| 5.1.1 | Pengujian Data Sekuen | 66 |
| 5.1.2 | Pengujian Label State Sekuen | 66 |
| 5.1.3 | Pengujian Emisi State | 69 |
| 5.2 | Hasil Pengujian Struktur Model dan Analisis | 70 |
| 5.2.1 | Pengujian Struktur Model | 70 |
| 5.2.2 | Analisis Pengujian Struktur Model | 71 |
| 5.2.3 | Pengujian Struktur Dasar Model | 71 |

| | | |
|----------------------------|--|------------|
| 5.2.4 | Analisis Pengujian Struktur Dasar Model | 76 |
| 5.2.5 | Pengujian dan Analisis Pengembangan Struktur Dasar Model | 78 |
| 5.2.6 | Pengujian Waktu Proses Simulasi | 83 |
| 5.2.7 | Analisis Waktu Proses Simulasi | 83 |
| BAB 6 | KESIMPULAN | 87 |
| 6.1 | Kesimpulan | 87 |
| 6.2 | Penelitian Lanjut | 88 |
| REFERENSI | | 89 |
| LAMPIRAN 1 | Data Sekuen DNA <i>Plasmodium falciparum</i> | 92 |
| LAMPIRAN 2 | Pengujian Data Sekuen | 100 |
| LAMPIRAN 3 | Hasil Uji Coba Struktur Dasar Model 5 State | 103 |
| LAMPIRAN 4 | Hasil Uji Coba Struktur Dasar Model 7 State | 106 |
| LAMPIRAN 5 | Hasil Uji Coba Struktur Dasar Model 9 State | 108 |
| UCAPAN TERIMA KASIH | | 110 |
| BIODATA | | 112 |

DAFTAR TABEL

| | Halaman |
|--|---------|
| Tabel 2.1 Kode Genetik | 24 |
| Tabel 5.1 Nilai CC rata-rata Uji Coba Struktur Model | 70 |
| Tabel 5.2 Nilai CC rata-rata Struktur Dasar Model 5 State | 73 |
| Tabel 5.3 Nilai CC rata-rata Struktur Dasar Model 7 State | 77 |
| Tabel 5.4 Nilai CC rata-rata Struktur Dasar Model 9 State | 77 |
| Tabel 5.5 Nilai CC rata-rata Struktur Dasar Model | 75 |
| Tabel 5.6 Nilai CC rata-rata Model 1 | 84 |
| Tabel 5.7 Nilai CC rata-rata Model 1 untuk jumlah ekson tertentu | 84 |
| Tabel 5.8 Nilai CC rata-rata Model 2 | 84 |
| Tabel 5.9 Nilai CC rata-rata Model 2 untuk jumlah ekson tertentu | 85 |
| Tabel 5.10 Nilai CC rata-rata Model 3 | 85 |
| Tabel 5.11 Nilai CC rata-rata Model 3 untuk jumlah ekson tertentu | 85 |
| Tabel 5.12 Nilai CC rata-rata Model 4 | 86 |
| Tabel 5.13 Nilai CC rata-rata Model 4 untuk jumlah ekson tertentu | 86 |
| Tabel 5.14 Waktu Proses <i>training</i> dan <i>testing</i> Model 1 ~ Model 4 | 86 |

DAFTAR GAMBAR

| | | Halaman |
|-------------|--|---------|
| Gambar 1.1 | <i>Gel image</i> sekuen DNA | 3 |
| Gambar 1.2 | Model sinyal sekuen DNA | 4 |
| Gambar 1.3 | Model sinyal sekuen DNA (<i>encoded</i>) | 4 |
| Gambar 1.4 | Model sinyal sekuen DNA (<i>non-encoded</i>) | 5 |
| Gambar 1.5 | Kode genetik pada bidang kompleks | 6 |
| Gambar 1.6 | Spektrum <i>coding</i> DNA | 7 |
| Gambar 1.7 | Spektrum daya DNA | 8 |
| Gambar 1.8 | Pemrosesan sekuen DNA dengan metode HMM | 9 |
| Gambar 1.9 | Struktur model HMM dengan sistem VEIL untuk penemuan suatu gen | 10 |
| Gambar 1.10 | Sistematika Kerangka Penelitian | 17 |
| Gambar 2.1 | Struktur DNA <i>double helix</i> | 20 |
| Gambar 2.2 | Urutan sintesis protein | 23 |
| Gambar 2.3 | Struktur gen eukariot | 26 |
| Gambar 2.4 | Data salah satu gen <i>P.falciparun</i> NC_004330 | 26 |
| Gambar 2.5 | <i>Plasmodium. falciparum</i> | 29 |
| Gambar 3.1 | Transisi <i>state</i> ekson dan intron | 33 |
| Gambar 3.2 | HMM dengan tiga <i>state</i> dan lima buah simbol observasi | 36 |
| Gambar 3.3 | Diagram <i>trellis</i> untuk seluruh kemungkinan transisi <i>state</i> | 36 |
| Gambar 3.4 | Ilustrasi tahap induksi pada prosedur <i>forward</i> | 39 |
| Gambar 3.5 | Ilustrasi tahap induksi pada prosedur <i>backward</i> | 40 |
| Gambar 3.6 | Ilustrasi prosedur <i>forward-backward</i> | 41 |
| Gambar 4.1 | Struktur <i>coding sequence</i> | 47 |
| Gambar 4.2 | Struktur Model 5 <i>state</i> | 47 |
| Gambar 4.3 | Struktur Model 8 <i>state</i> | 48 |
| Gambar 4.4 | Struktur model 14 <i>state</i> | 49 |
| Gambar 4.5 | Struktur model 23 <i>state</i> | 49 |

| | | |
|-------------|--|----|
| Gambar 4.6 | Struktur model 30 <i>state</i> | 50 |
| Gambar 4.7 | Struktur model 50 <i>state</i> | 50 |
| Gambar 4.8 | Struktur model 100 <i>state</i> | 51 |
| Gambar 4.9 | Pengembangan struktur dasar model | 51 |
| Gambar 4.10 | Struktur Model 1 | 52 |
| Gambar 4.11 | Struktur Model 2 | 52 |
| Gambar 4.12 | Struktur model 3 | 53 |
| Gambar 4.13 | Struktur model 4 | 53 |
| Gambar 4.14 | Skema blok proses simulasi | 54 |
| Gambar 4.15 | Data salah satu gen <i>P. falciparum</i> NC_004318 | 55 |
| Gambar 4.16 | <i>Label state</i> salah satu sekuen DNA gen <i>P. falciparum</i> NC_004318 | 57 |
| Gambar 4.17 | Nilai acak transisi <i>state</i> model 5 <i>state</i> | 59 |
| Gambar 4.18 | Ilustrasi TP, TN, FP dan FN | 62 |
| Gambar 4.19 | Diagram Alir Simulasi | 63 |
| Gambar 5.1 | Pemrograman <i>label state</i> sekuen | 67 |
| Gambar 5.2 | Hasil pengujian <i>label state</i> sekuen | 67 |
| Gambar 5.3 | Pemrograman emisi <i>state</i> | 69 |
| Gambar 5.4 | Hasil pengujian emisi <i>state</i> | 69 |
| Gambar 5.5 | Grafik Nilai CC rata-rata Struktur Model | 70 |
| Gambar 5.6 | Struktur Dasar Model 5 <i>state</i> | 72 |
| Gambar 5.7 | Struktur Dasar Model 7 <i>state</i> | 73 |
| Gambar 5.8 | Struktur Dasar Model 9 <i>state</i> | 74 |
| Gambar 5.9 | Grafik Nilai CC rata-rata Struktur Dasar Model | 75 |
| Gambar 5.10 | Grafik Nilai CC rata-rata Model 1 | 78 |
| Gambar 5.11 | Grafik Nilai CC rata-rata Model 2 | 79 |
| Gambar 5.12 | Grafik Nilai CC rata-rata Model 3 | 81 |
| Gambar 5.13 | Grafik Nilai CC rata-rata Model 4 | 82 |

DAFTAR SINGKATAN

| | | |
|----------|---|---|
| DSP | : | Digital Signal Processing |
| DSP | : | Digital Signal Processor |
| FPGAs | : | Field Programmable Gate Arrays |
| CAT scan | : | Computed Axial Tomography scan |
| MRI | : | Magnetic Resonance Imaging |
| PCR | : | Polymerase Chain Reaction |
| DNA | : | Deoxyribonucleic Acid |
| DFT | : | Discrete Fourier Transform |
| FIR | : | Finite Impulse Response |
| STFT | : | Short Time Fourier Transform |
| HMM | : | Hidden Markov Model |
| CC | : | Correlation Coefficient |
| VEIL | : | Viterbi Exon Intron Locator |
| pb | : | pasang basa (base pair/bp) |
| RNA | : | Rybonucleic Acid |
| A | : | Adenine |
| C | : | Cytosine |
| G | : | Guanine |
| T | : | Thymine |
| U | : | Urasil |
| rRNA | : | RNA ribosom |
| tRNA | : | RNA transfer |
| mRNA | : | RNA messenger |
| ORF | : | Open Reading Frame |
| TIGR | : | The Institute for Genomic Research |
| NMRC | : | National Microelectronics Research Centre |
| TP | : | True Positive |
| TN | : | True Negative |
| FP | : | False Positive |
| FN | : | False Negative |

DAFTAR ISTILAH

| | |
|-----------------|--|
| Triplet | : Pembentukan kelompok tiga molekul basa nitrogen pada Sekuen DNA yang merupakan kode genetik |
| Kodon | : Triplet basa pada RNA messenger (<i>mRNA</i>) |
| Ekson | : Rangkaian penyandi protein |
| Intron | : Rangkaian yang tidak menyandi protein |
| Kromosom | : Sarana pembawa informasi genetik pada makhluk hidup |
| Protein | : Kodon yang menghasilkan asam amino |
| Coding Sequence | : Rangkaian ekson tanpa intron yang diawali dengan kodon <i>start</i> (ATG) dan diakhiri dengan satu kodon <i>stop</i> (TAA atau TAG atau TGA) |
| Reading Frame | : Kerangka pembacaan pembentukan triplet pada untai basa DNA |
| True Positive | : Posisi <i>state</i> pada sekuen DNA asli merupakan <i>state</i> ekson dan posisi setelah tes tetap merupakan <i>state</i> ekson |
| True Negative | : Posisi <i>state</i> pada sekuen DNA asli merupakan <i>state</i> intron dan posisi setelah tes tetap merupakan <i>state</i> intron |
| False Positive | : Posisi <i>state</i> pada sekuen DNA asli merupakan <i>state</i> intron dan posisi setelah tes berubah menjadi <i>state</i> ekson |
| False Negative | : Posisi <i>state</i> pada sekuen DNA asli merupakan <i>state</i> ekson dan posisi setelah tes berubah menjadi <i>state</i> intron |

BAB 1

PENDAHULUAN

Sinyal merupakan suatu ukuran kuantitas fisik pada media tertentu antara lain seperti bentuk *audio (voice, speech, music)*, bentuk *visual (written documents, picture, video, gesture, facial expression)* dan sinyal dapat berisikan suatu informasi. Media adalah material fisik dimana sinyal dapat merambat seperti kertas untuk menulis, membuat gambar dan udara untuk suara, musik dan *sound pressure* serta berbagai alat untuk menampilkan gambar seperti *Cathode Ray Tube (CRT)*, *Liquid Crystal Display (LCD)*. Pemrosesan sinyal (*signal processing*) adalah suatu cara untuk memanipulasi sinyal pada media atau suatu representasi abstrak sebagai fungsi waktu atau ruang. Pemrosesan antara lain seperti proses transformasi, filterisasi, deteksi, estimasi, pengenalan, klasifikasi, pengkodean (kompresi), sintesis, reproduksi, perekaman, analisis dan pemodelan. Pemrosesan sinyal digital atau *Digital Signal Processing (DSP)* adalah pemrosesan sinyal sampling atau digital hasil konversi sinyal analog ke sinyal digital atau disebut dengan *Analog to Digital Converter (ADC)* dan menghasilkan sinyal digital yang di transformasikan kembali menjadi sinyal asli melalui *Digital to Analog Converter (DAC)*. Aplikasi DSP dalam teknologi dan ilmu pengetahuan antara lain dalam bidang-bidang seperti telekomunikasi, medik, bioteknik, militer, industri, pendekatan secara ilmiah, ruang angkasa dan komersil. Perkembangan aplikasi pada bidang bioteknik antara lain seperti pemrosesan sinyal genom (*Genomic Signal Processing*) dan pemrosesan sinyal biometrik (*Biometric Signal Processing*).

Pemrosesan sinyal genom seperti *Deoxyribonucleic Acid (DNA)* dan protein dengan teknologi digital basa-basa DNA dan asam amino yang membentuk protein dapat di asumsikan sebagai karakter *string* (huruf-huruf alfabet) sehingga dapat di manipulasi

menjadi bit-bit 1 dan 0. Pemrosesan sinyal genom seperti DNA dengan berbantuan komputer menyebabkan terjadinya *overload* data yang mengakibatkan dikembangkannya berbagai metode untuk memprediksi langsung daerah penyandi (ekson) seperti *GenScan*, *Fgenesh*, *Genie*, *Genmark*, *Morgan*, *Mzef*.

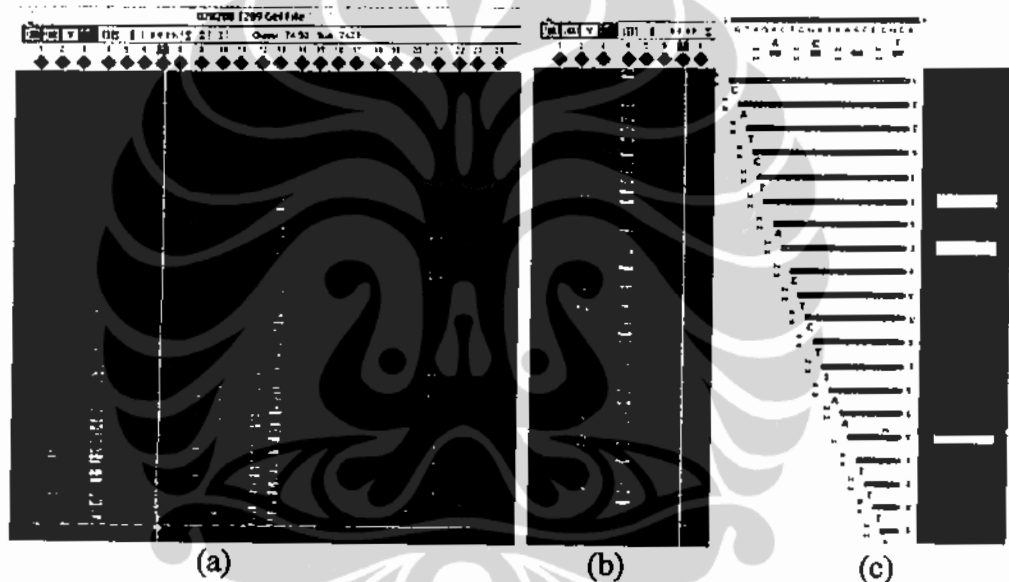
Penelitian ini menggunakan salah satu metode yaitu metode *Hidden Markov Model* (HMM) yang diaplikasikan pada rancangan struktur HMM untuk memprediksi ekson DNA *Plasmodium falciparum*.

1.1 Latar Belakang

Proses pengolahan sinyal digital diawali dengan proses pencuplikan sinyal masukan yang berupa sinyal kontinu menjadi sinyal diskrit yang dilakukan oleh suatu unit *Analog to Digital Converter* (ADC). Unit ADC ini terdiri dari sebuah bagian *Sample/Hold* dan sebuah bagian kuantisasi. Unit *sample/hold* merupakan bagian yang melakukan pencuplikan mulai dari 0 hingga satu periode waktu tertentu dan pencuplikan dilakukan pada setiap satu satuan waktu yang lazim disebut sebagai waktu cuplik (*sampling time*). Bagian kuantisasi akan mengubah hasil cuplikan menjadi beberapa level nilai kombinasi urutan bilangan 0 dan 1 (disebut juga dengan *biner*) sehingga proses pengolahan informasi berbantuan komputer menjadi lebih mudah, cepat dan akurat.

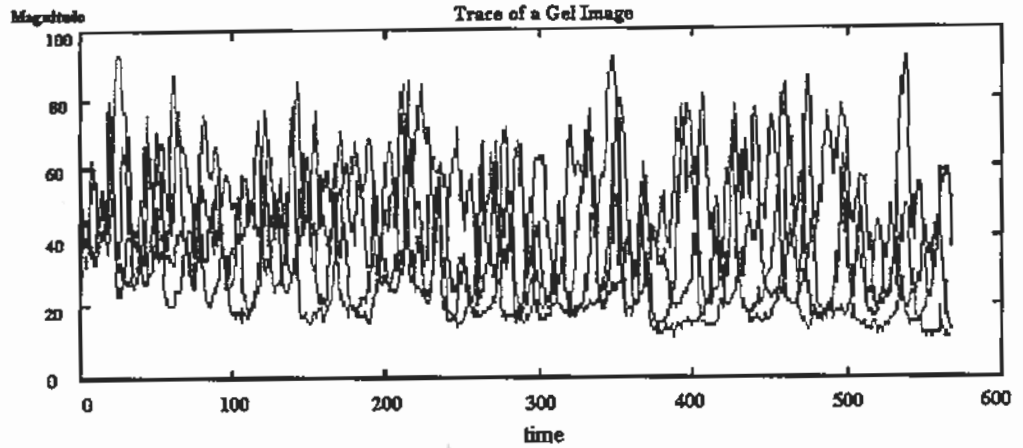
Aplikasi DSP dalam pemrosesan sinyal genom telah digunakan dalam proses analisis pengurutan basa *deoxyribonucleic acids* (DNA) dengan metode DNA *Sequencing*. Proses DNA *sequencing* memerlukan beberapa tahapan guna mendapatkan sinyal DNA seperti; isolasi DNA untuk mendapatkan DNA *template*, penggandaan DNA dengan teknik *Polymerase Chain Reaction* (PCR), pemisahan DNA secara *electrophoresis*, *gel image*, proses pembacaan *imaging* (*trace file*) dan proses deteksi basa-basa DNA (*base-caller*). *Gel image* yang dihasilkan dari proses *electrophoresis*

seperti pada Gambar 1.1a terdiri dari beberapa jalur atau *line* sample DNA dan memiliki warna yang berbeda untuk setiap basa. Proses pembacaan *imaging* dilakukan dengan *men-trace* setiap jalur pada *gel image* misalkan proses telah sampai pada jalur keempat seperti pada Gambar 1.1b dan bagaimana *base-caller* mendeteksi basa-basa yang ada pada jalur keempat itu maka ada bagian yang di *crop* menjadi Gbr 1.1c, terdapat sejumlah baris dari bagian jalur dengan warna yang berbeda yang menunjukkan basa-basa DNA terdeteksi seperti yang terdapat pada bagian kiri Gambar 1.1c tersebut.



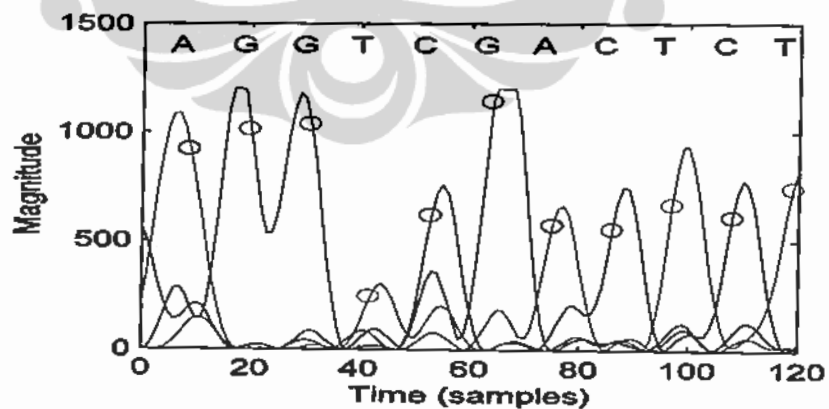
Gambar 1.1 *Gel image* sekuen DNA [1]

Basa-basa DNA yang telah terdeteksi dapat di representasikan sebagai model sinyal sekuen DNA dengan sumbu horizontal menunjukkan *time (samples)* dan sumbu vertikal menunjukkan *magnitude* setiap basa seperti pada Gambar 1.2. Pada proses deteksi setiap basa sekuen DNA harus mendapatkan *magnitude* maksimum dibandingkan dengan *magnitude* dari ketiga basa yang lain pada *time (sample)* yang sama sehingga basa itu terdeteksi dengan benar (*encoded*).

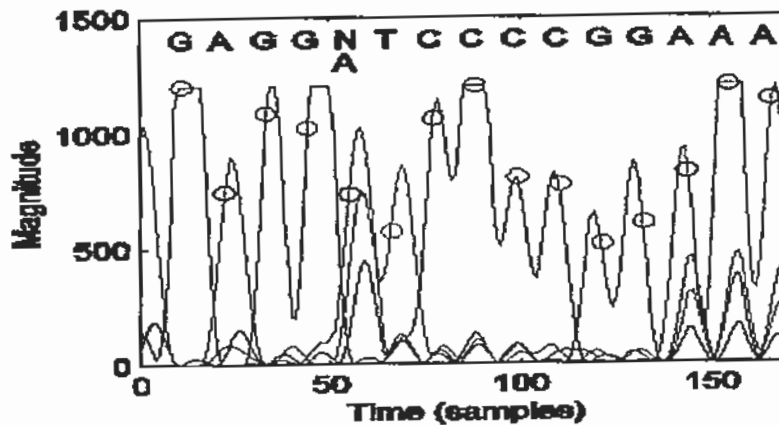


Gambar 1.2 Model sinyal sekuen DNA [2]

Banyak algoritma yang digunakan agar *base-caller* dapat mendeteksi setiap basa dengan benar dan salah satu algoritma tersebut adalah algoritma Viterbi dengan hasil model sinyal seperti pada Gambar 1.3 dan algoritma ini dapat pula mendeteksi basa-basa DNA yang tidak terdeteksi dengan benar (*non-encoded*) dimana pada *time (samples)* yang sama terdeteksi dua basa dengan besar *magnitude* yang tidak jauh berbeda sehingga kemudian diberi symbol N (*undetermined*) seperti model sinyal pada Gambar 1.4.



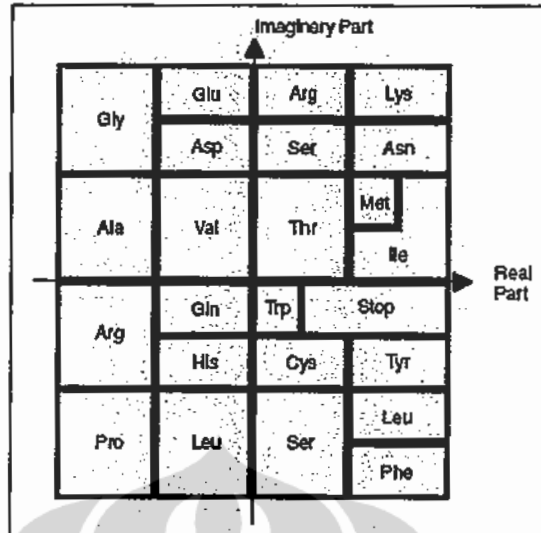
Gambar 1.3 Model sinyal sekuen DNA (*encoded*) [2]



Gambar 1.4 Model sinyal sekuen DNA (*non-encoded*) [2]

Beberapa aplikasi DSP pada bidang bioteknik hingga molekul sel membuat pemrosesan sinyal sangat diperlukan seperti pemrosesan sinyal untuk suatu genom hingga sekuen DNA gen atau protein yang disebut dengan *genomic signal processing*. Pengetahuan tentang dasar-dasar genom dan ekspresi genom diperlukan pada pemrosesan sinyal genom antara lain seperti panjang sekuen DNA, jumlah pasang basa (*base pairs*) atau pb, banyaknya gen dan informasi genetik yang terdapat pada DNA [3].

Menurut C.Q. Chang [3], Dimitris Anastasiou [4] dan P.D. Cristea [5], dengan menggunakan *Discrete Fourier Transform* (DFT) dan filter *Finite Impulse Response* (FIR) dalam *domain* frekuensi serta mengasumsikan koordinat-koordinat ke empat basa sekuen DNA maka informasi genetik pada DNA yang mengandung lokasi suatu kodon (asam amino) dapat di petakan ke dalam bidang kompleks seperti pada Gambar 1.5. Hasil uji coba ini pada akhirnya didapatkan informasi bahwa kodon *start* yaitu ATG atau dalam jenis asam amino disebut dengan Metionin (Met) berada pada kuadran satu dan ketiga kodon *stop* yaitu TAA, TAG dan TGA berada pada kuadran empat. Sehingga apabila rangkaian ekson yang terdiri dari rangkaian kodon tanpa intron dan diawali dengan kodon *start* dan diakhiri dengan satu kodon *stop* disebut dengan *coding sequence* (CDS).

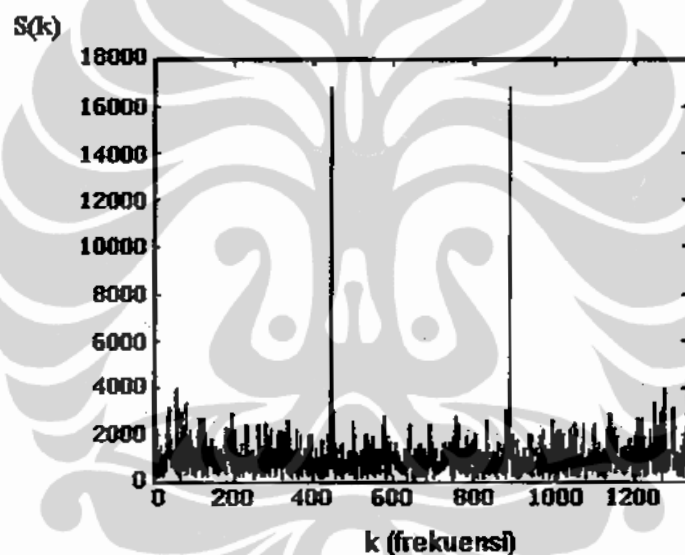


Gambar 1.5 Kode genetik pada bidang kompleks

Pembentukan kodon di dapat dari hasil kerangka pembacaan (*reading frame*) basa-basa DNA yang membentuk kelompok-kelompok tiga basa disebut juga dengan triplet dan terdapat tiga macam kerangka pembacaan tersebut yaitu *reading frame 1* yang di mulai dari urutan basa pertama, *reading frame 2* di mulai dari urutan basa kedua dan *reading frame 3* di mulai dari urutan basa ketiga sehingga jumlah jenis kodon yang dihasilkan dari ketiga kerangka pembacaan tersebut berbeda-beda dan maksimum adalah 64 yang diperoleh dari kombinasi empat basa yang dibentuk menjadi kelompok-kelompok tiga basa sehingga menghasilkan $4^3 = 64$ kodon. Setiap asam amino dibentuk paling sedikit oleh satu kodon yaitu asam amino Metionin (Met) dan paling banyak dibentuk oleh enam kodon antara lain asam amino Arginin (Arg), Serin (Ser) dan Leusin (Leu). Hasil uji coba ini telah di publikasikan pula pada *Proceeding of The 6th Quality in Research*, Fakultas Teknik Universitas Indonesia, Oktober 2003, dengan judul "*Proses Kode Genetik Pada Bidang Kompleks Menggunakan Filter Digital Finite Impulse Response*" [6].

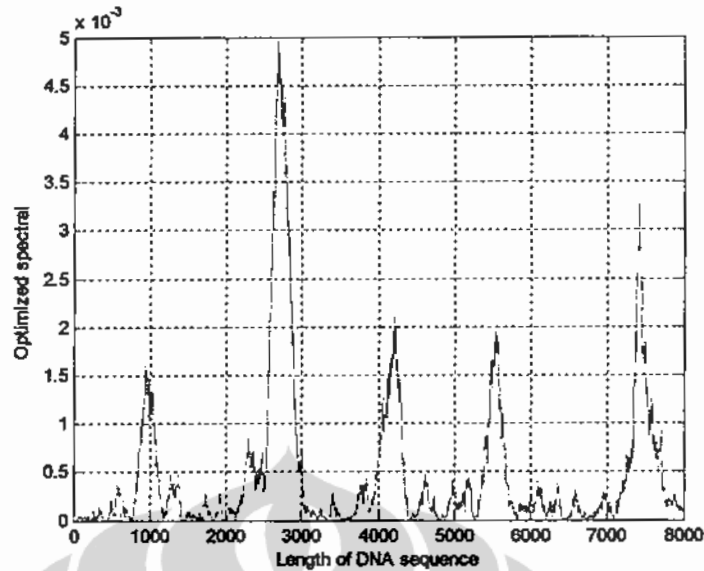
Struktur DNA yang terdiri dari ekson dan intron dan letaknya bergantian (*alternately located*) dengan proses *splicing* pada sekuen DNA tersebut maka diperoleh rangkaian

ekson dan rangkaian ini dapat diproses dengan DFT yang menghasilkan suatu spektrum ekson (*coding*). Sebagai contoh spektrum ekson DNA *S. cerevisiae* yang mempunyai panjang ekson (N) adalah 1320 dengan puncak nilai kuantitas dari keempat basa DNA tersebut berada pada frekuensi $k = N/3 = 440$ adalah periode dari ketiga basa pada kodon dan akan berulang pada $k = 880$ seperti yang terdapat pada Gambar 1.6 dengan sumbu horizontal adalah menyatakan frekuensi dan sumbu vertikal adalah menyatakan kuantitas total spektral basa ekson tersebut.



Gambar 1.6 Spektrum *coding* DNA

Sedangkan untuk spektrum daya yang dapat terjadi pada sekuen DNA diperoleh dari kuadrat *magnitude Short Time Fourier Transform* (STFT) rangkaian ekson dan intron. Sebagai contoh untuk menunjukkan spektrum daya dari bagian sekuen DNA *C. elegans* (*GenBank accession number AF 099922*) dengan panjang 8000 bp yang dimulai dari urutan 7021 pada sekuen asli dan memiliki lima ekson hasil dari *reading frame* 1 dan 2 seperti yang terdapat pada Gambar 1.7 [4] dengan sumbu horizontal adalah panjang sekuen dan sumbu vertikal adalah daya spektral optimal.



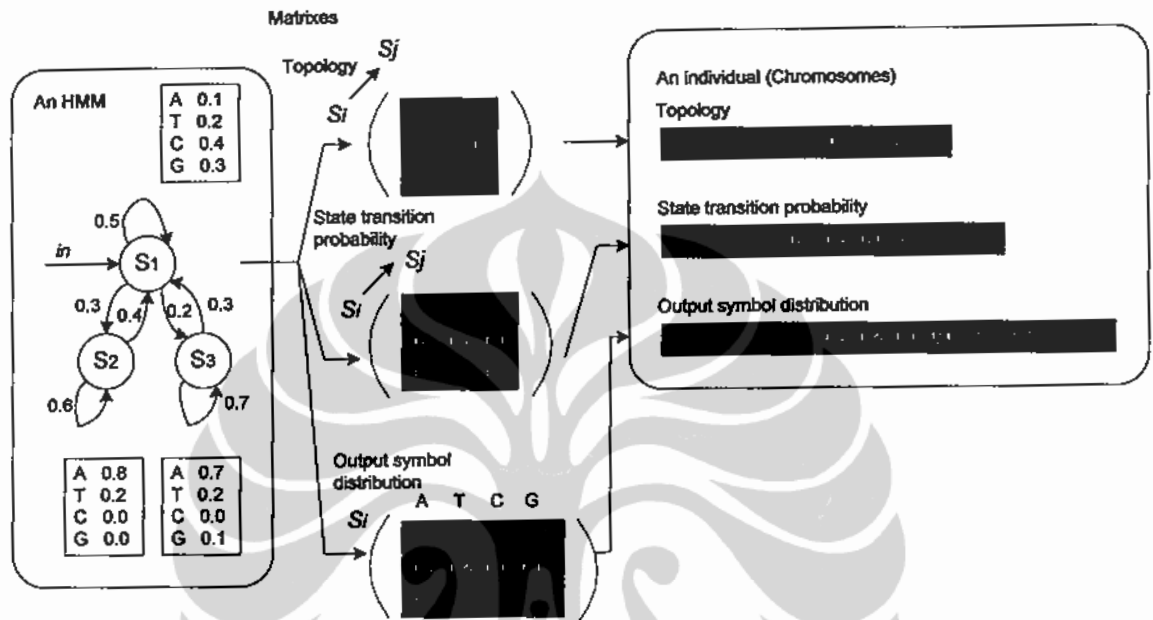
Gambar 1.7 Spektrum daya DNA

Gambar spektrum daya ini menjelaskan ada lima bagian puncak sesuai dengan kelima panjang ekson yang berbeda-beda yaitu ekson pertama 207 dari urutan 929-1135, ekson kedua 330 dari urutan 2528-2857, ekson ketiga 264 dari urutan 4114-4377, ekson keempat 180 dari urutan 5465-5644 dan ekson kelima 351 dari urutan 7255-7608. Setiap posisi puncak spektrum menunjukkan periode dari ketiga basa yang terdapat pada kodon yaitu $N/3$ yang ditambahkan pada urutan awal pada setiap ekson seperti contoh untuk puncak ekson pertama akan berada pada urutan $929 + 207/3 = 998$, dengan cara yang sama untuk keempat puncak lainnya.

Posisi dan jumlah ekson pada sekuen DNA sangat menentukan jenis asam amino yang terdapat pada protein setelah melalui proses translasi. Oleh karena itu banyak metode yang digunakan di dalam penelitian untuk mengidentifikasi atau memprediksi suatu genom dan salah satunya adalah dengan metode *Hidden Markov Model* (HMM).

Lawrence R. Rabiner [7], pada akhir tahun 1960 menjelang awal tahun 1970 banyak mempelajari metode statistik *Hidden Markov Model* (HMM) dan menjadi populer pada beberapa tahun terakhir. Ada dua alasan kuat mengapa banyak dipelajari metode ini, pertama model ini sangat kaya akan struktur matematik dan oleh sebab itu dapat menjadi

dasar teori untuk berbagai aplikasi. Alasan kedua adalah mengapa model sinyal itu penting karena model sinyal berpotensi membuat banyak belajar tentang sumber sinyal tanpa harus tersedianya sumber tersebut.

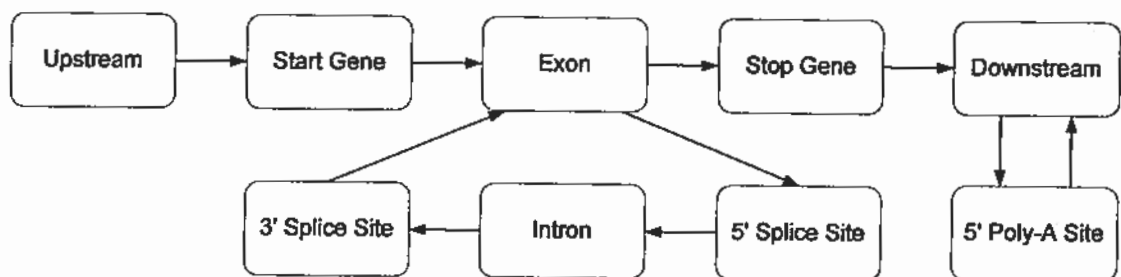


Gambar 1.8 Pemrosesan sekuen DNA dengan metode HMM [8]

Metode HMM yang digunakan untuk proses sekuen DNA dijelaskan dengan salah satu contoh seperti pada Gambar 1.8 dimana pada bagian struktur HMM terdiri dari tiga buah kromosom dengan masing-masing kromosom memiliki karakteristik parameter HMM yang berbeda. Parameter-parameter HMM antara lain seperti emisi yang terdapat pada setiap kromosom yang merupakan distribusi basa-basa DNA pada tiap kromosom serta nilai-nilai transisi dari satu kromosom ke kromosom lainnya yang tersusun dalam bentuk matrik topologi, matrik probabilitas dan matrik distribusi simbol keluaran dari model tersebut. Pada bagian akhir proses adalah setiap kromosom telah memiliki karakteristik probabilitas tertentu sehingga dengan mudah dapat terdeteksi bila terjadi mutasi (*mutation*), basa yang hilang (*deletion*), basa-basa yang berpindah (*cross over*) atau terjadi sisipan (*insertion*) diantara basa-basa yang ada pada setiap kromosom

tersebut. Model ini memberikan suatu hasil proses yang lebih rinci untuk setiap kromosom atau genom sehingga kemungkinan-kemungkinan kesalahan deteksi, prediksi atau identifikasi dapat diketahui [8].

Sekitar tahun 1990, mulailah digunakan metode HMM tersebut untuk proses identifikasi atau prediksi terhadap sekuen DNA untuk menentukan jenis gen atau genom. Yada dan Hirosawa [9], yang mengembangkan HMM untuk deteksi protein *coding regions* pada gen *cyanobacterium*, Strain PCC6808 *Synechocystis* sp yang terdapat pada GenBank, hasilnya 92,1 % dari *coding regions* berhasil di deteksi dengan nilai *Correlation Coefficient* (CC) sebesar 0,784 dan akurasi prediksi ini sebanding dengan menggunakan metode GeneMark. Kemudian Yada dan Hirosawa [10], melakukan deteksi *short protein coding regions* untuk sekuen yang sama dengan panjang sekuen dari 90 sampai 300 basa dengan metode GeneHacker lebih efektif untuk deteksi *very short CDSs* yang berbasis pada struktur HMM. Kemudian Henderson; Salzberg; Fasman [11], menjelaskan metode HMM untuk segmentasi sekuen *uncharacterized genomic* DNA ke dalam lokasi-lokasi ekson, intron dan intergenik dengan 92% berhasil di prediksi dengan nilai CC sebesar 0,73 dan menggunakan system *Viterbi Exon Intron Locator* (VEIL). Struktur HMM dengan menggunakan VEIL mempunyai beberapa model seperti untuk ekson, intron, intergenik, kodon *start*, kodon *stop*, donor dan akseptor terdapat pada Gambar 1.9.



Gambar 1.9 Struktur model HMM dengan sistem VEIL untuk penemuan suatu gen

Selanjutnya penelitian yang dilakukan oleh Daniel Nicorici [12] dengan menggunakan struktur model HMM seperti pada Gambar 1.9 untuk penemuan suatu gen tanpa menggunakan bagian 5' Poly-A site karena bagian tersebut menurut Michael Brudno[13] diperoleh setelah proses *splicing* dan merupakan awal proses translasi yang menghasilkan suatu protein.

Setelah mempelajari *domain* frekuensi dari penjelasan DFT yang digunakan untuk pemrosesan sekuen DNA dan *domain* waktu dari penjelasan tentang metode HMM, maka penelitian ini memilih metode HMM untuk memprediksi ekson DNA *Plasmodium falciparum*, agar pengembangan model dapat dirancang dengan penambahan jumlah *state* pada struktur model tersebut. Penambahan jumlah *state* dalam penelitian ini diasumsikan termasuk di dalam *domain* waktu yang dalam perancangan model dilakukan dengan memisahkan bagian kodon atau basa yang terdapat pada ekson atau intron.

Hasil pengujian elemen-elemen HMM seperti nilai acak transisi *state* sangat besar pengaruhnya terhadap kinerja model dan pada penelitian ini menggunakan salah satu parameter kinerja model yaitu nilai *Correlation Coefficient* (CC) yang dapat dihitung dari empat macam elemen yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN) [14]. Nilai acak transisi *state* sangat mempengaruhi hasil kinerja model ini dan sebagai hasil uji coba dari pengaruh nilai-nilai acak transisi *state* terhadap kinerja model telah di publikasikan dalam Seminar Internasional dan *Proceeding of BME DAYS 2006*, Institut Teknologi Bandung, November 2006, dengan judul "Correlation Coefficient Analysis of Exon region in DNA of Genes *Plasmodium falciparum* Using Hidden Markov Model" [15]. Sedangkan untuk penambahan jumlah *state* model digunakan struktur sistem VEIL yaitu dengan memisahkan kodon start dan kodon stop dari lokasi ekson pertama dan lokasi ekson terakhir. Begitu pula dengan penambahan *state* pada lokasi intron dilakukan dengan memisahkan *Donor site* (GT) dan *Acceptor site*

(AG) dari basa-basa lain yang terdapat pada lokasi intron tersebut. Perancangan struktur model benar-benar dilakukan mengikuti struktur ekson pada CDS sehingga untuk perancangan *state* intron mengikuti letak ekson dan intron yang bergantian (*alternately located*) seperti suatu rantai Markov sehingga dapat digunakan metode *forward-backward* yang terdapat pada kedua algoritma HMM yaitu Viterbi dan Baum-Welch. Hasil penelitian untuk struktur dasar model berbasis struktur ekson pada CDS telah diterima untuk di publikasikan pada *Jurnal Universa Medicina, Jurnal Kedokteran Trisakti, Vol. 26 No 4, (Oktober-Desember 2007)*, dengan judul "Exon Prediction on DNA-Genes *Plasmodium falciparum* based on Coding Sequence Structure by Using Hidden Markov Model" [16].

Pada penelitian ini juga dilakukan pengembangan struktur HMM yang berdasarkan struktur ekson pada CDS dengan menghasilkan model untuk sekuen yang memiliki lebih dari dua lokasi ekson pada CDS. Hasil pengembangan struktur HMM ini menghasilkan empat macam struktur model dengan penambahan jumlah *state* model dilakukan secara acak mulai dari 20, 30, 50 dan 100 *state* serta waktu proses simulasi untuk keempat jenis model. Pada Model 2 hasil dari pengembangan struktur model menghasilkan kinerja model lebih baik bila dibandingkan dengan ketiga model lain dan waktu proses *testing* dengan algoritma Baum-Welch rata-rata dua kali lebih lama dibandingkan dengan algoritma Viterbi. Hasil penelitian ini telah di publikasikan dalam Seminar Internasional dan di publikasikan pada *Proceeding of 4th International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2007, Orlando, Florida, USA, July 2007*, dengan judul "Exon Prediction on DNA-Gen *Plasmodium falciparum* by Using Hidden Markov Model" [17]. Perbedaan struktur pada pengembangan struktur model dilakukan dengan penambahan secara acak jumlah *state* hingga mencapai 100 *state* dan dilakukan bertahap pada bagian ekson dan intron

funksinya (*encoded*) dan tersimpan sebagai *database* yang dapat di akses melalui salah satu situs seperti <http://www.plasmodb.org>. [20].

Pengujian dilakukan dengan proses simulasi untuk struktur HMM yang telah dirancang sesuai dengan struktur dasar ekson pada CDS. Penambahan jumlah *state* pada model dilakukan secara acak sehingga dapat mempengaruhi penentuan nilai-nilai acak transisi *state*. Simulasi juga dilakukan untuk menghitung nilai emisi *state* sesuai dengan struktur model, melakukan proses *training* HMM dengan menggunakan algoritma Viterbi dan *testing* HMM dengan menggunakan algoritma Viterbi dan Baum-Welch sehingga memperoleh kinerja model yaitu dengan menggunakan nilai *Correlation Coefficient* (CC).

1.5 Kontribusi Penelitian

Kontribusi pertama yang dapat diberikan dari penelitian ini adalah rancangan model yang berbeda dari penelitian sebelumnya dimana penentuan *state* untuk lokasi intron diawali dengan *state* GT dan *state* AG yang lazim disebut sebagai *Donor site* dan *Acceptor site*. Dua basa GT dan AG telah ditetapkan sebagai bagian awal dan akhir lokasi intron sesuai dengan struktur umum gen eukariot. Oleh karena itu *state* GT dan *state* AG pada perancangan model penelitian ini tidak disertakan dengan beberapa basa seperti struktur *Donor site* dan *Acceptor site* yang selama ini digunakan untuk memprediksi ekson dengan metoda HMM.

Kontribusi kedua adalah pengujian pada saat menentukan nilai acak transisi *state* yang pada penelitian sebelumnya hal ini tidak dapat diketahui. Penentuan nilai ini sangat sulit dan kemungkinan berbeda pada setiap penelitian karena harus sesuai dengan struktur model dengan metode *forward-backward* dan rantai Markov yang berlaku pada HMM. Oleh karena itu penelitian ini mencoba untuk mengungkapkan nilai-nilai acak transisi

state setiap model dengan beberapa kali pengujian untuk mendapatkan kinerja model yang optimal.

Kontribusi ketiga adalah pengujian yang dilakukan dengan simulasi memungkinkan hasil diperoleh untuk memprediksi ekson khusus sekuen DNA *Plasmodium falciparum* dapat lebih cepat dan optimal. Biasanya pengujian tentang genom seperti DNA atau protein suatu organisme dilakukan dengan tes pada laboratorium dan membutuhkan waktu relatif lebih lama untuk mendapatkan hasilnya dibandingkan dengan proses simulasi. Sehingga pengujian dengan proses simulasi tetap perlu dilakukan karena dapat membantu dan mendukung proses tes laboratorium untuk memprediksi keberadaan ekson pada sekuen.

1.6 Sistematika Kerangka Penulisan

Sistematika kerangka penulisan penelitian ini dapat dijelaskan seperti Gambar 1.10 yang terdiri dari 6 Bab yaitu Bab 1 berisi pendahuluan antara lain tentang latar belakang, tujuan penelitian, metodologi penelitian, batasan masalah dan sistematika penulisan. Bab 2 berisi penjelasan tentang struktur dan organisasi genom *Plasmodium falciparum* yang digunakan sebagai masukan pada proses simulasi perancangan model *hidden* Markov yang diusulkan pada penelitian ini. Bab 3 berisi tentang teori dasar dan elemen model *hidden* Markov yang diaplikasikan untuk memprediksi ekson. Bab 4 berisi bentuk perancangan model yaitu struktur dasar model dan pengembangan struktur dasar model serta perancangan proses simulasi dengan menggunakan metoda HMM. Bab 5 berisi tentang pengujian metoda HMM untuk beberapa perancangan struktur model dan analisis hasil penelitian.

bergantian sesuai struktur sekuen aslinya. Penambahan *state* pada lokasi ekson bisa dilakukan dengan memisahkan kodon start dari ekson pertama pada CDS dan kodon stop pada ekson terakhir pada CDS. Begitu pula dengan lokasi intron, penambahan *state* dilakukan dengan memisahkan nukleotida GT dan AG dari untaian nukleotida intron pada CDS. Sedangkan pemilihan sekuen DNA *Plasmodium falciparum* disebabkan karena semua struktur ekson dan intron pada kromosom *Plasmodium* telah dapat diketahui fungsinya (*encoded*) dan tersimpan sebagai *database* yang dapat di akses melalui salah satu situs seperti <http://www.plasmodb.org>. [20].

Pengujian dilakukan dengan proses simulasi untuk struktur HMM yang telah dirancang sesuai dengan struktur dasar ekson pada CDS. Penambahan jumlah *state* pada model dilakukan secara acak sehingga dapat mempengaruhi penentuan nilai-nilai acak transisi *state*. Simulasi juga dilakukan untuk menghitung nilai emisi *state* sesuai dengan struktur model, melakukan proses *training* HMM dengan menggunakan algoritma Viterbi dan *testing* HMM dengan menggunakan algoritma Viterbi dan Baum-Welch sehingga memperoleh kinerja model yaitu dengan menggunakan nilai *Correlation Coefficient* (CC).

1.5 Kontribusi Penelitian

Kontribusi pertama yang dapat diberikan dari penelitian ini adalah rancangan model yang berbeda dari penelitian sebelumnya dimana penentuan *state* untuk lokasi intron diawali dengan *state* GT dan *state* AG yang lazim disebut sebagai *Donor site* dan *Acceptor site*. Dua basa GT dan AG telah ditetapkan sebagai bagian awal dan akhir lokasi intron sesuai dengan struktur umum gen eukariot. Oleh karena itu *state* GT dan *state* AG pada perancangan model penelitian ini tidak disertakan dengan beberapa basa seperti

struktur *Donor site* dan *Acceptor site* yang selama ini digunakan untuk memprediksi ekson dengan metoda HMM.

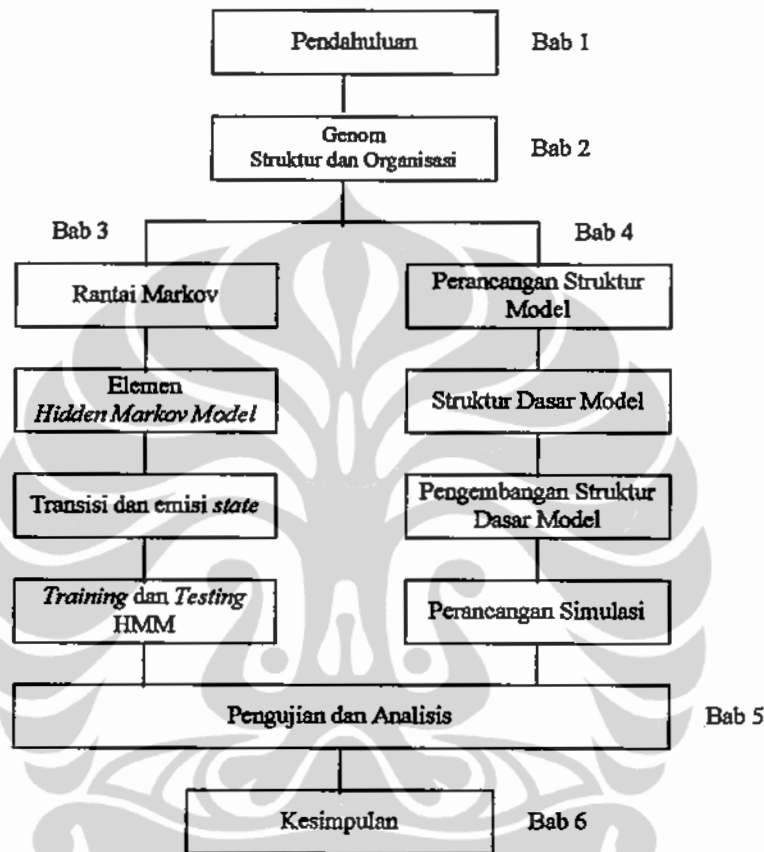
Kontribusi kedua adalah pengujian pada saat menentukan nilai acak transisi *state* yang pada penelitian sebelumnya hal ini tidak dapat diketahui. Penentuan nilai ini sangat sulit dan kemungkinan berbeda pada setiap penelitian karena harus sesuai dengan struktur model dengan metode *forward-backward* dan rantai Markov yang berlaku pada HMM. Oleh karena itu penelitian ini mencoba untuk mengungkapkan nilai-nilai acak transisi *state* setiap model dengan beberapa kali pengujian untuk mendapatkan kinerja model yang optimal.

Kontribusi ketiga adalah pengujian yang dilakukan dengan simulasi memungkinkan hasil diperoleh untuk memprediksi ekson khusus sekuen DNA *Plasmodium falciparum* dapat lebih cepat dan optimal. Biasanya pengujian tentang genom seperti DNA atau protein suatu organisme dilakukan dengan tes pada laboratorium dan membutuhkan waktu relatif lebih lama untuk mendapatkan hasilnya dibandingkan dengan proses simulasi. Sehingga pengujian dengan proses simulasi tetap perlu dilakukan karena dapat membantu dan mendukung proses tes laboratorium untuk memprediksi keberadaan ekson pada sekuen.

1.6 Sistematika Kerangka Penulisan

Sistematika kerangka penulisan penelitian ini dapat dijelaskan seperti Gambar 1.10 yang terdiri dari 6 Bab yaitu Bab 1 berisi pendahuluan antara lain tentang latar belakang, tujuan penelitian, metodologi penelitian, batasan masalah dan sistematika penulisan. Bab 2 berisi penjelasan tentang struktur dan organisasi genom *Plasmodium falciparum* yang digunakan sebagai masukan pada proses simulasi perancangan model *hidden* Markov yang diusulkan pada penelitian ini. Bab 3 berisi tentang teori dasar dan

elemen model *hidden* Markov yang diaplikasikan untuk memprediksi ekson. Bab 4 berisi bentuk perancangan model yaitu struktur dasar model dan pengembangan struktur dasar model serta perancangan proses simulasi dengan menggunakan metoda HMM.



Gambar 1.10 Sistematika Kerangka Penulisan

Bab 5 berisi tentang pengujian metoda HMM untuk beberapa perancangan struktur model dan analisis hasil penelitian. Proses pengujian dilakukan dengan mengintegrasikan semua bagian menjadi sebuah perangkat lunak dengan bahasa pemrograman yang digunakan untuk memprediksi ekson pada sekuen DNA *Plasmodium falciparum*. Hasil proses pengujian merupakan kesimpulan dan menjadi kinerja model penelitian ini yang dituliskan pada Bab 6 disamping penjelasan tentang kegiatan penelitian lanjut yang dapat digunakan untuk memprediksi ekson DNA *Plasmodium falciparum* dengan menggunakan metode HMM.

BAB 2

GENOM: STRUKTUR DAN ORGANISASI

Human Genome Project (HGP) merupakan suatu upaya selama 13 tahun, yang di koordinasikan oleh Departemen Energi dan Institusi Kesehatan Nasional Amerika Serikat dimulai pada Oktober 1990 sampai dengan 2003, untuk mencari sekitar 20.000 – 25.000 gen manusia dan membuatnya agar dapat diakses untuk kebutuhan studi biologi di masa datang. Tujuan proyek ini adalah untuk menentukan untaian lengkap dari 3 milyar pasang basa *Deoxyribonucleic acid (DNA)* yang terdapat pada manusia. Sebagai bagian dari HGP, studi studi paralel telah dilakukan pada model organisme tertentu seperti bakteri *E.coli* dan tikus untuk membantu mengembangkan teknologi dan menginterpretasikan fungsi dari gen makhluk hidup serta mengetahui etika, hukum dan masalah sosial yang mungkin terjadi dari proyek ini. *Department of Energy (DOE) Human Genome Program* dan *the National Institutes of Health (NIH) National Human Genome Research Institute (NHGRI)* secara bersama sama mensponsori HGP.

Genom adalah semua DNA dalam satu organisme termasuk gen yang merupakan pembawa informasi untuk menyandi protein yang diperlukan oleh semua organisme. Genom manusia di publikasikan sejak dimulainya HGP pada tahun 1989 sampai 2002, publikasi ini memfasilitasi komunikasi, membantu mencegah penggandaan terhadap usaha penelitian, dan menginformasikan kepada para peneliti yang tertarik pada penelitian tentang genom dan disponsori oleh program genom dari *Department of Energy Office of Biological and Environmental Research*.

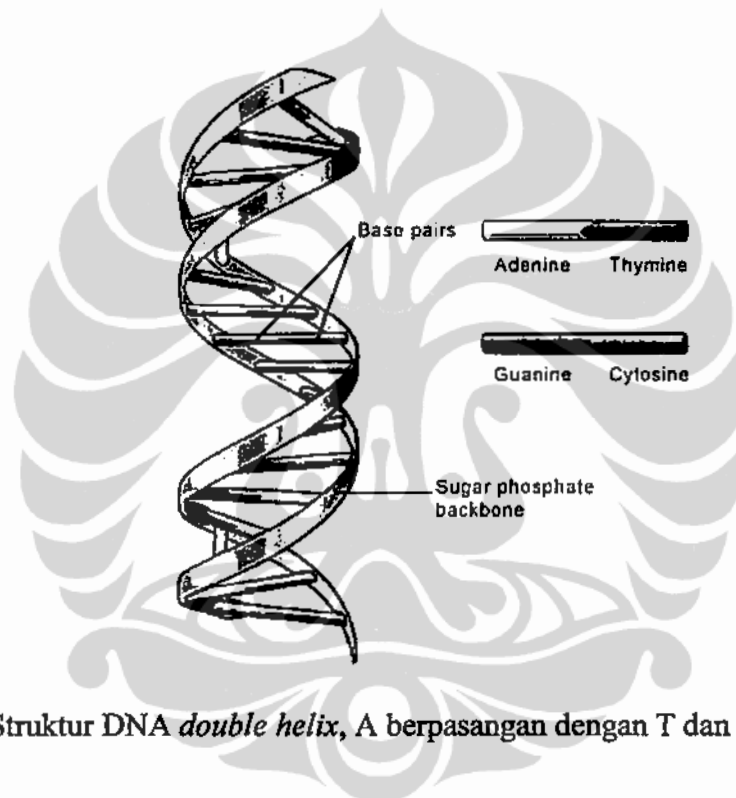
2.1 Deoxyribonucleic Acid (DNA)

Serangkaian kode genetik yang terdapat pada setiap individu memiliki sifat tertentu yang dapat diwariskan kepada keturunannya dan ilmu yang mempelajari tentang pewarisan sifat individu kepada keturunannya disebut genetika karena berkaitan dengan gen atau faktor keturunan. Setiap gen berfungsi mengontrol atau menentukan satu macam sifat, misalnya gen jenis dan warna rambut, gen warna mata, gen warna kulit, dan sebagainya. Gen terdapat berderet di dalam kromosom pada tempat-tempat yang disebut lokus. Setiap gen disusun oleh substansi genetik yang dikenal sebagai asam nukleat, yaitu DNA dan *Ribonucleid Acid* (RNA).

Dari sejumlah organisme yang terdapat di dalam sel, nukleus (inti sel) merupakan organ yang paling banyak mendapat perhatian dan berdasarkan membran (selaput pelindung) pada inti sel, makhluk hidup di kelompokkan menjadi 2 yaitu *prokariot* dengan nukleus tidak di selubungi oleh membran serta DNA tidak berada di dalam kromosom (contohnya bakteri) dan *eukariot* yang merupakan kelompok makhluk hidup dengan *nucleus* di selubungi oleh membran dengan DNA yang tersusun teratur di dalam kromosom (contohnya manusia, hewan, dan tumbuhan).

DNA terdiri dari dua jenis untaian yaitu untai tunggal (*single strand*) dan untai ganda (*double strand*). Setiap nukleotida disusun oleh basa nitrogen, gula de-oksiribosa dan asam fosfat. Antara nukleotida yang satu dengan nukleotida lainnya dihubungkan oleh suatu ikatan kimia antara gula dan fosfat. Ada 4 macam basa nitrogen yang ditemukan pada DNA, yaitu Adenin (A), Sitosin (C), Guanin (G) dan Timin (T). Keempat macam basa nitrogen ini menyusun DNA secara berpasangan, Guanin hanya dapat berpasangan dengan Sitosin atau sebaliknya, sedangkan Adenin berpasangan dengan Timin atau sebaliknya dengan struktur yang dikenal sebagai DNA *double helix* seperti pada Gambar 2.1 [21].

Informasi yang ada dalam DNA ditentukan oleh urutan dari huruf (basa) sepanjang untai tersebut. Misalnya urutan ACGCT menunjukkan informasi yang berbeda dengan urutan AGTCC. Ciri khas DNA makhluk hidup adalah informasi yang mengandung kode DNA. Perbedaan urutan DNA membedakan makhluk hidup yang satu dengan yang lain atau karakter mereka karena perbedaan struktur asam amino yang membentuk protein.



Gambar 2.1 Struktur DNA *double helix*, A berpasangan dengan T dan G dengan C [21]

2.2 Ribonucleic Acid (RNA)

Pada organisme *eukariot*, RNA bersama DNA berfungsi dalam pembentukan protein dan nukleotida pada RNA di susun oleh komponen yang sama dengan penyusun nukleotida DNA, yaitu gula, fosfat dan basa nitrogen. Secara ringkas, komponen penyusun RNA terdiri dari satu molekul gula berupa ribosa, gugus fosfat, dan 4 basa nitrogen yaitu Adenin (A) dan Guanin (G), Sitosin (C) dan Urasil (U).

RNA terdiri dari 3 jenis yaitu *RNA ribosom* (rRNA) sebagai RNA penyusun ribosom bersama protein, *RNA transfer* (tRNA) sebagai RNA pemindah dan berfungsi

untuk mengenali dan mengikat satu macam asam amino, mengenali dan mengikat urutan khusus nukleotida pada molekul mRNA serta *RNA messenger* (mRNA) adalah RNA yang bertugas menyampaikan informasi dari DNA ke ribosom [21,22].

2.3 Ekspresi Gen

Segala kegiatan sel (seperti metabolisme, reproduksi, dan perkembangan) dipengaruhi oleh enzim dan jika tubuh kekurangan salah satu enzim maka sel akan menanggapinya dan memerintahkan DNA untuk membuat enzim yang diperlukan sedangkan enzim merupakan molekul protein.

Satu jenis protein minimal tersusun atas 20 macam asam amino dalam satu urutan yang spesifik dan masing-masing asam amino harus mempunyai kode tertentu untuk memudahkan instruksi, sehingga diperlukan minimal 20 kode. Untuk keperluan pengkodean ini, sekitar tahun 1940 dua orang ahli genom, Beadle dan Tatum, membuat hipotesis "triplet basa". Satu triplet basa adalah 3 molekul basa nitrogen yang terletak berurutan pada DNA dan merupakan kode bagi satu jenis asam amino tertentu. Triplet basa ini disebut juga dengan kodon yang merupakan kode genetik.

DNA berada dan terikat pada kromosom di dalam nukleus, sedangkan sintesis protein terjadi pada ribosom di dalam sitoplasma. Ada dua kemungkinan untuk penyampaian kode genetik dari DNA ke ribosom, yaitu:

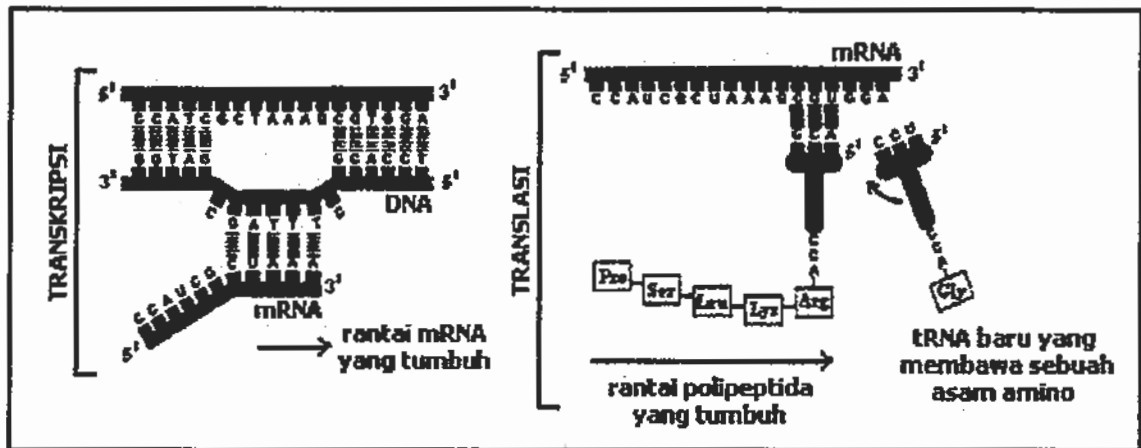
- a. asam deoksiribonukleat (DNA) keluar dari nukleus dan menuju ke sitoplasma;
- b. terdapat satu molekul yang berperan sebagai kurir, yaitu membawa perintah dari DNA ke sitoplasma.

Kemungkinan pertama tidak mungkin terlaksana karena tidak pernah ditemukan DNA di dalam sitoplasma. Sehingga kemungkinan kedua yang akan terjadi dimana peristiwa yang ditemukan adalah DNA membentuk RNA dan RNA inilah yang kemudian

menyampaikan instruksi kepada sitoplasma. RNA yang berfungsi menyampaikan instruksi dari DNA disebut mRNA.

Pada pembentukan mRNA, bagian yang dikehendaki dari DNA akan membuka (pasangan basa terpisah) dan salah satu bagian berfungsi sebagai cetakan untuk pembentukan RNA. Setelah mRNA terbentuk kemudian mRNA mengelupas dari cetakannya dan bergerak keluar meninggalkan inti melalui pori-pori pada membran inti menuju ke dalam sitoplasma. Dalam sitoplasma mRNA akan melekat pada permukaan ribosom. Pembentukan protein dibantu oleh tRNA yang merupakan benang tunggal yang berlipat membentuk lilitan-lilitan. Pada tRNA terdapat 3 basa nitrogen dengan urutan yang bervariasi dan pada salah satu ujung benangnya terikat satu jenis asam amino. Setiap sel mempunyai gudang penyimpanan RNA (misalnya *mukleolus*) yang berisi bermacam RNA, termasuk tRNA. Pada pembentukan protein, antikodon tRNA akan berkaitan pada kodon mRNA. Selanjutnya, asam amino yang datang bersama tRNA akan berderet sesuai dengan kodon mRNA, kemudian asam-asam amino ini akan berikatan membentuk protein [21,22]

Secara garis besar langkah pencetakan protein itu terjadi melalui 2 tahap, yaitu transkripsi dan translasi. Transkripsi adalah pencetakan mRNA oleh DNA untuk melakukan sintesis protein. Sedangkan translasi adalah penerjemahan kode oleh tRNA, berupa urutan asam amino yang dikehendaki. mRNA menyampaikan informasi kepada ribosom untuk menyintesis protein yang diperlukan dengan menggunakan kode-kode perintah atau informasi yang tercermin pada urutan dan pengulangan basa nitrogen yang teratur dalam RNA. Proses sintesis protein dijelaskan pada Gambar 2.2.



Gambar 2.2 Urutan sintesis protein

DNA, RNA dan polipeptida adalah molekul-molekul yang berbentuk panjang dan tidak bercabang. Sampai saat ini telah dikenal dua puluh macam asam amino sebagai bahan dasar untuk protein. Karena terdapat empat basa nitrogen yang berbeda dalam DNA dan RNA, dan terdapat dua puluh macam asam amino yang berbeda dalam protein, maka basa tersebut tidak dapat memberikan satu kode untuk setiap asam amino. Oleh karena itu, kode genetik dapat menggunakan urutan singkat dari basa untuk memberikan kode bagi masing-masing asam amino. Apabila suatu urutan tiga basa memberikan kode untuk satu asam amino, maka akan terjadi $4^3 = 64$ kemungkinan kombinasi dari basa sehingga menghasilkan 64 macam kelompok tiga basa atau triplet atau kodon. Tabel 2.1 adalah tabel kode genetik yang dihasilkan dari 64 kode triplet atau kodon mRNA [4,21].

Kemungkinan adanya terjadi kesalahan didalam menerjemahkan kode-kode genetik yang diterima dari DNA dapat mengakibatkan protein yang disusun juga menjadi salah sehingga dengan sendirinya enzim yang dihasilkan juga salah. Oleh karena itu kekeliruan tRNA menafsirkan kode-kode genetik yang diterima dari DNA merupakan salah satu mekanisme mutasi gen dan mempunyai akibat yang luas sehingga dapat pula menyebabkan adanya keanekaragaman organisme hidup di dunia ini.

Tabel 2.1 Kode genetik

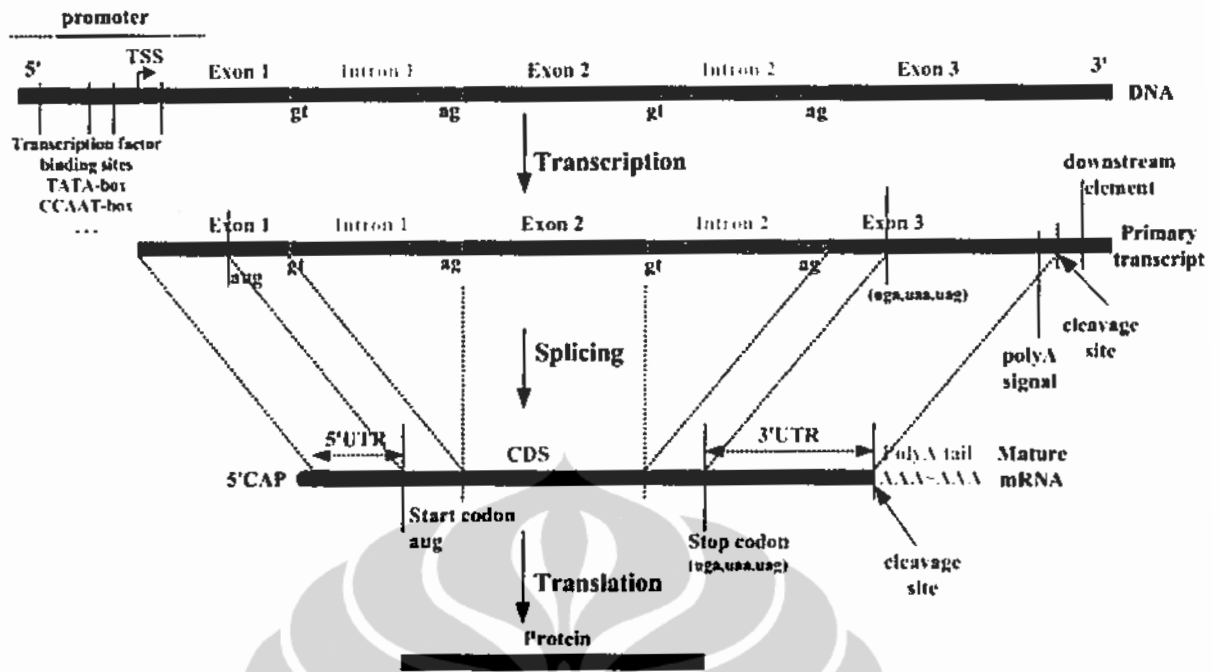
| Basa Pertama | Basa Kedua | | | | Basa Ketiga |
|--------------|-------------------|---------------|-----------------|-------------|-------------|
| Ujung 5' | A | G | T | C | Ujung 3' |
| A | AAA Lisin | AGA Arginin | ATA Isoleusin | ACA Treonin | A |
| | AAG Lisin | AGG Arginin | ATG Metionin | ACG Treonin | G |
| | AAT Asparagin | AGT Serin | ATT Isoleusin | ACT Treonin | T |
| | AAC Asparagin | AGC Serin | ATC Isoleusin | ACC Treonin | C |
| G | GAA Asam glutamat | GGA Glisin | GTA Valin | GCA Alanin | A |
| | GAG Asam glutamat | GGG Glisin | GTG Valin | GCG Alanin | G |
| | GAT Asam aspartat | GGT Glisin | GTT Valin | GCT Alanin | T |
| | GAC Asam aspartat | GGC Glisin | GTC Valin | GCC Alanin | C |
| T | TAA "Stop" | TGA "Stop" | TTA Leusin | TCA Serin | A |
| | TAG "Stop" | TGG Triptofan | TTG Leusin | TCG Serin | G |
| | TAT Tirosin | TGT Sistein | TTT Fenilalanin | TCT Serin | T |
| | TAC Tirosin | TGC Sistein | TTC Fenilalanin | TCC Serin | C |
| C | CAA Glutamin | CGA Arginin | CTA Leusin | CCA Prolin | A |
| | CAG Glutamin | CGG Arginin | CTG Leusin | CCG Prolin | G |
| | CAT Histidin | CGT Arginin | CTT Leusin | CCT Prolin | T |
| | CAC Histidin | CGC Arginin | CTC Leusin | CCC Prolin | C |

Pada Tabel 2.1 terdapat 20 macam asam amino dan pembentukan setiap asam amino terjadi dari kodon yang berbeda, misalnya kodon AAA dan AAG membentuk asam amino Lisin sedangkan kodon AAT dan AAC membentuk asam amino Asparagin. Sedikitnya ada satu kodon yang dapat membentuk asam amino seperti asam amino Metionin (ATG) dan paling banyak ada enam kodon yang membentuk asam amino seperti Leusin, Arginin dan Serin, sedangkan kodon-kodon Stop yaitu TAA, TAG dan TGA bukan merupakan asam amino.

2.4 Ekson dan Intron

Suatu rantai DNA tersusun atas rangkaian penyandi protein (ekson) dan rangkaian bukan penyandi protein (intron). Ekson didefinisikan sebagai rangkaian kodon yang dapat ditranslasi menjadi protein (asam amino) atau dengan kata lain ekson adalah daerah pengkodean protein. Sedangkan intron adalah daerah pengkodean yang tidak dapat ditranslasi menjadi protein atau disebut sebagai daerah yang tidak menyandi. Intron harus dihilangkan saat akan ditranslasi menjadi protein dan kemunculan intron diawali oleh basa-basa GT dan diakhiri oleh basa-basa AG. Dalam suatu rangkaian DNA, posisi intron berada diantara dua ekson sehingga jumlah intron adalah satu lebih sedikit dari jumlah ekson.

Pada sel-sel bakteri (*prokariot*) kebanyakan protein disandikan oleh sebuah rangkaian DNA tunggal yang disalin tanpa perubahan untuk memproduksi sebuah molekul mRNA. Sedangkan struktur gen *eukariot* memiliki rangkaian-rangkaian penyandi (ekson) yang diselingi oleh rangkaian-rangkaian bukan penyandi (intron) seperti dijelaskan pada Gambar 2.3 [14]. Untuk memproduksi sebuah protein, keseluruhan panjang gen yang meliputi ekson dan intron, mula-mula ditranskripsikan ke dalam sebuah molekul RNA yang sangat besar. Sebelum molekul RNA meninggalkan *nukleus*, dengan sebuah kompleks enzim pengolah RNA dapat menghilangkan semua rangkaian intron sehingga menjadikan molekul RNA jauh lebih pendek. Setelah selesai proses menghilangkan intron (*RNA splicing*) maka molekul RNA bergerak ke sitoplasma sebagai molekul mRNA yang mensintesis protein tertentu.



Gambar 2.3 Struktur gen eukariot

Salah satu contoh sekuen DNA *Plasmodium falciparum* yang diperoleh dari database GenBank mempunyai bentuk format GenBank yang diuraikan seperti pada Gambar 2.4 dibawah ini.

| | |
|------------|---|
| LOCUS | NC_004330 861 bp DNA linear INV 30-JAN-2006 |
| DEFINITION | Plasmodium falciparum 3D7 chromosome 9, PF11250w gene, complete CDS. |
| ACCESSION | NC_004330 REGION: 1..1541723 |
| VERSION | NC_004330.1 GI:23613523 |
| KEYWORDS | . |
| SOURCE | Plasmodium falciparum 3D7 |
| ORGANISM | Plasmodium falciparum 3D7 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. |
| REFERENCE | 1 (bases 1 to 861) |
| AUTHORS | Sanders, M., Hauser, H., Baker, S., Unwin, L., Mungall, K., Berriman, M., Pain, A., Hall, N., Bowman, S., Churcher, C., Quail, M. and Barrell, B. |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 861) |
| AUTHORS | Sanders, M., Hauser, H., Baker, S., Unwin, L., Mungall, K., Berriman, M., Pain, A., Hall, N., Bowman, S., Churcher, C., Quail, M. and Barrell, B. |
| TITLE | Direct Submission |
| JOURNAL | Submitted(20-SEP-2002)P.falciparum Genome Sequencing Consortium, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK |

Gambar 2.4 Data salah satu gen *P.falciparum* NC_004330

```

COMMENT      PROVISIONAL REFSEQ: This record has not yet been subject
              to final NCBI review. The reference sequence was derived
              from AL844508.
FEATURES     Location/Qualifiers
     source   1..861
              /organism="Plasmodium falciparum 3D7"
              /mol_type="genomic DNA"
              /sub_species="3D7"
              /db_xref="taxon:36329"
              /chromosome="9"
     gene     1..861
              /gene="PFI1250w"
              /db_xref="GeneID:813530"
     CDS     join(1..78,303..377,457..523,620..700,809..861)
              /gene="PFI1250w"
              /note="Similar to Ictalurus punctatus thioredoxin
              TR:Q9DGI3 (EMBL:AF293651) (107 aa) fasta scores: E():
              2.7e-09, 38.202% id in 89 aa. Weak similarities to
              Plasmodium falciparum thioredoxin 1 txrl TR:Q9NIR2
              (EMBL:AF202664) (104 aa) fasta scores: E(): 3.7e-08,
              34.884% id in 86 aa. HMMPfam hit to PF00085,
              Thioredoxin; ProfileScan hit to PS50223,
              Thioredoxin-domain (does not find all)."
```

```

              /codon_start=1
              /product="thioredoxin, putative"
              /protein_id="NP_704793.1"
              /db_xref="GI:23613772"
              /db_xref="GeneID:813530"
              /translation="MFFLQNLKKITHPKYLCTTIRNVYTVNGNKLVVA
              QFGASWCAPCKKMKPVIEKLGEDNDNIESLYIDIDEFPELGENEDINEL
              PTILLRKNKGKYLDKIIGMNESDLIKAVEKHQSD"
ORIGIN
     1 atgttttttc ttcaaaattt aaaaaaaatt acccatccaa aatatttatg cacaacaatt
     61 agaagaaatg tgtacacagt aaggataaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
    121 aatacataaa tatatatata atatatatta taatatggct ggaacaaact caattataac
    181 aattcttttt tttttttttt tttttttttt ttgtaatatg tatttgtttt tgcTcaaaaa
    241 caataatttg tttattataa aaaatatagg aattaaataa aattgatgat tatcttagta
    301 aggttaatgg aaataaattg gttgttgctc aatttggtgc ttcatggtgt gcccttgta
    361 aaaaaatgaa accagtggta aataaataaa taaatatata tatatatata tatatatata
    421 tatatacata tttatatatc tgttatatat ttatagattg aaaagctggg agaagataac
    481 gataacattg aatctttata tatagacatt gacgaattc caggtatatt actagaaaaa
    541 aatgaaact atagtaatac ctatctttgt ttatgatata tttattcatt taaattattt
    601 attttgttta atttaacaga actaggtgaa aatgaagata ttaatgaatt acctacaatt
    661 ttgttgagaa aaaatgggaa atatttgac aagattatag gtgaccttat atctcacact
    721 aaaaatataa atataataaa catagaagtg tgtctttttc gttctatata tatatatata
    781 tatatataat tttttctttt ttttttaggg atgaatgaga gtgatttaat aaaagccgtt
    841 gaaaaacatc agagtgattg a
//
Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

```

Gambar 2.4 Data salah satu gen *P.falciparum* NC_004330 (Lanjutan)

Pada Gambar 2.4, jenis gen tersebut memiliki CDS yang terdiri dari lima lokasi ekson (warna biru) termasuk kodon *start* pada awal sekuen dan kodon *stop* pada akhir

sekuen tetapi untuk membedakan dari basa-basa ekson yang lain maka warna kodon tersebut di bedakan. Sedangkan untuk lokasi intron dapat diketahui dari sekuen asli gen tersebut ada sebanyak 4 lokasi yang berwarna merah dan biasanya lokasi-lokasi intron diawali dengan basa-basa GT dan diakhiri dengan basa-basa AG.

2.5 Reading Frame

Rangkaian nukleotida dalam molekul mRNA yang berfungsi sebagai bahan perantara harus dibaca secara berurutan dalam kelompok-kelompok tiga yang disebut triplet nukleotida atau kodon dan menyandi sebuah asam amino. Pada prinsipnya, setiap rangkaian nukleotida dapat ditranslasikan dalam suatu *reading frame* (kerangka pembacaan). Ada tiga macam *reading frame* yaitu *reading frame 1*, *reading frame 2*, dan *reading frame 3*. Ketiga *reading frame* tersebut menghasilkan pengkodean protein yang berbeda. Dalam hampir setiap kasus, hanya salah satu dari ketiga *reading frame* itu yang akan memproduksi sebuah protein yang fungsional. Karena tidak adanya "tanda baca" kecuali pada awal dan akhir pesan RNA, kerangka pembacaan ditentukan sejak proses translasi dimulai dan selanjutnya tetap demikian. Salah satu contoh yang menjelaskan bagaimana *reading frame* diterapkan pada suatu rangkaian nukleotida yang terdapat dalam molekul mRNA suatu organisme misalkan tersusun seperti contoh dibawah ini :

'CATTGCAGTACTGGC'. Kemudian rangkaian nukleotida tersebut ditranslasikan ke dalam tiga macam *reading frame* sehingga diperoleh tiga buah urutan pengkodean protein yang berbeda seperti berikut ini [4]:

reading frame 1 : CAT TGC AGT ACT GGC

His Cys Ser Thr Gly

reading frame 2 : C ATT GCA GTA CTG GC

Ile Ala Val Leu

reading frame 3 : CA TTG CAG TAC TGC C

Leu Gln Tyr Cys

Dari contoh di atas untuk basa-basa yang tidak tersusun secara triplet maka tidak dapat diterjemahkan menjadi protein. Kumpulan ekson hasil *splicing* yang membentuk rangkaian nukleotida baru yang diawali dengan kodon start dan diakhiri dengan satu kodon *stop* dinamakan *open reading frame* (ORF).

2.6 Organisasi Genom *Plasmodium falciparum*

Penyakit malaria pada manusia disebabkan karena infeksi parasit di dalam sel gen *Plasmodium* yang ditularkan oleh vektor nyamuk *Anopheles*. *Plasmodium* yang dapat menyebabkan penyakit malaria pada manusia yaitu *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale* dan *Plasmodium malariae*. 80% penyakit malaria pada manusia disebabkan oleh *P. falciparum* dan 90% dari kematian disebabkan oleh *P. falciparum* dengan bentuk seperti pada Gambar 2.5 [23], oleh karena itu spesies ini yang paling mematikan dibandingkan dengan spesies lainnya. Genom inti *P. falciparum* memiliki panjang kurang lebih 25 Mb dan terdiri dari 14 kromosom dengan panjang bervariasi ~0.6-3.4 Mb [24,25].



Gambar 2.5 *Plasmodium falciparum*

Kromosom ini dapat diurai pada media gel elektroforesis yang bergetar dan pada media yang berbeda seringkali kromosom terpisah dalam berbagai ukuran, hal ini terjadi karena rekombinasi yang melibatkan bagian subtelomerik kromosom. Genom ini sangat kaya akan nukleotida A+T (80%) dan kepingan besar dari DNA *P. falciparum* menjadi tidak stabil di dalam organisme *E. coli*. Parasit malaria dan parasit lain yang terkait seperti *Toxoplasma* juga memiliki genom mitokondria kira-kira 6kb, dan sebuah DNA sirkular 35kb yang telah terlokalisasi menjadi sebuah organell baru yang disebut *apicoplast* [26].

The Institute for Genomic Research (TIGR) dan *National Microelectronics Research Centre (NMRC)* berhasil mensekuen kromosom 2, 10, 11, dan 14, terdiri dari kurang lebih 9 Mb. Tata cara yang dipakai pada tiga pusat perangkaian agak beragam namun melibatkan lima langkah dasar yakni pembentukan *library*, perangkaian acak, pemasangan, penutupan dan pemberian keterangan. Kromosom 2 telah disempurnakan dan diterbitkan pada 1998 [27], sementara kromosom 10, 11, dan 14 sedang dalam penyelesaian dan hampir semua kromosom berakhir dengan ciri sebagai suatu seri bagian berulang.

Telomer merupakan struktur serupa dari eukariot lain, mengandung rangkaian heptanukleotide berulang yang jumlahnya beragam di antara kromosom. Bagian kompleks berulang (*complex repeats*) dan bagian tidak-berulang (*non-repetitive repeats*) mengikuti bagian telomerik berulang. Bagian rep-20 (Rep20) terdiri dari perulangan 21 bp, kemudian di ikuti oleh bagian tidak-berulang dan dipisahkan oleh sisi batasan ApaI serta Rep-20 akan berulang-ulang sampai mengandung 5% genom parasit.

Pemisahan kromosom *P. falciparum* oleh media gel elektroforesis bergetar dapat menjadi beberapa kelompok seperti berikut:

1. Kromosom-kromosom dengan berat molekul ringan yaitu kromosom 1, kromosom 2, kromosom 3 dan kromosom 4.

2. Kromosom-kromosom dengan berat molekul medium yaitu kromosom 10, kromosom 11 dan kromosom 12.
3. Kromosom-kromosom dengan berat molekul besar yaitu kromosom 13 dan kromosom 14.
4. Kromosom dengan bentuk gumpalan yaitu kromosom 5, kromosom 6, kromosom 7, kromosom 8 dan kromosom 9.

Pada sekuen DNA *Plasmodium falciparum* dengan format Genbank dapat diketahui bahwa setiap sekuen DNA memiliki satu CDS yang merupakan rangkaian ekson tanpa intron. Selain CDS, pada format tersebut dapat pula diketahui untaian basa asli sekuen yang terdiri dari rangkaian ekson dan intron yang letaknya bergantian (*alternately located*) yang menyerupai suatu rantai serta setiap intron diawali dengan basa-basa GT dan diakhiri basa-basa AG.

Penelitian ini menggunakan sekuen DNA *Plasmodium falciparum* dengan format GenBank dan pada untaian basa asli sekuen DNA tersebut harus merupakan untaian basa yang telah terdeteksi dengan benar (*encoded*) atau tidak terdapat basa-basa yang *non-encoded* (N). Hal ini penting agar proses kerangka pembacaan (*reading frame*) yang membentuk kelompok-kelompok tiga basa atau triplet menghasilkan rangkaian kodon sebagai ekson dapat di prediksi.

BAB 3

MODEL HIDDEN MARKOV

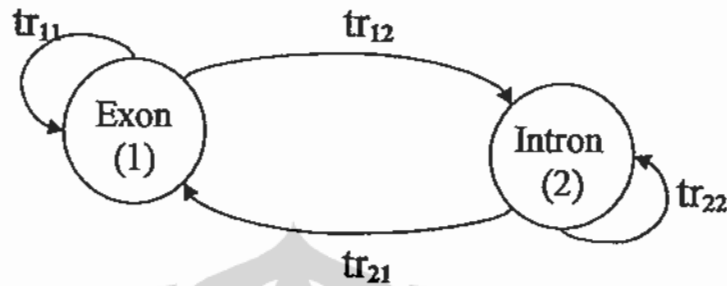
Hidden Markov Model (HMM) merupakan suatu model statistik yang digunakan untuk membuat karakteristik suatu *frame* sinyal DNA. Asumsi dasar dari HMM adalah sinyal DNA dapat di karakterisasi sebagai suatu representasi proses *random* parametrik. Parameter-parameter dari proses stokastik ini diharapkan dapat di prediksi dengan tepat. HMM terdiri dari serangkaian proses stokastik rangkap, dimana proses stokastik utamanya tidak dapat di observasi secara langsung (*Hidden*) tetapi dapat diobservasi hanya melalui serangkaian proses stokastik lain yang menghasilkan suatu deretan observasi.

Dalam penelitian ini, HMM dapat menjadi salah satu cara yang digunakan untuk mengidentifikasi atau memprediksi daerah *coding* dan *non-coding* pada sekuen DNA suatu organisme atau makhluk hidup. Daerah *coding* dan *non-coding* pada DNA disebut juga dengan ekson dan intron. Menurut Tapas Kanungo [28], Tulyakov [29], Andrew W Moore [30] dan Jose Renau [31] beberapa garis besar dalam metoda HMM seperti rantai Markov, elemen HMM, permasalahan dasar HMM dan solusi permasalahan dasar HMM dapat dijelaskan pada bagian berikut ini.

3.1. Rantai Markov

Rantai Markov (*Markov Chain*) merupakan probabilitas transisi dari satu *state* ke *state* yang lain yang dapat ditentukan sesuai dengan jumlah *state* dan topologi HMM yang digunakan. Sifat rantai Markov ini bila diaplikasikan pada suatu gen eukariot dimana terdapat ekson dan intron maka ekson dan intron tersebut dapat dianggap sebagai suatu *state*. Selain itu rantai Markov juga memiliki sifat probabilitas emisi yang

terkandung pada masing-masing *state*. Sebagai penjelasan rantai Markov dapat digambarkan seperti pada Gambar 3.1.



Gambar 3.1 Transisi *state* ekson dan intron

3.2 Elemen Model *Hidden Markov*

Beberapa elemen yang digunakan pada HMM diuraikan sebagai berikut :

1. N (banyaknya *state* pada HMM).

HMM terdiri dari N buah *state* q yang diberi indeks $\{1,2,3,\dots,N$. *State* pada saat t dinyatakan sebagai q_t)

2. M (banyaknya simbol observasi yang berbeda pada setiap *state*)

Simbol-simbol observasi individual dari setiap *state* yang akan membentuk suatu deretan observasi dinyatakan dengan

$$V = \{ v_1, v_2, \dots, v_M \} \quad (3.1)$$

3. A (distribusi peluang antar *state*)

$A = \{ a_{ij} \}$ adalah distribusi peluang-peluang transisi dari *state* i menuju *state* j yang secara matematis dinyatakan sebagai:

$$a_{ij} = P[q_{t+1} = j | q_t = i], 1 \leq (i, j) \leq N$$

$$\sum_{j=1}^N a_{ij} = 1, 1 \leq i \leq N \quad (3.2)$$

4. B (distribusi peluang simbol observasi dalam suatu *state*)

$B = \{b_j(k)\}$ adalah distribusi peluang simbol observasi ke- k pada *state* j , yang secara matematis dinyatakan sebagai :

$$b_j(k) = P[o_t = v_k | q_t = j], \quad 1 \leq k \leq M \quad \text{dan} \quad 1 \leq j \leq N$$

$$\sum_{k=1}^M b_j(k) = 1, \quad 1 \leq j \leq N \quad (3.3)$$

5. π (distribusi peluang *state* inisial)

$\pi = \{\pi_i\}$ menyatakan distribusi peluang suatu sistem berada pada *state* i saat $t = 1$ (keadaan awal) yang secara matematis dinyatakan sebagai:

$$\pi_i = P[q_1 = i], \quad 1 \leq i \leq N$$

$$\sum_{i=1}^N \pi_i = 1 \quad (3.4)$$

6. *State* set dinotasikan : $Q = \{1, 2, \dots, N\}$.

7. Model dinotasikan : λ .

Sebuah HMM dapat direpresentasikan dengan $\lambda = (A, B, \pi)$, dimana A, B , dan π berturut-turut menyatakan distribusi peluang transisi antar *state*, distribusi peluang simbol observasi, dan distribusi peluang *state* inisial.

3.3 Permasalahan dasar HMM

Suatu model $\lambda = (A, B, \pi)$ mempunyai tiga permasalahan dasar yang harus dipecahkan supaya dapat diaplikasikan secara optimal, yaitu:

3.3.1 Evaluasi

Masalah pertama dalam HMM yang harus dipecahkan adalah bagaimana caranya menghitung nilai $P(O|\lambda)$ secara efisien untuk setiap model λ , yaitu peluang *sequence* observasi $O = \{o_1, o_2, \dots, o_T\}$ apabila diberikan model λ . Probabilitas ini menyatakan seberapa cocok HMM tersebut dalam memodelkan *sequence*

observasi yang terjadi. Solusi dari masalah pertama ini akan menentukan model mana yang paling sesuai untuk menghasilkan *sequence* observasi yang bersangkutan. $P(O|\lambda)$ sering disebut sebagai *likelihood*.

3.3.2 Decoding

Masalah kedua dari HMM yang harus dipecahkan adalah bagaimana caranya menentukan *state* $Q = \{q_1, q_2, \dots, q_T\}$ yang optimal (paling cocok dalam memodelkan kejadian sebenarnya) apabila diberikan *sequence* observasi $O = \{o_1, o_2, \dots, o_T\}$. *Sequence state* ini merupakan bagian yang tersembunyi (*hidden*) dari model $\lambda = (A, B, \pi)$ karena beberapa buah *state* bisa memiliki simbol observasi yang sama.

3.3.3 Training

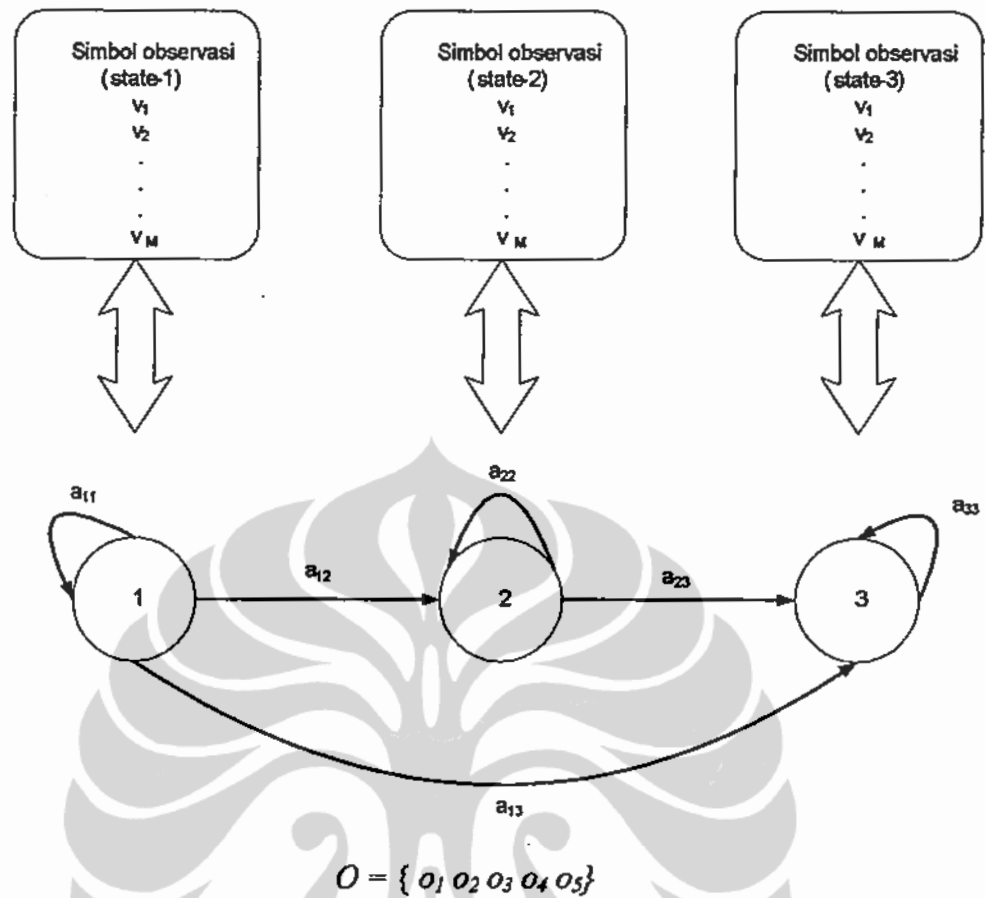
Masalah ketiga dalam HMM yang harus dipecahkan adalah bagaimana caranya memilih model $\lambda = (A, B, \pi)$ yang sesuai untuk mengoptimalkan $P(O|\lambda)$ atau *likelihood* sehingga model tersebut paling cocok dalam merepresentasikan bagaimana suatu *sequence* observasi terjadi. Untuk bisa memperoleh model yang paling sesuai, HMM harus di *training* menggunakan *sequence* observasi *training* yang akan melakukan adaptasi terhadap parameter-parameter HMM sehingga menghasilkan model yang terbaik.

3.4 Solusi untuk setiap permasalahan dasar HMM

Berikut ini adalah solusi-solusi pemecahan masalah atas ketiga hal dasar permasalahan dalam HMM yang telah disebutkan di atas.

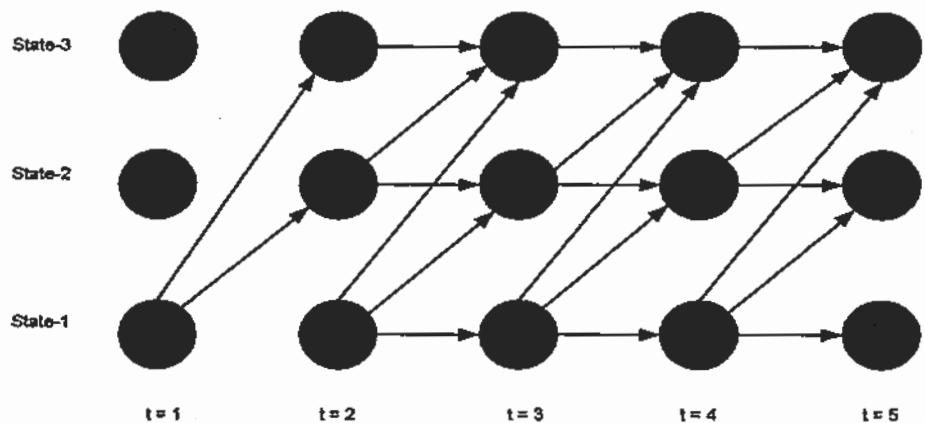
3.4.1 Solusi Evaluasi

Misalkan ada suatu topologi HMM yang memiliki tiga buah *state* ($N = 3$) dan lima buah simbol observasi ($T = 5$) seperti Gambar 3.2 di bawah ini.



Gambar 3.2 HMM dengan tiga *state* dan lima buah simbol observasi

Dari model seperti Gambar 3.2 akan dicari berapa besarnya nilai $P(O|\lambda)$ dan semua kemungkinan *state* yang bisa terjadi pada model ini untuk $T = 5$ dapat digambarkan dalam suatu diagram *trellis* seperti Gambar 3.3.



Gambar 3.3 Diagram *trellis* untuk seluruh kemungkinan transisi *state*

Jalur diagram *trellis* diatas yang ditandai dengan garis (merupakan salah satu kemungkinan yang bisa terjadi di antara seluruh kemungkinan yang ada), pada Gambar 3.3 akan digunakan sebagai ilustrasi untuk mencari nilai $P(O|\lambda)$. Misalkan jalur tersebut adalah kemungkinan ke- m yang dapat terjadi dari seluruh p kemungkinan yang ada ($m \leq p$), maka $P(O^{(m)}, Q^{(m)} | \lambda)$ dapat dihitung dengan cara:

$$1. P(O^{(m)}|Q^{(m)}, \lambda) = \prod_{t=1}^5 P(O_t^{(m)}|q_t, \lambda) \quad (3.5)$$

$$P(O^{(m)}|Q^{(m)}, \lambda) = P(O_1^{(m)}|q_1, \lambda) \times P(O_2^{(m)}|q_2, \lambda) \times \dots \times P(O_5^{(m)}|q_5, \lambda)$$

$$P(O^{(m)}|Q^{(m)}, \lambda) = b_{q_1}^{(m)} b_{q_2}^{(m)} b_{q_3}^{(m)} b_{q_4}^{(m)} b_{q_5}^{(m)} = b_1^{(m)} b_1^{(m)} b_2^{(m)} b_3^{(m)} b_3^{(m)}$$

$$2. P(Q^{(m)} | \lambda) = \pi_1 \prod_{t=1}^4 a_{q_t, q_{t+1}} \quad (3.6)$$

$$P(Q^{(m)} | \lambda) = \pi_1 a_{q_1 q_2} a_{q_2 q_3} a_{q_3 q_4} a_{q_4 q_5} = \pi_1 a_{11} a_{12} a_{23} a_{33}$$

$$3. P(O^{(m)}, Q^{(m)} | \lambda) = P(O^{(m)}|Q^{(m)}, \lambda) \times P(Q^{(m)} | \lambda) \quad (3.7)$$

$$P(O^{(m)}, Q^{(m)} | \lambda) = \pi_1 b_1^{(m)} a_{11} b_1^{(m)} a_{12} b_2^{(m)} a_{23} b_3^{(m)} a_{33} b_3^{(m)}$$

$P(O|\lambda)$ adalah nilai probabilitas keseluruhan yang harus dihitung dengan cara menjumlahkan semua nilai $P(O^{(m)}, Q^{(m)} | \lambda)$ untuk setiap jalur yang mungkin. Secara matematis :

$$P(O|\lambda) = \sum_{m=1}^p P(O^{(m)}, Q^{(m)} | \lambda) = \sum_{m=1}^p P(O^{(m)}|Q^{(m)}, \lambda) \times P(Q^{(m)} | \lambda) \quad (3.8)$$

Dimana p menyatakan banyaknya kemungkinan jalur yang dapat ditempuh pada diagram *trellis* di atas. Komputasi yang diperlukan untuk prosedur perhitungan $P(O|\lambda)$ di atas sangat rumit. Ada suatu cara yang dapat digunakan untuk

menyederhanakan seluruh rangkaian komputasi di atas, cara tersebut di kenal dengan nama prosedur *forward-backward*.

A. Prosedur *forward*

Misalkan terdapat suatu variabel probabilitas *forward* pada waktu ke- t dan suatu *state* ke- i yang dinotasikan dengan $\alpha_t(i)$ dimana secara matematis :

$$\alpha_t(i) = P(o_1 o_2 \dots o_t, q_t = i | \lambda) \quad (3.9)$$

Fungsi probabilitas *forward* ini dapat diselesaikan untuk N buah *state* dan T buah simbol observasi secara iteratif dengan algoritma sebagai berikut:

1. Inisialisasi ($t=1$)

$$\alpha_1(i) = \pi b_i(o_1), 1 \leq i \leq N \quad (3.10)$$

2. Induksi

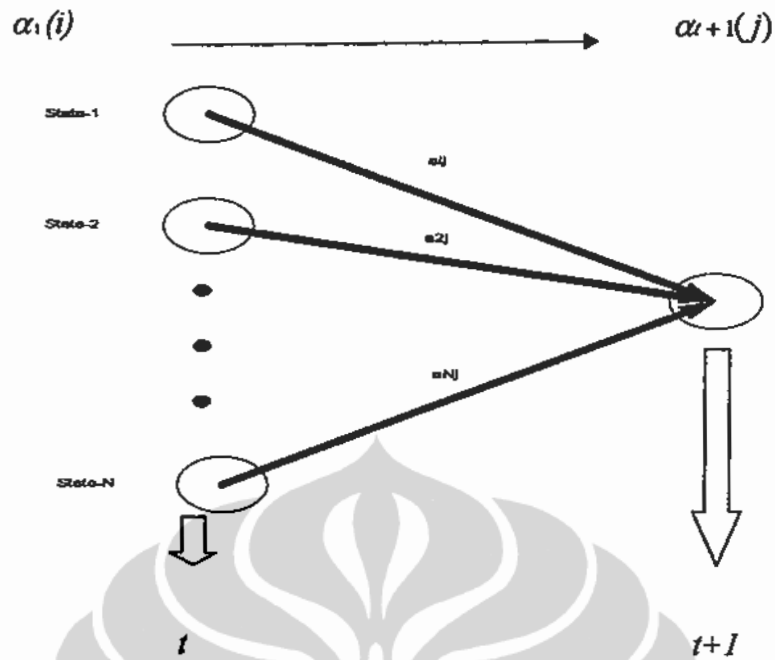
$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N (\alpha_t(i) \times a_{ij}) \right] \times b_j(o_{t+1}), 1 \leq (i, j) \leq N, 1 \leq t \leq T-1 \quad (3.11)$$

pada tahap ini suatu *state* j dapat di capai dari setiap *state* i . Ilustrasi dari tahap induksi dapat dijelaskan pada Gambar 3.4.

3. Terminasi

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (3.12)$$

Tahap terakhir dengan prosedur *forward* adalah terminasi. Pada tahap ini $P(O|\lambda)$ dihitung dengan cara menjumlahkan seluruh probabilitas *forward* untuk setiap *state* yang ada pada waktu observasi ke- T . Nilai dari $P(O|\lambda)$ ini menyatakan seberapa tepat HMM memodelkan suatu deretan observasi yang terjadi.



Gambar 3.4 Ilustrasi tahap induksi pada prosedur *forward*

B. Prosedur *backward*

Metode ini disebut prosedur *backward* karena perhitungan nilai probabilitas $P(O|\lambda)$ dicari secara mundur (*backward*) dari waktu observasi maksimum ($t=T$) sampai dengan waktu observasi awal ($t=1$).

Misalkan terdapat suatu variabel probabilitas *backward* pada waktu ke- t dan pada state ke- i yang dinotasikan dengan $\beta_t(i)$ dimana secara matematis:

$$\beta_t(i) = P(O_{t+1} O_{t+2} \dots O_T, q_t = i | \lambda) \quad (3.13)$$

Ada dua tahap yang harus dilakukan untuk memperoleh nilai probabilitas $P(O|\lambda)$ menggunakan prosedur *backward*, yaitu :

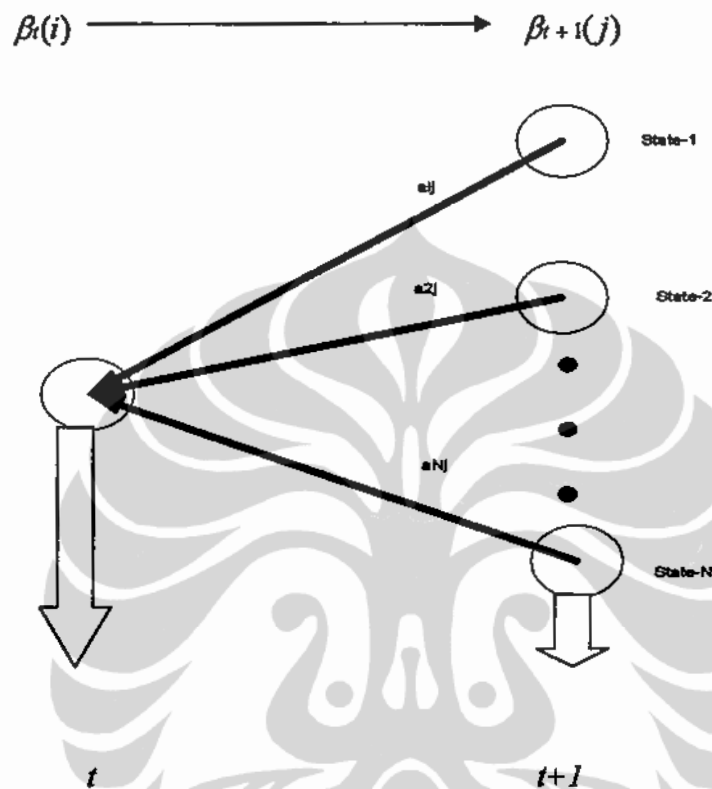
1. Inisialisasi

$$\beta_t(i) = 1, \quad 1 \leq i \leq N \quad (3.14)$$

2. Induksi

$$\beta_t(i) = \sum_{j=1}^N a_{ij}(o_{t+1}) \beta_{t+1}(j), t = T-1, T-2, \dots, 1, 1 \leq (i, j) \leq N \quad (3.15)$$

tahap induksi pada prosedur *backward* ini akan lebih mudah dipahami dengan bantuan Gambar 3.5 berikut ini.



Gambar 3.5 Ilustrasi tahap induksi pada prosedur *backward*

C. Prosedur *forward-backward*

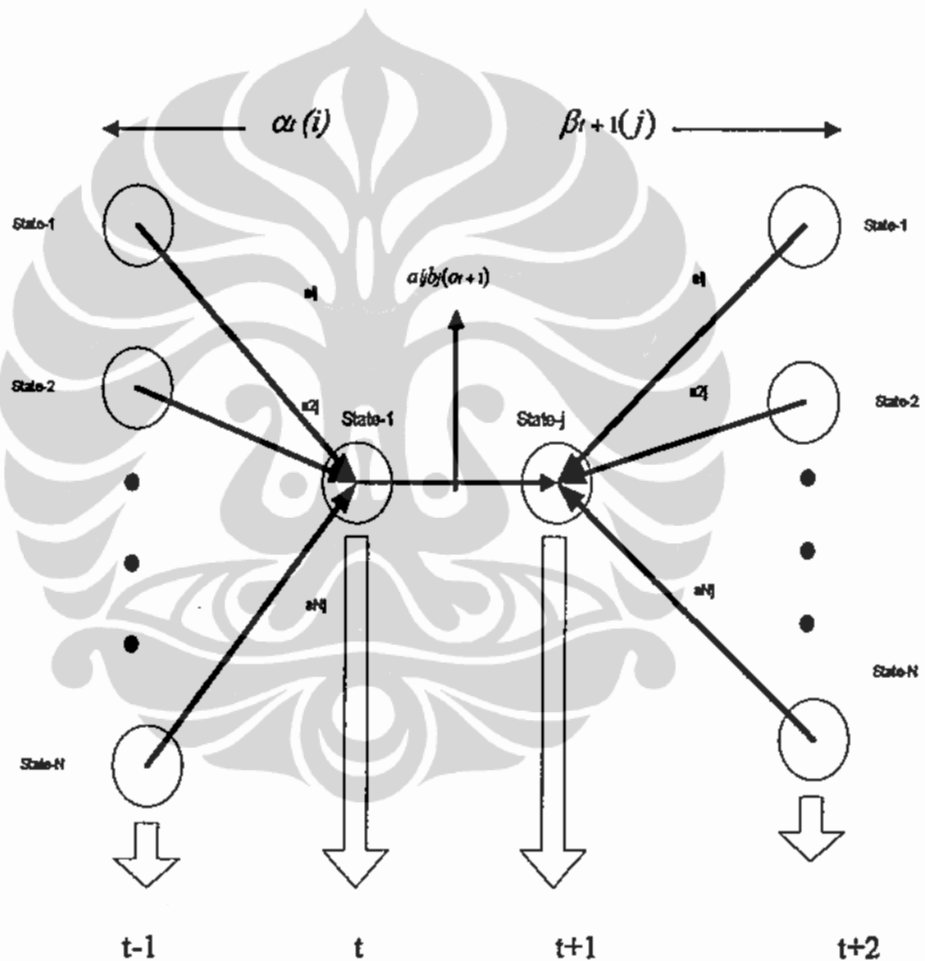
Nilai $P(O|\lambda)$ dapat dihitung menggunakan prosedur gabungan antara fungsi probabilitas *forward* $\alpha(i)$ dan fungsi *backward* $\beta(i)$. Ilustrasi prosedur ini dapat dijelaskan dengan Gambar 3.6.

Peluang berada pada saat *state* ke- i pada waktu t dari N buah *state* sebelumnya pada waktu $t-1$ di hitung menggunakan fungsi probabilitas *forward* $\alpha(i)$. Peluang transisi dari *state* ke- i pada waktu t menuju *state* ke- j pada waktu $t+1$ dan mengambil sebuah simbol observasi pada *state* ke- j tersebut adalah $a_{ij}b_j(o_{t+1})$. Fungsi probabilitas

probabilitas *backward* $\beta_{t+1}(j)$ digunakan untuk menghitung peluang munculnya deretan simbol observasi yang dimulai dari waktu $t+1$ sampai T .

Secara matematis perhitungan $P(O|\lambda)$ menggunakan prosedur *forward-backward* dinyatakan sebagai berikut :

$$P(O|\lambda) = \sum_{i=1}^N \sum_{j=1}^N \alpha_i(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) = \sum_{i=1}^N \alpha_i(i) \beta_i(i) \quad (3.16)$$



Gambar 3.6 Ilustrasi prosedur *forward-backward*

3.4.2 Solusi *decoding*

Inti dari masalah kedua ini adalah mencari deretan *state* yang tersembunyi (*hidden*) untuk suatu deretan observasi yang dihasilkan dari model λ . Metode yang

yang digunakan untuk mencari deretan *state optimal* ini adalah menggunakan algoritma *Viterbi* dan *posterior* sebagai berikut :

1. Algoritma Viterbi

Algoritma Viterbi memaksimalkan nilai probabilitas $P(Q|O, \lambda)$ sehingga akan dihasilkan deretan *state* yang optimal. Secara matematis dinyatakan sebagai berikut:

$$P(Q|O, \lambda) = \frac{P(Q, O | \lambda)}{P(O | \lambda)} \quad (3.17)$$

Di dalam algoritma Viterbi, probabilitas maksimum berada pada *state* ke- i pada waktu t dimana sebelumnya telah menempuh deretan *state* $q_1 q_2 \dots q_{i-1}$, apabila diberikan deretan observasi O dan suatu model λ dinotasikan dengan $\delta(i)$ dimana secara matematis :

$$\delta(i) = \max P[q_1 q_2 \dots q_{t-1}, q_t = i, o_1 o_2 \dots o_T | \lambda] \quad (3.18)$$

deretan *state* yang *optimal* kemudian diurutkan mundur (*backtracked*) menggunakan fungsi $\psi_t(j)$. Tahap tahap lengkap dari algoritma Viterbi adalah sebagai berikut:

1. Inisialisasi

$$\begin{aligned} \delta_1(i) &= \pi_i b_i(o_1), \quad 1 \leq i \leq N \\ \psi_1(i) &= 0 \end{aligned} \quad (3.19)$$

2. Rekursi

$$\begin{aligned} \delta_t(j) &= \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(o_t), 1 \leq (i, j) \leq N, 2 \leq t \leq T \\ \psi_t(j) &= \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], 1 \leq (i, j) \leq N, 2 \leq t \leq T \end{aligned} \quad (3.20)$$

3. Terminasi

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)]$$

$$q_t^* = \underset{1 \leq i \leq N}{\operatorname{arg\,max}} [\delta_T(i)] \quad (3.21)$$

4. Backtracking

$$q_t^* = \psi_{t+1}(q_{t+1}^*), t=T-1, T-2, \dots, 1 \quad (3.22)$$

2. Algoritma Posterior

Algoritma posterior digunakan juga untuk memaksimalkan nilai probabilitas $P(Q|O, \lambda)$. Algoritma posterior biasa mempunyai nilai matematis.

$$\text{Jika } \gamma_t(i) = P(X_t = s_i | O = 0; \lambda)$$

Dimana $\gamma_t(i)$ merupakan probabilitas pada *state* i dan waktu t dengan deretan observasi O , dan model λ dan $\xi_t(i, j)$ menyatakan probabilitas berada pada *state* ke- i dan berpindah menuju *state* ke- j pada waktu $t+1, X_t$ yang secara matematis dituliskan sebagai berikut:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O=o|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^n \alpha_t(i)\beta_t(i)} = t = \overline{1, L}; i = \overline{1, N} \quad (3.23)$$

$$X_t = \operatorname{arg\,max} \{ \gamma_t(i) \} = t = \overline{1, L} \quad (3.24)$$

State tersebut memaksimalkan probabilitas *marginal posterior*, yang memberikan solusi kepada *marginal decoding*.

Algoritma posterior memiliki beberapa kelebihan, diantaranya adalah :

1. Probabilitas posterior menghasilkan nilai tengah alternatif, untuk men-*decode sequence* observasi. Algoritma ini biasanya lebih baik dari Viterbi, terutama jika diketahui banyak jalur (*paths*) yang mempunyai nilai probabilitas yang hampir sama.

2. *Posterior decoding* berguna jika kita ingin mengetahui lebih dalam nilai *state* pada posisi tertentu, bukan pada keseluruhan *state* dari *sequence*.

3.4.3 Solusi *training*

Masalah ketiga ini adalah masalah yang tersulit untuk dipecahkan dibandingkan dengan dua masalah sebelumnya. Inti dari masalah ketiga ini adalah menyesuaikan (*training*) parameter-parameter (A, B, π) berdasarkan kriteria *optimal* tertentu. Metode yang biasa digunakan untuk memecahkan masalah ketiga ini adalah algoritma Baum-Welch atau sering disebut dengan algoritma *forward-backward*. Algoritma ini merupakan suatu metode iteratif yang berfungsi untuk mencari nilai-nilai maksimum lokal (*local maximas*) dari fungsi probabilitas $P(O|\lambda)$. Proses *training* ini berlangsung terus sampai kondisi kritis (*local minimal*) terpenuhi. Model hasil *training* harus lebih baik daripada model sebelumnya.

Algoritma Baum-Welch ini merupakan implikasi ide dasar dari Expectation-Maximization yang dituliskan dalam rumus-rumus seperti dibawah ini :

jika $\xi(i, j)$ merupakan probabilitas dari *state* i pada waktu t dan *state* j pada waktu $t+1$

maka:

$$\begin{aligned} \xi(i, j) &= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \end{aligned} \quad (3.25)$$

Dimana $\gamma_t(i)$ Sebagai probabilitas dari *state* i pada waktu t dari *sequence* observasi

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j)$$

$\sum_{i=1}^T \gamma_t(i)$ banyaknya nilai yang diharapkan pada waktu *sate* *i*

$\sum_{i=1}^{T-1} \xi_t(i, j)$ banyaknya nilai yang diharapkan pada transisi dari *state* *i* ke *state* *j*

$\bar{\pi}_i$ = frekwensi yang diharapkan dari *state* *i* pada waktu *t* ($t=1$) = $\gamma_1(i)$

\bar{a}_{ij} = (banyaknya nilai yang diharapkan pada perpindahan transisi dari *state* *i* ke *state* *j*)/(banyaknya ilai yang diharapkan dari *state* *i*)

$$\bar{a}_{ij} = \frac{\sum \xi_t(i, j)}{\sum \gamma_t(i)} \quad (3.26)$$

$\bar{b}_j(k)$ = (banyaknya nilai yang diharapkan dari *state* *j* dan simbol observasi *k*)/
(banyaknya nilai yang diharapkan dari *state* *j*)

$$\bar{b}_j(k) = \frac{\sum_{t, o_t = k} \gamma_t(j)}{\sum \gamma_t(j)} \quad (3.27)$$

Elemen-elemen HMM, permasalahan dasar dan solusi permasalahan dasar yang dijabarkan dengan rumus-rumus untuk solusi *training* dan solusi *decoding* (*testing*) dengan menggunakan kedua algoritma yaitu algoritma Viterbi dan Posterior (Baum-Welch) sesuai dengan struktur DNA yang terdiri dari ekson dan intron yang dapat di asumsikan sebagai suatu *state*. Oleh karena itu, penelitian ini menggunakan metode HMM untuk memprediksi rangkaian ekson DNA *Plasmodium falciparum* dengan simbol observasi adalah simbol yang terdapat pada ekson dan intron yang terdiri dari empat basa DNA yaitu A, T, G dan C. Nilai transisi *state* ditentukan secara acak dan merupakan nilai transisi dari satu *state* ke *state* lainnya atau nilai transisi pada *state* itu sendiri yang di

dalam sekuen DNA merupakan transisi dari *state* ekson ke *state* intron. Kemudian nilai emisi *state* adalah nilai distribusi basa-basa DNA pada tiap *state*.

Rantai Markov pada metode HMM sesuai dengan struktur ekson dan intron yang terdapat pada DNA dimana letak ekson dan intron harus bergantian (*alternately located*) sehingga berpengaruh pula pada nilai-nilai transisi *state* yang diberikan secara acak sehingga akan mempengaruhi pula pada hasil kinerja model. Kinerja model ditentukan oleh perbandingan antara estimasi *state* hasil *decoding* dengan *state* asli sekuen sesuai dengan jumlah *state* yang di rancang pada struktur model.

Pengembangan struktur model pada penelitian ini dilakukan dengan penambahan jumlah *state* secara acak dengan cara memisahkan kodon-kodon atau basa-basa pada *state* ekson maupun *state* intron pada sekuen DNA tersebut.

BAB 4

PERANCANGAN STRUKTUR MODEL HIDDEN MARKOV DAN SIMULASI

4.1 Perancangan Struktur Model

Struktur model *Hidden Markov* yang dirancang pada penelitian ini sesuai dengan struktur DNA gen *eukariot* yang terdiri dari lokasi ekson dan intron yang letaknya bergantian pada region *coding sequence* (CDS) dan minimal terdapat 2 lokasi ekson dan 1 lokasi intron seperti pada Gambar 4.2.

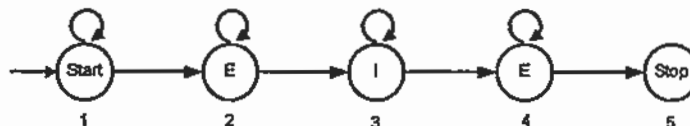


Gambar 4.2 Struktur *Coding Sequence* (CDS)

Beberapa rancangan struktur model dapat dilakukan dengan jumlah *state* yang berbeda.

4.1.1 Struktur Dasar Model 5 *state*

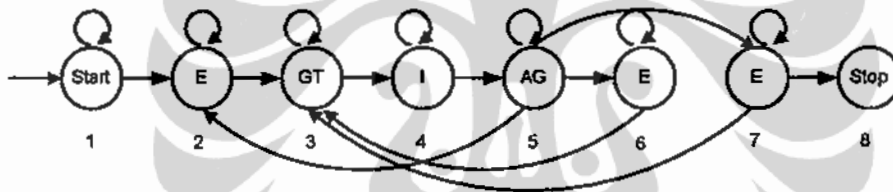
Pada Gambar 4.1 ekson pertama dan ekson terakhir masing-masing dapat dibuat menjadi 2 *state* yaitu dengan cara memisahkan kodon start dari ekson pertama dan kodon stop dari ekson terakhir. Selanjutnya diantara kedua ekson terdapat *state* intron sehingga struktur model memiliki 5 *state* dan menjadi struktur dasar model seperti pada Gambar 4.2.



Gambar 4.2 Struktur Model 5 *state*

4.1.2 Struktur Model 8 state

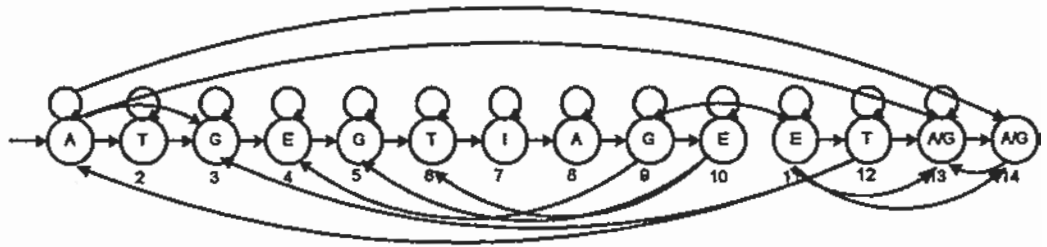
Setiap adanya lokasi intron pada sekuen DNA biasanya diawali dengan dua basa GT dan diakhiri dengan dua basa AG. Model ini membuat setiap lokasi intron menjadi 3 state sedangkan untuk ekson pertama dan terakhir tetap masing-masing dua state. Pada model ini ada penambahan satu state ekson tetapi state ini tidak dapat dihubungkan ke state ekson terakhir tetapi bisa *backward* ke state intron. Model ini mencoba parameter *forward-backward* metode HMM dengan tetap mengacu pada lokasi ekson dan intron pada CDS yang letaknya bergantian tersebut. Sehingga keseluruhan model ini memiliki jumlah state 8 dan struktur model seperti pada Gambar 4.3 dibawah ini.



Gambar 4.3 Struktur Model 8 state

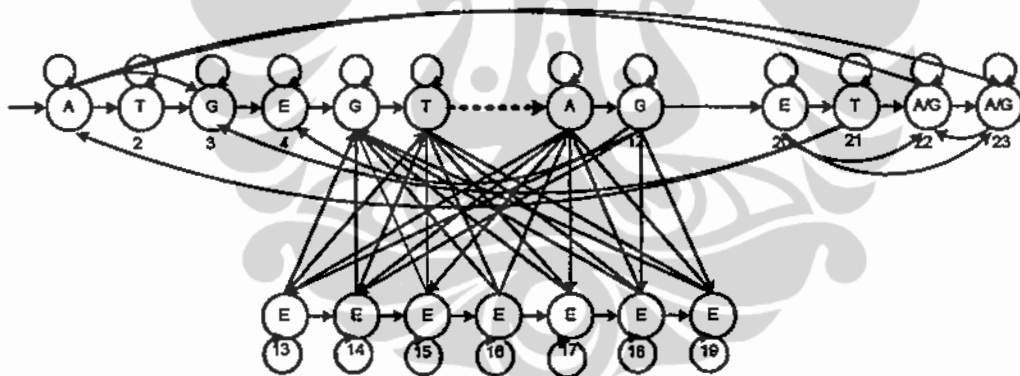
4.1.3 Struktur Model 14 state

Model ini adalah pengembangan dari model 8 state dengan 1 state kodon start (ATG) yang terdiri dari 3 buah basa dipisah menjadi 3 state yang masing-masing terdiri dari satu basa. Kodon stop pada DNA ada 3 yaitu TAA, TAG dan TGA, tetapi setiap sekuen DNA pada CDS hanya memiliki satu dari ketiga kodon stop sehingga 1 state kodon stop dapat pula dipisah menjadi 3 state. Sedangkan state GT dan AG pada intron dipisah menjadi masing-masing 2 state sehingga pada bagian intron menjadi 5 state. Model ini total memiliki 14 state dengan *forward-backward* seperti yang terdapat pada Gambar 4.4.

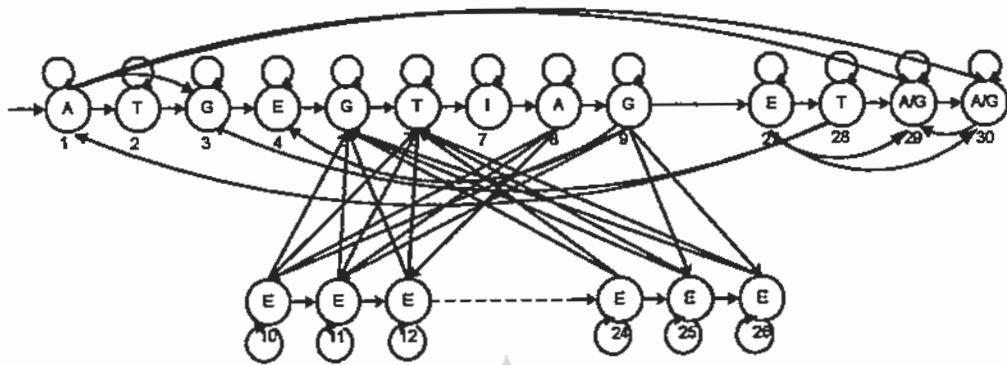


Gambar 4.4 Struktur model 14 *state*

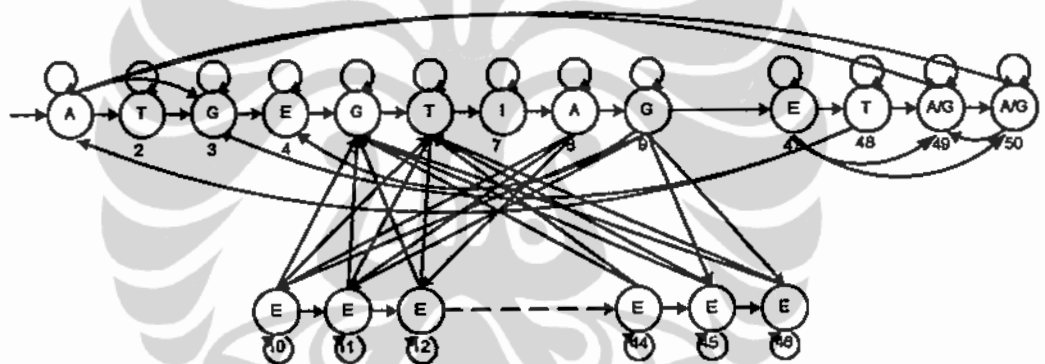
Proses penambahan *state* pada model dapat dilakukan pada bagian ekson dan intron yang terdapat pada CDS dengan jumlah sembarang/acak tetapi tetap mengikuti struktur DNA gen tersebut. Model-model tersebut mempunyai jumlah *state* 23, 30, 50 dan 100 dengan struktur seperti pada Gambar 4.5, Gambar 4.6, Gambar 4.7 dan Gambar 4.8.



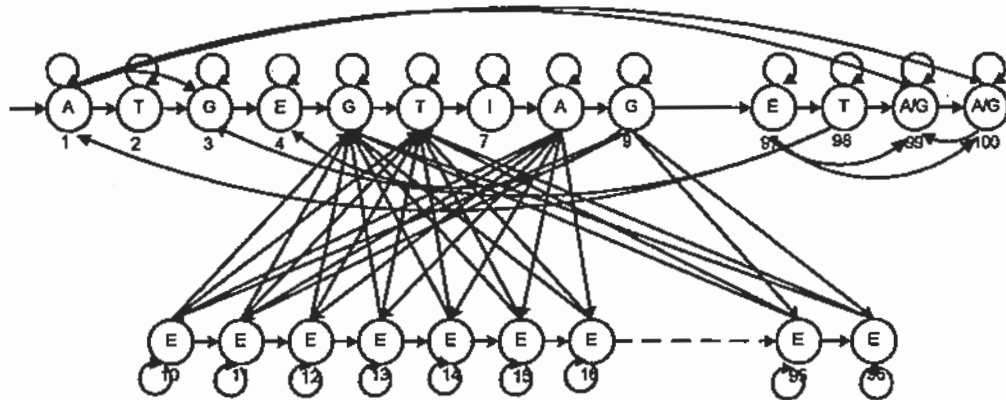
Gambar 4.5 Struktur model 23 *state*



Gambar 4.6 Struktur model 30 state



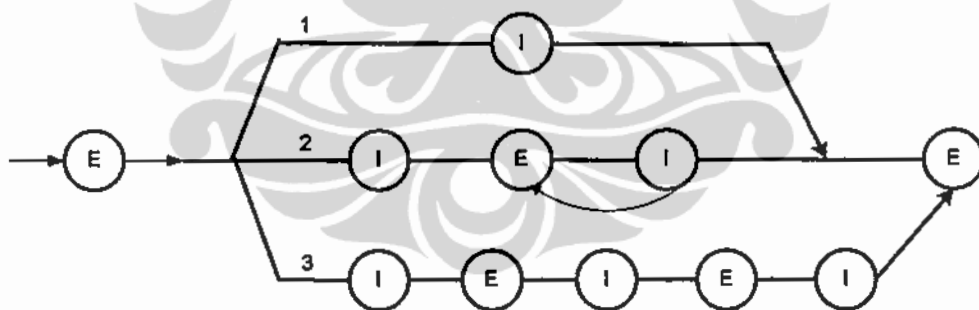
Gambar 4.7 Struktur model 50 state



Gambar 4.8 Struktur model 100 state

4.1.4. Pengembangan Struktur Dasar Model

Pengembangan struktur model *hidden* Markov selain untuk model dengan minimal 2 ekson yang ada pada CDS dapat pula dirancang untuk model dengan jumlah ekson lebih dari 2 seperti Gambar 4.9 pada jalur 2 dan 3.



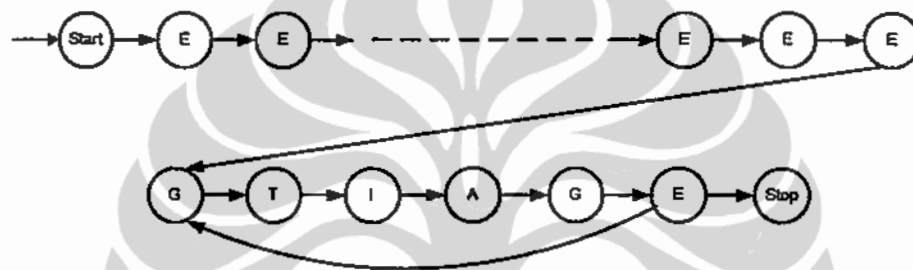
Gambar 4.9 Pengembangan struktur dasar model

Pada Gambar 4.9 diatas, jalur 1 menggambarkan model dengan jumlah lokasi ekson minimal yang terdapat pada CDS yaitu 2 lokasi ekson. Selanjutnya untuk model-model ini dicoba dengan arah *forward-backward* berbeda dari model pada Gambar 4.2 sampai dengan Gambar 4.8. Sedangkan untuk jalur 2 dan 3 menggambarkan model untuk

sekuen yang memiliki jumlah lokasi ekson pada CDS lebih dari 2 dan mempunyai bentuk umum model yang berbeda.

A. Struktur Model 1

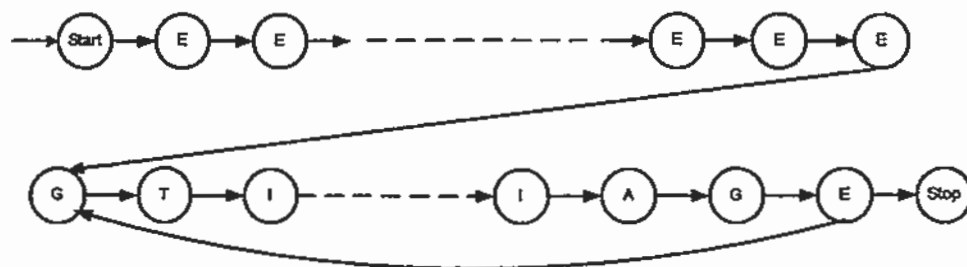
Model ini adalah ekspansi model pada Gambar 4.9 jalur 1 dengan penambahan *state* hanya dilakukan pada lokasi ekson pertama saja. Sedangkan untuk intron terdiri dari 5 *state* dan ekson terakhir terdiri dari 2 *state*. Gambar 4.10 menggambarkan struktur umum untuk model 1 ini.



Gambar 4.10 Struktur umum Model 1

B. Struktur Model 2

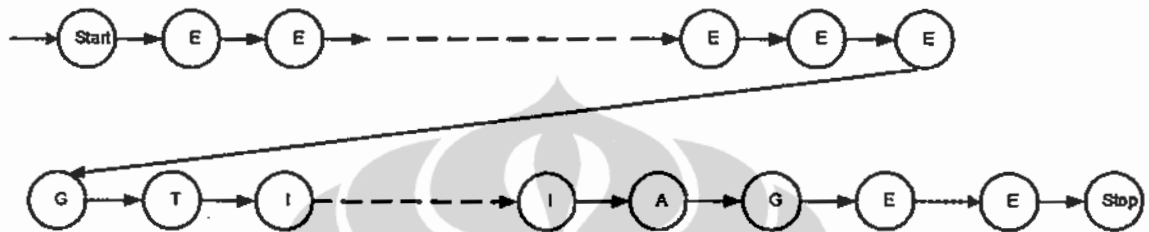
Model ini merupakan ekspansi dari struktur umum Model 1 pada Gambar 4.10 dengan penambahan *state* dilakukan pada bagian intron setiap sekuen sehingga struktur model seperti Gambar 4.11 dibawah ini.



Gambar 4.11 Struktur umum Model 2

C. Struktur Model 3

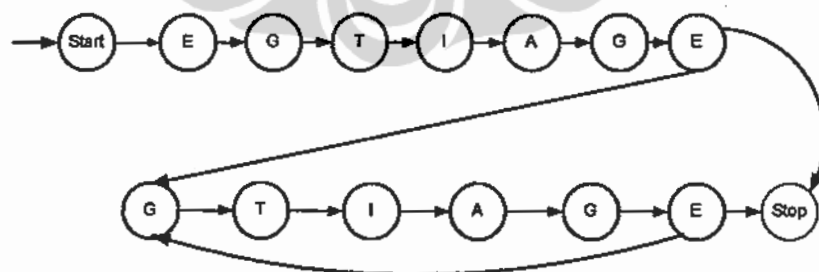
Model ini merupakan ekspansi dari struktur umum Model 2 pada Gambar 4.11 dengan penambahan *state* dilakukan pada bagian ekson terakhir setiap sekuen sehingga semua lokasi ekson dan intron terdapat penambahan jumlah *state*. Struktur umum model ini seperti Gambar 4.12.



Gambar 4.12 Struktur umum Model 3

D. Struktur Model 4

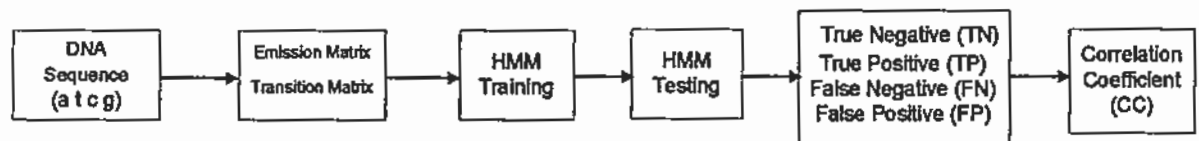
Model ini merupakan struktur model pada Gambar 4.9 untuk jalur 2 dan 3 dan juga merupakan pengembangan struktur umum Model 1, 2 dan 3. Penambahan *state* pada lokasi intron pertama dari tiap sekuen terpisah dari bagian intron yang berulang dengan struktur umum model ini seperti Gambar 4.13.



Gambar 4.13 Struktur umum Model 4

4.2. Perancangan Simulasi

Perancangan simulasi dapat dijelaskan dengan skema blok proses simulasi seperti Gambar 4.14 dibawah ini.



Gambar 4.14 Skema blok proses simulasi

4.2.1 Input Sekuen DNA

Sebagai input proses simulasi sesuai Gambar 4.14 diatas yaitu sekuen-sekuen DNA *Plasmodium falciparum* yang diperoleh dari situs <http://www.ncbi.nlm.nih.gov/entrez/> [3,18]. Pada halaman situs, dengan mengganti kata 'PubMed' pada kolom 'search' dengan kata 'Nucleotide' dan mengisi kolom 'for' dengan *accession number* (nomor akses) atau nama suatu organisme yang akan dicari (dalam bahasa Inggris atau Latin). Setelah itu dengan menekan tombol ikon 'Go' atau tekan Enter (↵) maka data sekuen tersebut diperoleh dengan format GenBank. Data masukan sekuen DNA yang digunakan pada simulasi sebanyak 152 sekuen dan rincian data terdapat pada Lampiran 1. Sebagai salah satu contoh data sekuen DNA yang diperoleh dari Genbank untuk salah satu jenis gen organisme *Plasmodium falciparum* dengan *accession number*: NC_004318 yang mempunyai keterangan seperti pada Gambar 4.15.

```

LOCUS      NC_004318      862 bp      DNA      linear      INV 30-JAN-2006
DEFINITION Plasmodium falciparum 3D7 chromosome 4, PFD0750w gene,
complete CDS.
ACCESSION  NC_004318 REGION: 1..1204112
VERSION    NC_004318.1 GI:23509994
KEYWORDS   .
SOURCE     Plasmodium falciparum 3D7
ORGANISM   Plasmodium falciparum 3D7
           Eukaryota;Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 862)
AUTHORS    Hamlin,N., Pain,A., Berriman,B., Hall,N., Bowman,S.,
Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M. And Barrell,B.
TITLE      Direct Submission
JOURNAL    Submitted (24-FEB-1999) P.falciparum Genome Sequencing
Consortium,The Sanger Centre, Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, UK
COMMENT    PROVISIONAL REFSEQ: This record has not yet been subject to
final NCBI review. The reference sequence was derived from AL844503.
FEATURES   Location/Qualifiers
           source          1..862
                        /organism="Plasmodium falciparum 3D7"
                        /mol_type="genomic DNA"
                        /isolate="3D7"
                        /db_xref="taxon:36329"
                        /chromosome="4"
           gene           1..862
                        /gene="PFD0750w"
                        /db_xref="GeneID:812555"
           CDS            join(1..156,314..862)
                        /gene="PFD0750w"
                        /note="Similarity to several nuclear cap-binding
                        proteins Similar to Arabidopsis thaliana nuclear
                        cap-binding protein cbp20 SWALL:Q9XFD1
                        (EMBL:AF140219) (257 aa) fasta scores: E(): 3.4e-
                        21, 44.3% id in 158 aa, and to Homo
                        sapiens 20 kDa nuclear cap binding protein ncbp2 or
                        cbp20 SWALL:AAH01255 (EMBL:X84157) (156 aa) fasta
                        scores: E():1.3e-18, 46.87% id in 128 aa, and to
                        Drosophila melanogaster cbp20 protein cbp20 or
                        cg12357 SWALL:Q9V3L6
                        (EMBL:AE003721) (154 aa) fasta scores: E(): 2.3e-
                        18, 48.03% id in 127 aa"

```

Gambar 4.15 Data salah satu gen *P. falciparum* NC_004318

```

/codon_start=1
/product="nuclear cap-binding protein, putative"
/protein_id="NP_702805.1"
/db_xref="GI:23510139"
/db_xref="GeneID:812555"
/translation="MAHLYEEVYKKRKYFDRALCNDYEEWQDKIRESKTVYIG
NLSIYTTQQQIYEHMSKAGDVENIIMGLHRTEKSPCGFCFVVYKKKEGYTQAV
NFLNNSILDGRIIRVDEDLGIIGKRKYGRGKTGVQKRDERNKFYDEDRPKVLD
NLVDHTINKKRKLNNNHYYERNIKQKILLPSNYTLNTTDMKNSRMVYIKPPVT
LYPNAQHMMHINKKKHQNRNLYKFHVKKDRQNTPT"

ORIGIN
  1 atggcacatt tatatgaaga ggtttataaa aagagaaaat attttgatag agctttatgt
 61 aatgattatg aagaatggca agataaaata agagaatcaa aaacggctca tataggaaat
121 ttatccatat atacaacaca acaacaaata tatgaggtaa gcaaaaagga aaatatgata
181 taatcatagg ttgtaatata tatatatata tatattatat atatataatt atttatttat
241 ttattttatt ttattttatt ttatttttat ttatcgaatg tcttcacatt atataccatt
301 taatttttcc aagcatatgt ctaaggctgg ggatgtagaa aacatcatca tgggtttaca
361 cagaaccgag aaatctcctt gtggattttg ttttgtagtt tataaaaaga aagaaggata
421 tactcaagca gtaaattttt taataaatc aatttttagat ggtagaatta taagagttga
481 tgaagacctt ggtattattg gtaaacgaaa atatggaaga gggaaaacag gagttcaaaa
541 aagagatgaa cgtaataaat tttatgatga agatagacca aaagttttag ataacttagt
601 tgatcataca attaacaaaa aaaggaaact taataataat cattattatg aaagaaatat
661 aaaacaaaaa atactcttac catcaatta tactcttaac acaaccgata tgaaaaatag
721 tagaatggtc tatataaaac caccgtaac tttatatcgg aacgctcaac atatgatgca
781 tataaataaa aaaaaacatc aaaatagaaa tttatataaa tttcacgtaa aaaaagatag
841 acaaaataat acgccaacat aa

//
Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

```

Gambar 4.15 Data salah satu gen *P. falciparum* NC_004318 (Lanjutan)

Sekuen DNA *Plasmodium falciparum* sebagai masukan diperoleh dari kromosom 1, 2, 3, 4, 5, dan kromosom 9. Di dalam kromosom tersebut terdapat banyak gen dan harus diteliti karena tidak semua sekuen DNA tersebut baik untuk di proses. Oleh karena itu ada beberapa kriteria yang harus diperhatikan waktu mencari sekuen agar dapat digunakan dengan baik seperti yang sudah dijelaskan pada bagian identifikasi masalah pada bagian pendahuluan.

Label state adalah suatu parameter HMM yang dapat diaplikasikan pada DNA gen sebagai ekson dan intron dalam CDS. Setiap lokasi ekson dan intron pada CDS terdiri dari jumlah basa yang berbeda sehingga dapat digunakan untuk penambahan jumlah *state* pada model. Jika model dirancang dengan menggunakan banyak *state* maka akan didapat

banyaknya kombinasi transisi *state* dari satu *state* ke *state* lainnya. Pada proses simulasi ini di asumsikan bahwa *state* ekson itu adalah sebagai label *state* 1 dan untuk *state* intron adalah sebagai label *state* 2. Oleh karena itu di dalam perancangan simulasi diperlukan suatu indentifikator untuk mengidentifikasi banyaknya kombinasi yang terjadi sehingga hasil yang diperoleh tetap hanya ada 2 *state* yaitu *state* 1 untuk ekson dan *state* 2 untuk intron. Hal ini diperlukan karena sebagai keluaran proses simulasi adalah prediksi *state* ekson dan intron dan hasil keluaran ini akan dibandingkan dengan *state* ekson dan intron pada masukan. Sebagai contoh pemberian label *state* ekson dan intron pada salah satu sekuen asli DNA *Plasmodium falciparum* seperti Gambar 4.15 yang memiliki CDS join(1....156, 314....862) dapat dijelaskan pada Gambar 4.16.

| | | | | | | |
|-------|-------------|-------------|------------|-------------|------------|------------|
| Asli | atggcacatt | tatatgaaga | ggtttataaa | aagagaaaat | attttgatag | agctttatgt |
| State | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |
| Asli | aatgattatg | aagaatggca | agataaaata | agagaatcaa | aaacggctca | tataggaat |
| State | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |
| Asli | ttatccatat | atacaacaca | acaacaaata | tatgaggtaa | gcaaaaagga | aaatatgata |
| State | 1111111111 | 1111111111 | 1111111111 | 1111112222 | 2222222222 | 2222222222 |
| Asli | taatcatagg | ttgtaataata | tatatatata | tatattatat | atatatattt | atattattat |
| State | 2222222222 | 2222222222 | 2222222222 | 2222222222 | 2222222222 | 2222222222 |
| Asli | ttattttattt | ttattttattt | ttatttttat | ttatcgaatg | tottcacatt | atataccatt |
| State | 2222222222 | 2222222222 | 2222222222 | 2222222222 | 2222222222 | 2222222222 |
| Asli | taatttttcc | aagcatatgt | ctaaggctgg | ggatgtagaa | aacatcatca | tgggtttaca |
| State | 2222222222 | 2221111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |
| Asli | cagaaccgag | aatctcctt | gtggattttg | ttttgtagtt | tataaaaaga | aagaaggata |
| State | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |
| Asli | tactcaagca | gtaaattttt | taataattc | aatttttagat | ggtagaatta | taagagttga |
| State | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |

Gambar 4.16 Label *state* salah satu sekuen DNA *P. falciparum* NC_004318

| | | | | | | |
|-------|------------|------------|------------|------------|------------|------------|
| Asli | tgaagaccta | ggtattattg | gtaaacgaaa | atatggaaga | gggaaaacag | gagttcaaaa |
| State | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |
| Asli | aagagatgaa | cgtaataaat | tttatgatga | agatagacca | aaagttttag | ataacttagt |
| State | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |
| Asli | tgatcataca | attaacaaaa | aaaggaaact | taataataat | cattattatg | aaagaaatat |
| State | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |
| Asli | aaaacaaaaa | atactcttac | catcaaatta | tactcttaac | acaaccgata | tgaaaaatag |
| State | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |
| Asli | tagaatggtc | tatataaac | caccgtaac | tttatatccg | aacgctcaac | atatgatgca |
| State | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |
| Asli | tataaataaa | aaaaaacatc | aaaatagaaa | tttatataaa | tttcacgtaa | aaaaagatag |
| State | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 | 1111111111 |
| Asli | acaaaataat | acgccaacat | aa | | | |
| State | 1111111111 | 1111111111 | 11 | | | |

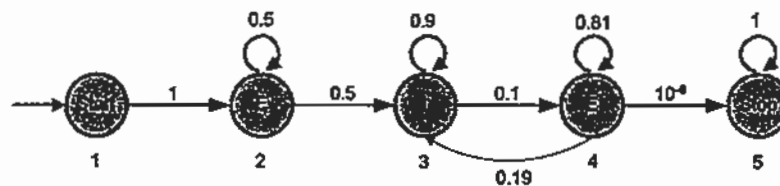
Gambar 4.16 Label *state* salah satu sekuen DNA *P. falciparum* NC_004318 (Lanjutan)

Pada Gambar 4.16 diatas menjelaskan bahwa sekuen tersebut memiliki dua lokasi ekson dan satu lokasi intron yang dapat diketahui dari CDS sekuen pada urutan 1 sampai dengan urutan 156 dan urutan 314 sampai dengan 862 adalah posisi untuk ekson dengan label *state* 1 dan berwarna biru. Sedangkan untuk posisi lokasi intron dimulai dari urutan sekuen 157 sampai dengan 313 yang diberi label *state* 2 dan berwarna merah.

4.2.2 Transisi *State* dan Emisi *State*

Transisi *state* dan emisi *state* adalah parameter-parameter yang digunakan pada proses training HMM. Nilai matriks transisi adalah random dan nilai ini merupakan probabilitas transisi *state* satu ke *state* lainnya dengan nilai total seluruh kemungkinan adalah satu. Karena nilai matriks ini random maka nilai ini sangat besar pengaruhnya terhadap hasil keluaran yang dilakukan pada proses simulasi. Sedangkan nilai komponen matrik emisi adalah persentase distribusi nukleotida DNA yang terdapat pada setiap *state* sesuai model. Untuk mencari nilai matriks ini diperlukan suatu sub-program simulasi

yang dapat mengikuti perubahan struktur *state* pada model. Sebagai contoh model Hidden Markov dengan 5 *state* dengan nilai acak transisi *state* untuk model tersebut ditentukan seperti pada Gambar 4.17.



Gambar 4.17 Nilai acak transisi *state* model 5 *state*

Nilai acak transisi *state* model dijadikan dalam bentuk matrik disebut dengan matriks transisi (t_r) yang mempunyai jumlah baris dan kolom sama banyak, sedangkan untuk emisi *state* dalam bentuk matrik disebut dengan matriks emisi (E) yang berarti banyak kolom menunjukkan jumlah basa yang pada DNA dan banyak baris menunjukkan jumlah *state* model. Sebagai contoh matrik transisi dan emisi untuk Gambar 4.17 diatas dapat dituliskan sebagai berikut :

$$t_r = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.9 & 0.1 & 0 \\ 0 & 0 & 0.19 & 0.81 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$E = \begin{bmatrix} 0.3333 & 0 & 0.3333 & 0.3333 \\ 0.4342 & 0.1094 & 0.1507 & 0.3056 \\ 0.4210 & 0.0599 & 0.0703 & 0.4487 \\ 0.4215 & 0.1173 & 0.1622 & 0.2990 \\ 0.5592 & 0 & 0.1075 & 0.3333 \end{bmatrix}$$

4.2.3 HMM Training

Proses *training* pada metode HMM untuk simulasi ini hanya menggunakan algoritma Viterbi karena waktu proses dengan menggunakan algoritma ini relatif lebih singkat bila dibandingkan dengan menggunakan algoritma Baum-Welch. Kedua matriks transisi dan emisi digunakan pada proses ini dan hasil dari proses ini adalah estimasi matriks transisi ($\text{est } t_r$) dan estimasi matriks emisi ($\text{est } E$). Misalkan untuk contoh model 5 *state* dengan nilai matriks transisi dan emisi yang terdapat seperti pada Gambar 4.17 menghasilkan nilai $\text{est } t_r$ dan nilai $\text{est } E$ dalam bentuk matriks sebagai berikut ini :

$$\text{est } t_r = \begin{bmatrix} 0 & 1.0000 & 0 & 0 & 0 \\ 0 & 0.9983 & 0.0017 & 0 & 0 \\ 0.0112 & 0 & 0.9871 & 0.0017 & 0 \\ 0 & 0 & 0 & 0.1220 & 0.8780 \\ 1.0000 & 0 & 0 & 0 & 0 \end{bmatrix}$$

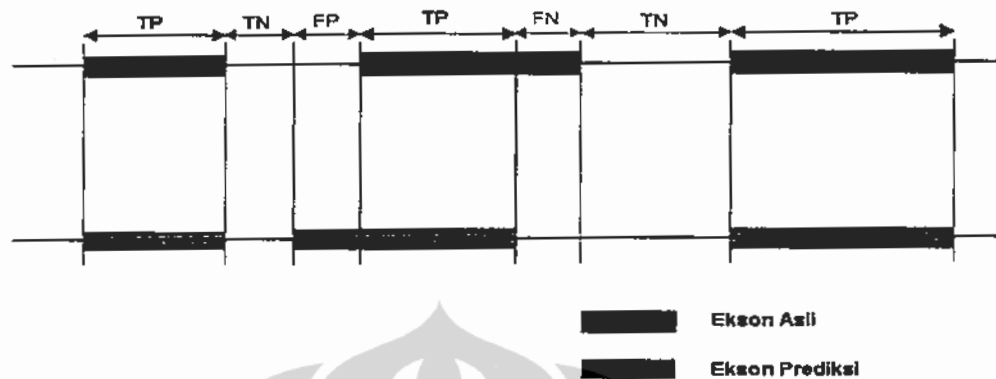
$$\text{est } E = \begin{bmatrix} 0 & 0 & 1.0000 & 0 \\ 0.4311 & 0.1114 & 0.1513 & 0.3063 \\ 0.3285 & 0.0362 & 0.0281 & 0.6072 \\ 0 & 1.0000 & 0 & 0 \\ 0.7333 & 0 & 0 & 0.2667 \end{bmatrix}$$

4.2.4 HMM Testing

Kedua nilai matriks estimasi hasil dari proses *training* digunakan untuk proses HMM *testing* terhadap sejumlah sekuen yang sama yang digunakan pada proses *training*. Algoritma yang digunakan pada HMM *testing* ini adalah algoritma Viterbi dan Baum-Welch karena waktu yang diperlukan untuk proses *testing* tidak memerlukan waktu yang lama bila dibandingkan dengan waktu proses *training*. Tetapi waktu proses *testing* dengan algoritma Viterbi relatif lebih singkat dibandingkan dengan menggunakan algoritma Baum-Welch. Hasil yang diperoleh dari HMM *testing* ini adalah estimasi *state* ekson dan estimasi *state* intron. Untuk memudahkan kombinasi *state* yang terjadi apabila model menggunakan banyak *state* maka diperlukan indentifikator untuk mengidentifikasi *state* hanya menjadi *state* ekson dan *state* intron saja. Hal ini memudahkan untuk membandingkan hasil estimasi *state* ini dengan *state* asli dari sekuen tersebut. Hasil perbandingan kedua posisi *state* (asli dan prediksi) ditentukan dengan parameter-parameter sebagai berikut :

1. *True Positive* (TP), jika posisi *state* pada sekuen asli merupakan *state* ekson dan posisi setelah hasil tes tetap merupakan *state* ekson.
2. *True Negative* (TN), jika posisi *state* pada sekuen asli merupakan intron dan posisi setelah hasil tes tetap merupakan *state* intron.
3. *False Positive* (FP), jika posisi *state* pada sekuen asli merupakan *state* intron dan posisi setelah tes berubah menjadi *state* ekson.
4. *False Negative* (FN), jika posisi *state* pada sekuen asli merupakan *state* ekson dan posisi setelah tes berubah menjadi *state* intron.

Sebagai ilustrasi mengenai TP, TN, FP dan FN dapat di jelaskan seperti Gambar 4.18.



Gambar 4.18 Ilustrasi TP, TN, FP dan FN

4.2.5 Keluaran Simulasi

Besar nilai dari parameter-parameter TP, TN, FP dan FN digunakan untuk menghitung nilai Koefisien Korelasi / *Correlation Coefficient* (CC) model dan merupakan keluaran simulasi yang menjadi sebagai kinerja model dan dituliskan dengan rumus (4.1) berikut ini [13].

$$CC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FN) \cdot (TN + FP) \cdot (TP + FP) \cdot (TN + FN)}} \quad (4.1)$$

Nilai CC berkisar dari 0 ~ 1 dan dari rumus (4.1) pada bagian numerator harus menghasilkan nilai positif sedangkan untuk nilai pada bagian denominator lebih besar dari numerator. Sehingga komponen TP harus besar dan komponen FP harus kecil agar nilai CC masih menghasilkan nilai positif. Jika nilai CC negatif berarti asumsi menjadi terbalik atau berlawanan. Oleh karena itu ketepatan prediksi ekson harus lebih banyak dibandingkan intron yang diprediksi menjadi ekson.

4.2.6 Spesifikasi Perangkat Simulasi

Spesifikasi perangkat komputer yang digunakan dalam simulasi dibagi dua bagian yaitu perangkat keras (*hardware*) dan perangkat lunak (*software*) seperti berikut :

1. Perangkat keras (*hardware*)

Sebuah komputer PC IBM Standard dengan spesifikasi sebagai berikut :

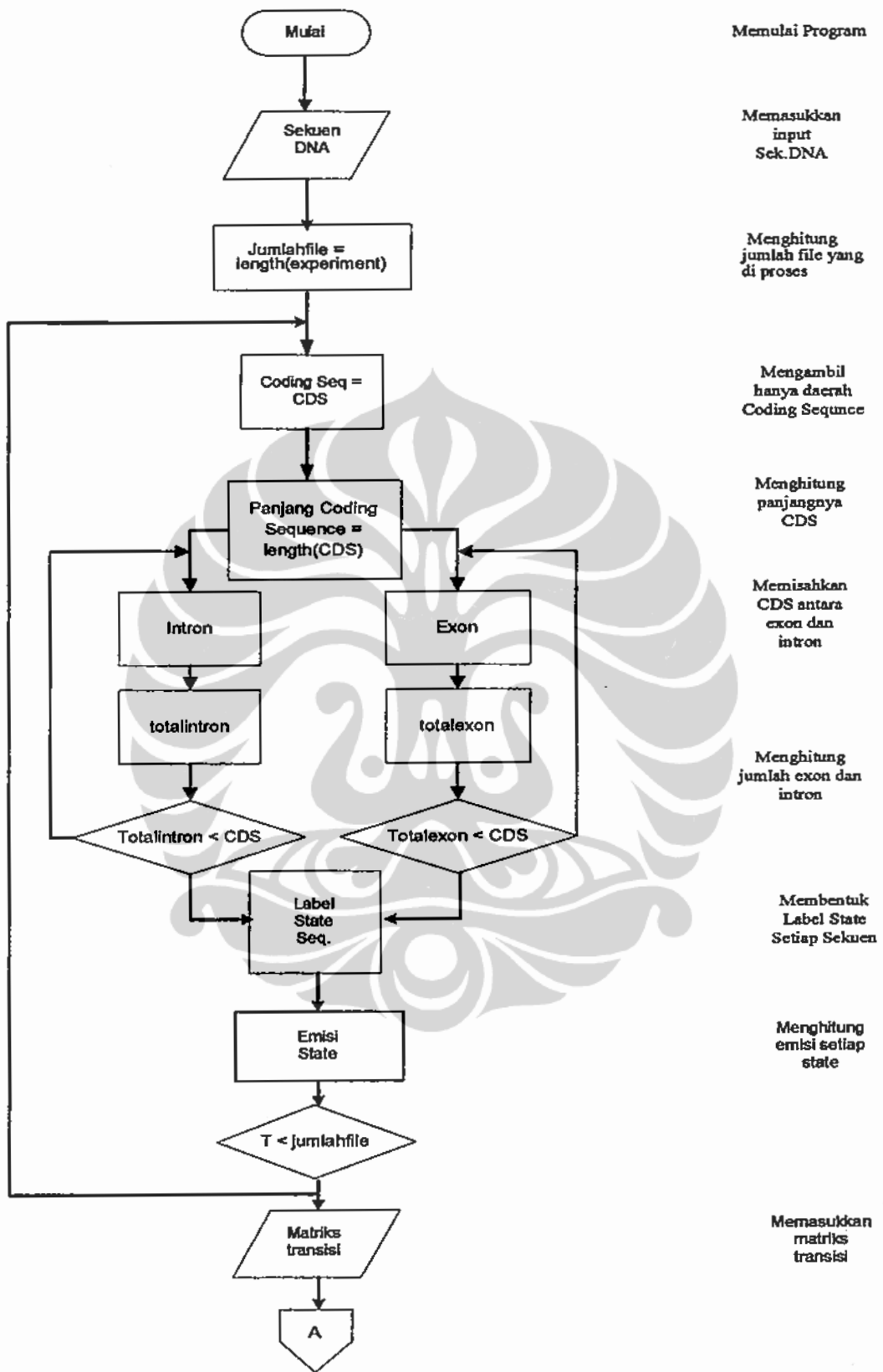
- a. *Processor* : Intel(R) Pentium(R) 4 CPU 2.8 GHz.
- b. *Memory* : 1.99 GB of RAM.
- c. *Harddisk* : 40 GB.
- d. *Operating system* : Microsoft Windows XP, 2002 version.

2. Perangkat lunak (*software*)

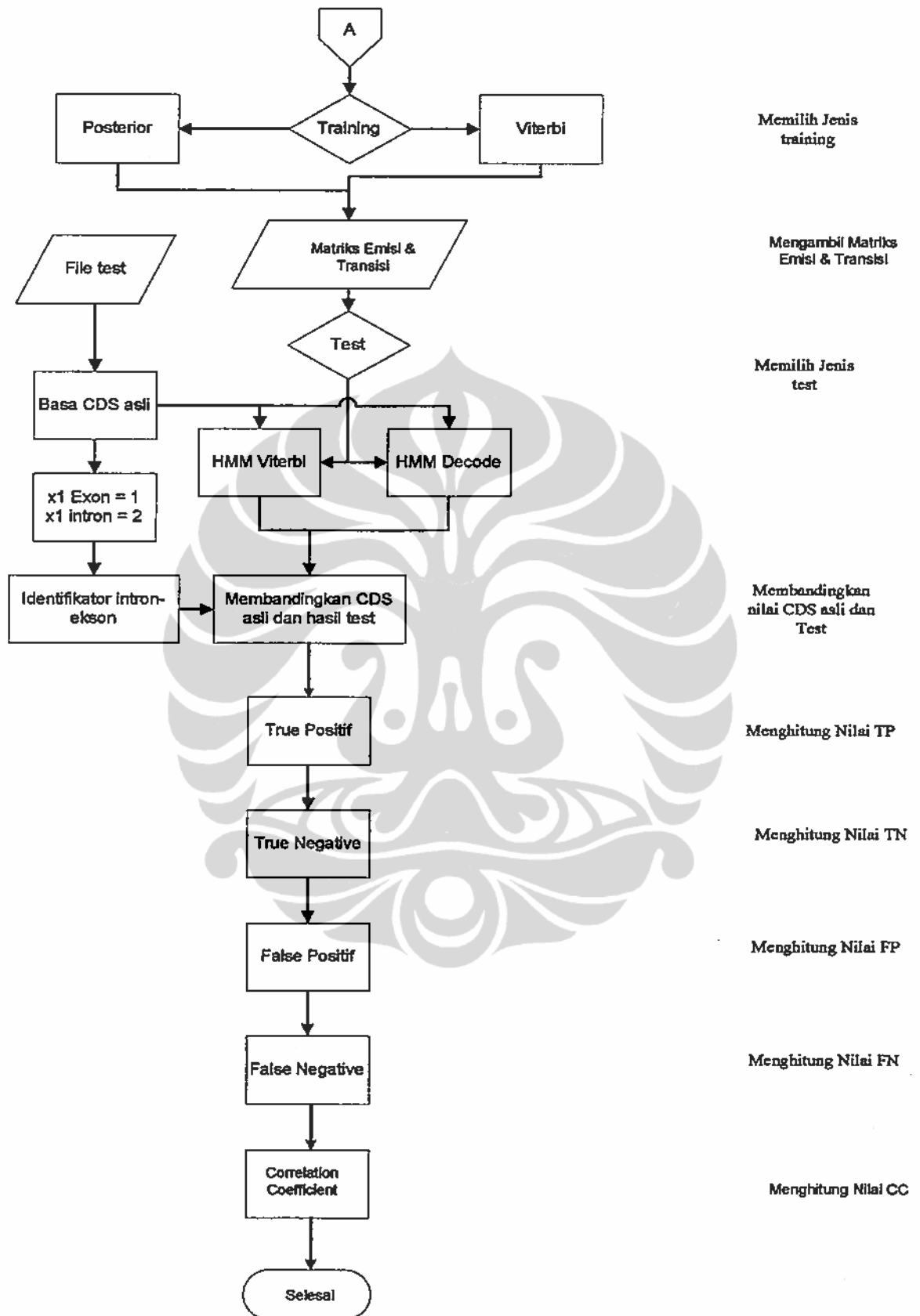
Perangkat lunak adalah bahasa pemrograman yang digunakan pada proses simulasi yaitu Matlab versi 7.0 yang memiliki *tool-box bioinformatics* sehingga dapat memproses sekuen DNA dalam bentuk format GenBank.

4.2.7 Diagram Alir Simulasi

Proses simulasi penelitian ini mempunyai diagram alir (*flowchart*) seperti Gambar 4.19.



Gambar 4.19 Diagram Alir Simulasi



Gambar 4.19 Diagram Alir Simulasi (Lanjutan)

BAB 5

PENGUJIAN DAN ANALISIS

Pengujian dilakukan dengan uji coba simulasi perancangan model yang menggunakan model *hidden* Markov untuk memprediksi ekson DNA *Plasmodium falciparum*. Hasil pengujian merupakan hasil simulasi dari beberapa uji coba sesuai dengan struktur model. Analisis hasil pengujian dilakukan terhadap hasil-hasil uji coba dan teori dasar yang digunakan.

5.1 Pengujian

Pengujian dilakukan terhadap sistem proses simulasi dengan menggunakan bahasa pemrograman Matlab versi 7.0 antara lain untuk pengujian data-data sekuen input, pembentukan *label* sekuen input dan pengujian emisi *state* model.

5.1.1 Pengujian Data Sekuen

Pada pengujian data sekuen mempunyai tujuan agar dapat diketahui karakteristik data dari sekuen untuk proses *training* HMM. Hasil pengujian ini antara lain seperti nama sekuen, urutan basa sekuen, urutan basa total ekson, urutan basa total intron, jumlah lokasi ekson dan intron. Hasil pengujian ini terdapat pada Lampiran 2 yang menampilkan contoh pemrograman untuk pengujian data sekuen dan hasil pengujian yang antara lain menampilkan nama sekuen, basa-basa ekson, basa-basa intron, basa-basa sekuen asli, banyaknya ekson dan intron sekuen tersebut.

5.1.2 Pengujian Label State Sekuen

Pengujian ini dilakukan setelah menentukan jumlah *state* pada model, sehingga nomor urut *label* sekuen sesuai dengan jumlah *state* model. Sebagai contoh untuk model dengan 5 *state* maka pada sekuen akan terdapat nomor urut *label state* dari 1 sampai

dengan 5 sepanjang CDS sekuen. Pemrograman untuk pengujian ini terdapat pada Gambar 5.1 dan hasil pengujian ini terdapat pada Gambar 5.2.

```

% Membentuk struktur label sekuen sesuai dengan jumlah statemodel(5)
% =====
m = test;
o = length(m);
for ff = 1:o;
    t = m(ff);
    Sgab = char(t);
    s = genbankread(Sgab);
    cds = s.CDS;
    pj = length(cds);
    SS = s.Sequence;
    SeQ{ff} = nt2int(SS);
    x1=[1,length(SS)];
    x1(1:3)=1;
    x1(4:cds(2))=2;
    for x = 1:2:pj-2;
        for y=cds(x+1)+1:1:cds(x+2)-1
            x1(y)=3;
        end
    for x = 2:2:pj-1;
        for y=cds(x+1):1:cds(x+2)
            x1(y)=4;
        end
    end
end
end
x1(length(SS)-2:length(SS))=5;
rx{ff}=x1;
rx{ff}=IntronExonIdentificator(x1,4,[2 1 2]);
%disp(rx{ff});
end

```

Gambar 5.1 Pemrograman *label state* sekuen

| | | | | | | |
|-------|------------|------------|------------|------------|------------|------------|
| Asli | atggcacatt | tatatgaaga | ggtttataaa | aagagaaaat | attttgatag | agctttatgt |
| State | 1112222222 | 2222222222 | 2222222222 | 2222222222 | 2222222222 | 2222222222 |
| Asli | aatgattatg | aagaatggca | agataaaata | agagaatcaa | aaacggctca | tataggaat |
| State | 2222222222 | 2222222222 | 2222222222 | 2222222222 | 2222222222 | 2222222222 |
| Asli | ttatccatat | atacaacaca | acaacaaata | tatgaggtaa | gcaaaaagga | aaatatgata |
| State | 2222222222 | 2222222222 | 2222222222 | 2222223333 | 3333333333 | 3333333333 |
| Asli | taatcatagg | ttgtaatata | tatatatata | tatattatat | atatatattt | atttatttat |
| State | 3333333333 | 3333333333 | 3333333333 | 3333333333 | 3333333333 | 3333333333 |
| Asli | ttatttattt | ttatttattt | ttatttttat | ttatcgaatg | tcttcacatt | atataccatt |
| State | 3333333333 | 3333333333 | 3333333333 | 3333333333 | 3333333333 | 3333333333 |

Gambar 5.2 Hasil pengujian *label state* sekuen

| | | | | | | |
|-------|------------|------------|------------|-------------|-------------|------------|
| Asli | taatttttcc | aagcatatgt | ctaaggctgg | ggatgtagaa | aacatcatca | tgggtttaca |
| State | 3333333333 | 3334444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 |
| Asli | cagaaccgag | aaatctcctt | gtggattttg | ttttgtagtt | tataaaaaga | aagaaggata |
| State | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 |
| Asli | tactcaagca | gtaaattttt | taaataattc | aatttttagat | ggtagaatta | taagagttga |
| State | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 |
| Asli | tgaagaccta | ggtattattg | gtaaacgaaa | atatggaaga | gggaaaacag | gagttcaaaa |
| State | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 |
| Asli | aagagatgaa | cgtaataaat | tttatgatga | agatagacca | aaagtttttag | ataacttagt |
| State | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 |
| Asli | tgatcataca | attaacaaaa | aaaggaaact | taataataat | cattattatg | aaagaaatat |
| State | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 |
| Asli | aaaacaaaaa | atactcttac | catcaaatta | tactcttaac | acaaccgata | tgaaaaatag |
| State | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 |
| Asli | tagaatggtc | tataaaaac | caccgtaac | tttatatccg | aacgctcaac | atatgatgca |
| State | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 |
| Asli | tataaataaa | aaaaaacatc | aaaatagaaa | tttatataaa | tttcacgtaa | aaaaagatag |
| State | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 | 4444444444 |
| Asli | acaaaataat | acgccaacat | aa | | | |
| State | 4444444444 | 4444444444 | 55 | | | |

Gambar 5.2 Hasil Pengujian *label state* sekuen (Lanjutan)

Dari Gambar 5.2 dapat dijelaskan bahwa model mempunyai *label state* 1 untuk kodon *start*, *label state* 2 untuk panjang lokasi ekson pertama tanpa kodon *start*, *label* 3 untuk awal lokasi intron yang biasa ditandai dengan basa GT sampai ketemu basa AG sebagai batas akhir lokasi intron, *label* 4 untuk lokasi ekson terakhir tanpa salah satu dari kodon *stop*. Jika sekuen memiliki jumlah ekson lebih dari dua maka *label state* akan berulang untuk *label state* 3 dan 4 saja karena model ini mempunyai struktur arah *backward* menuju awal lokasi intron kemudian menuju lokasi ekson sehingga membentuk seperti struktur CDS sekuen tersebut dimana ekson dan intron *alternately located*.

5.1.3 Pengujian Emisi State

Pengujian emisi *state* dilakukan sebagai kontrol terhadap distribusi basa DNA pada setiap *state* sesuai model. Sebagai contoh pemrograman untuk pengujian bentuk struktur model 5 *state* terdapat pada Gambar 5.3 dan hasilnya terdapat pada Gambar 5.4.

```
function[E, ihi]=MakeMTE(SeqGab, STGab, JSequence, m)
che=0;
cho=0;
for i=1:1:JSequence
    if che<max(SeqGab{i})
        che=max(SeqGab{i});
    end
    if cho<max(STGab{i})
        cho=max(STGab{i});
    end
end
Phi=zeros(cho, che);
for i=1:1:JSequence
    HL=length(SeqGab{i});
    for j=1:1:HL
        if (STGab{i}(j)==5)&&(SeqGab{i}(j)==2)
            disp('Check this Out');
            disp([STGab{i} ; SeqGab{i}])
            disp(sprintf('Ada di sequence %d %s', i, m{i}));
            pause
        end
        Phi(STGab{i}(j), SeqGab{i}(j))=Phi(STGab{i}(j), SeqGab{i}(j))+1;
    end
end
for i=1:1:cho
    GL=sum(Phi(i, :));
    Phi(i, :)=Phi(i, :)/GL;
end
E=Phi;
```

Gambar 5.3 Pemrograman emisi *state*

```
e = [ 0.3333    0    0.3333    0.3333;
      0.4342    0.1094    0.1507    0.3056;
      0.4210    0.0599    0.0703    0.4487;
      0.4215    0.1173    0.1622    0.2990;
      0.5592    0    0.1075    0.3333 ]
```

Gambar 5.4 Hasil pengujian emisi *state*

5.2 Hasil Pengujian Struktur Model dan Analisis

Beberapa struktur model yang terdapat pada perancangan model seperti yang diuraikan pada BAB 4 telah di ujicoba dan menghasilkan nilai-nilai CC seperti penjelasan di bawah ini.

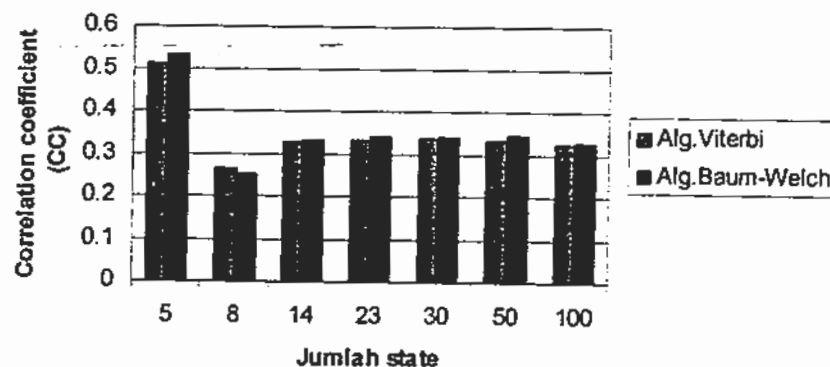
5.2.1 Pengujian Struktur Model

Ujicoba model ini menggunakan struktur model seperti pada Gambar 4.2 sampai dengan Gambar 4.8 pada BAB 4 dengan jumlah *state* 5, 8, 14, 23, 30, 50 dan 100. Simulasi ini mencoba komponen nilai random matriks transisi dan metode *forward-backward* HMM seperti yang terdapat pada struktur model tersebut. Hasil simulasi yang diperoleh dari ujicoba ini dengan nilai CC rata-rata terdapat pada Tabel 5.1. dan grafik nilai CC rata-rata seperti Gambar 5.5.

Tabel 5.1 Nilai CC rata-rata Uji Coba Struktur Model

| HMM Test | Jumlah State | | | | | | |
|------------------|--------------|--------|--------|--------|--------|--------|--------|
| | 5 | 8 | 14 | 23 | 30 | 50 | 100 |
| CC of Viterbi | 0.5094 | 0.2626 | 0.3270 | 0.3304 | 0.3356 | 0.3339 | 0.3220 |
| CC of Baum-Welch | 0.5336 | 0.2523 | 0.3298 | 0.3391 | 0.3416 | 0.3429 | 0.3275 |

Nilai CC rata-rata



Gambar 5.5 Grafik Nilai CC rata-rata Struktur Model

5.2.2 Analisis Pengujian Struktur Model.

Simulasi dengan bentuk struktur model ini dilakukan untuk jumlah *state* 5, 8, 14, 23, 30, 50 dan 100 dengan menghasilkan nilai CC rata-rata untuk jumlah *state* 8 sampai dengan 100 hasilnya lebih kecil dari model dengan jumlah *state* 5. Hal ini diakibatkan karena pengujian bentuk struktur model 5 *state* masih sesuai dengan struktur CDS tetapi untuk struktur model 8 *state* sampai dengan 100 *state* terdapat penambahan satu ekson sehingga struktur tidak lagi seperti CDS, terdapat jumlah ekson adalah dua lebih banyak dari jumlah intron. Transisi *state* dengan arah *forward-backward* yang dirancang pada model dilakukan dari beberapa *state* yang ada pada lokasi ekson menuju beberapa *state* yang terdapat pada intron. Sehingga mengakibatkan estimasi emisi *state* hasil *training* jauh berbeda dari emisi *state* awal (sebelum proses *training*).

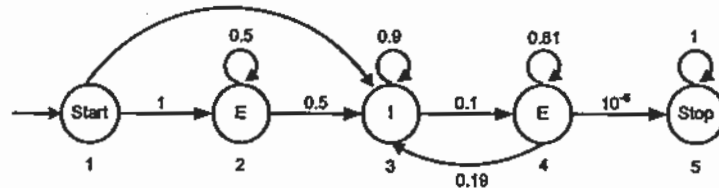
5.2.3 Pengujian Struktur Dasar Model

Simulasi untuk pengujian struktur model pada Bab 5.2.1 dengan hasil simulasi yang terdapat pada Tabel 5.1 diatas, menghasilkan nilai CC rata-rata semakin kecil dengan bertambahnya jumlah *state*. Oleh karena itu penelitian ini melakukan ujicoba kembali dengan struktur model dasar seperti pada struktur DNA *Plasmodium falciparum*. Pada model ini tidak ditambah dengan *state* ekson sehingga struktur model seperti pada CDS gen dan *forward-backward* hanya dilakukan dari akhir lokasi ekson menuju awal lokasi intron. Model dirancang untuk jumlah *state* 5, 7 dan 9 dengan penjelasan struktur model seperti dibawah ini.

1. Struktur Dasar Model 5 *state*

Kodon start dan kodon stop dari ekson pertama dan ekson terakhir dipisah dari rangkaian basa yang terdapat pada bagian ekson tersebut sehingga jumlah *state* menjadi

5. Proses *backward* dari *state* 4 menuju ke *state* 3 dengan nilai transisi *state* model dan struktur model dapat dijelaskan seperti Gambar 5.6.



Gambar 5.6 Struktur Dasar Model 5 *state*

Sedangkan nilai komponen emisi *state* untuk struktur model ini diperoleh dari distribusi basa DNA pada tiap *state* dan hasil dalam bentuk matriks emisi *state* model ini adalah sebagai berikut :

$$e = \begin{bmatrix} 0.3333 & 0 & 0.3333 & 0.3333 \\ 0.4342 & 0.1094 & 0.1507 & 0.3056 \\ 0.4210 & 0.0599 & 0.0703 & 0.4487 \\ 0.4215 & 0.1173 & 0.1622 & 0.2990 \\ 0.5592 & 0 & 0.1075 & 0.3333 \end{bmatrix}$$

Beberapa ujicoba dilakukan dengan menggunakan beberapa nilai acak dari transisi *state* yang menghasilkan nilai CC rata-rata seperti pada Tabel 5.2.

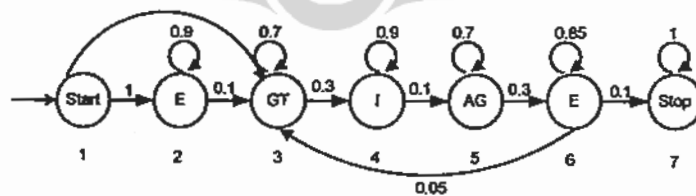
Selain hasil ujicoba yang terdapat pada Tabel 5.2, dilakukan pula ujicoba dengan beberapa nilai acak transisi *state* yang lain dengan hasil nilai CC rata-rata terdapat pada Lampiran 3.

Tabel 5.2 Nilai CC rata-rata Struktur Dasar Model 5 *state*

| Ujicoba | Transition Matrix | | | | | | | Corr.Coeff.(CC) | | Iterations |
|---------|-------------------|----------|----------|----------|----------|----------|----------|-----------------|---------|------------|
| | t_{11} | t_{22} | t_{33} | t_{44} | t_{55} | t_{13} | t_{43} | Viterbi | B-Welch | |
| 1 | 0 | 0.9 | 0.9 | 0.8 | 1 | 0.1 | 0.1 | 0.5688 | 0.5968 | 12 |
| 2 | 0 | 0.9 | 0.9 | 0.86 | 1 | 0.05 | 0.14 | 0.5708 | 0.5992 | 36 |
| 3 | 0 | 0.9 | 0.9 | 0.85 | 1 | 0.1 | 0.1 | 0.6091 | 0.6225 | 23 |
| 4 | 0 | 0.7 | 0.9 | 0.88 | 1 | 0.1 | 0.1 | 0.6988 | 0.6648 | 25 |
| 5 | 0 | 0.1 | 0.9 | 0.85 | 1 | 0.1 | 0.05 | 0.7018 | 0.6710 | 18 |
| 6 | 0 | 0.5 | 0.9 | 0.88 | 1 | 0 | 0.1 | 0.7042 | 0.6724 | 22 |
| 7 | 0 | 0.5 | 0.9 | 0.86 | 1 | 0.1 | 0.14 | 0.7044 | 0.6741 | 20 |
| 8 | 0 | 0.5 | 0.9 | 0.84 | 1 | 0 | 0.16 | 0.7045 | 0.6741 | 20 |
| 9 | 0 | 0.5 | 0.9 | 0.85 | 1 | 0 | 0.15 | 0.7054 | 0.6750 | 25 |
| 10 | 0 | 0.5 | 0.9 | 0.81 | 1 | 0 | 0.19 | 0.7061 | 0.6759 | 25 |
| 11 | 0 | 0.5 | 0.9 | 0.86 | 1 | 0.05 | 0.14 | 0.7062 | 0.6744 | 27 |

2. Struktur Dasar Model 7 *state*

Model ini merupakan pengembangan dari struktur model 5 *state* dengan memisahkan basa-basa GT dan basa-basa AG dari *state* intron menjadi *state* GT dan *state* AG. Sehingga pada bagian intron memiliki jumlah *state* sebanyak 3, sedangkan bagian ekson untuk model ini mempunyai total jumlah *state* adalah empat sehingga total jumlah *state* model ini menjadi 7 dan struktur model serta nilai komponen transisi *state* seperti pada Gambar 5.7.



Gambar 5.7 Struktur Dasar Model 7 *state*

Sedangkan nilai emisi *state* untuk struktur model ini diperoleh sebagai berikut :

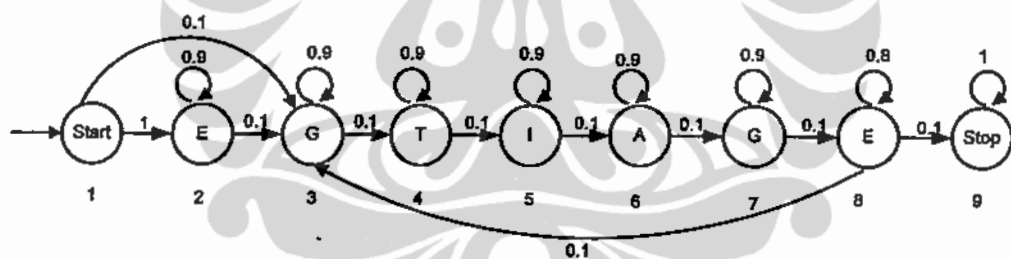
$$e = \begin{bmatrix} 0.3333 & 0 & 0.3333 & 0.3333 & ; \\ 0.4342 & 0.1094 & 0.1507 & 0.3056 & ; \\ 0 & 0 & 0.5000 & 0.5000 & ; \\ 0.4210 & 0.0599 & 0.0703 & 0.4487 & ; \\ 0.5000 & 0 & 0.5000 & 0 & ; \\ 0.4215 & 0.1173 & 0.1622 & 0.2990 & ; \\ 0.5592 & 0 & 0.1075 & 0.3333 &]$$

Hasil simulasi model dengan menggunakan nilai emisi *state* dan beberapa nilai acak transisi *state* menghasilkan nilai CC rata-rata seperti Tabel 5.3.

Selain hasil ujicoba yang terdapat pada Tabel 5.3, dilakukan pula ujicoba dengan beberapa nilai acak matriks transisi yang lain dengan hasil nilai CC rata-rata terdapat pada Lampiran 4.

3. Struktur Dasar Model 9 state

Seperti pada struktur model 7 *state* maka model ini juga dikembangkan dengan penambahan *state* masih pada bagian intron. *State* GT dipisah menjadi *state* G dan *state* T begitu pula pada *state* AG dipisah menjadi *state* A dan *state* G sehingga pada bagian intron terdapat 5 *state* dan total *state* model ini menjadi 9 *state*. Struktur model dan nilai komponen transisi *state* model dijelaskan seperti Gambar 5.8 dibawah ini.



Gambar 5.8 Struktur Dasar Model 9 *state*

Nilai emisi *state* untuk struktur model ini diperoleh sebagai berikut :

$$e = \begin{bmatrix} 0.3333 & 0 & 0.3333 & 0.3333 & ; \\ 0.4342 & 0.1094 & 0.1507 & 0.3056 & ; \\ 0 & 0 & 0.1000 & 0 & ; \\ 0 & 0 & 0 & 1.0000 & ; \\ 0.4210 & 0.0599 & 0.0703 & 0.4487 & ; \\ 1.0000 & 0 & 0 & 0 & ; \\ 0 & 0 & 1.0000 & 0 & ; \\ 0.4215 & 0.1173 & 0.1622 & 0.2990 & ; \\ 0.5592 & 0 & 0.1075 & 0.3333 &]$$

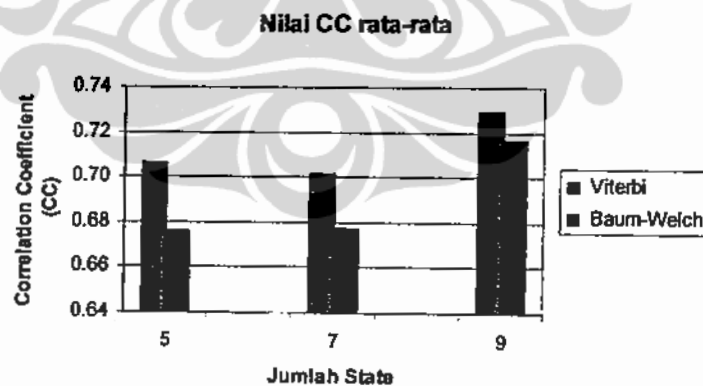
Hasil simulasi model dengan menggunakan nilai emisi *state* dan beberapa nilai acak transisi *state* menghasilkan nilai CC rata-rata seperti Tabel 5.4.

Selain hasil ujicoba yang terdapat pada Tabel 5.4, dilakukan pula ujicoba dengan beberapa nilai acak transisi *state* yang lain dan menghasilkan nilai CC rata-rata seperti terdapat pada Lampiran 5.

Hasil simulasi untuk ketiga model struktur dasar diatas menghasilkan nilai-nilai CC rata-rata terbaik untuk masing-masing model terdapat pada Tabel 5.5 dan grafik nilai CC rata-rata ketiga model tersebut terdapat pada Gambar 5.9.

Tabel 5.5 Nilai CC rata-rata Struktur Dasar Model

| HMM Test | Jumlah State | | |
|---------------|--------------|--------|--------|
| | 5 | 7 | 9 |
| CC of Viterbi | 0.7061 | 0.7016 | 0.7289 |
| CC of B-Welch | 0.6759 | 0.6774 | 0.7166 |



Gambar 5.9 Grafik Nilai CC rata-rata Struktur Dasar Model

Pada Gambar 5.9 di atas, nilai CC rata-rata untuk model 7 *state* dengan menggunakan algoritma Viterbi lebih rendah dibandingkan dengan hasil tes untuk model 5 *state* dan 9 *state*. Sedangkan hasil tes HMM dengan menggunakan algoritma Baum-

Welch untuk ketiga model ini menunjukkan nilai CC rata-rata lebih baik dengan bertambahnya jumlah *state*.

5.2.4 Analisis Pengujian Struktur Dasar Model

Pada simulasi ini struktur model dirancang dengan jumlah *state* 5, 7 dan 9 yang merupakan pengembangan struktur model tetapi tetap mengikuti struktur ekson dan intron yang terdapat pada CDS. Penambahan jumlah *state* dari CDS tersebut tetap mengikuti dogma yang ada pada struktur gen khususnya DNA *Plasmodium falciparum*. Sedangkan rantai Markov pada metode HMM dilakukan dari *state* akhir pada lokasi ekson menuju *state* awal lokasi intron. Dari hasil simulasi menunjukkan nilai CC rata-rata model lebih baik dari hasil simulasi pada uji coba struktur model walaupun untuk model 7 *state* nilai CC rata-rata tersebut lebih kecil dari model 5 *state* dan 9 *state*. Oleh karena itu pada uji coba selanjutnya penambahan jumlah *state* pada struktur model dirancang dengan menggunakan dasar-dasar pengembangan model seperti struktur dasar model 9 *state*.

Tabel 5.3 Nilai CC rata-rata Struktur Dasar Model 7 state

| Ujicoba | Transition Matrix | | | | | | | | | | | | | | | Corr. Coeff. (CC) | | Iterations |
|---------|-------------------|-----|-----|------|-----|------|-----|------|-----|--------|--------|-----|-----|-----|-----|-------------------|---------|------------|
| | t11 | t12 | t13 | t14 | t15 | t16 | t17 | t18 | t19 | t20 | t21 | t22 | t23 | t24 | t25 | Viterbi | B-Weich | |
| 1 | 0 | 0.6 | 0.2 | 0.65 | 0.2 | 0.7 | 0 | 0.1 | 1 | 0.3740 | 0.3608 | 22 | | | | | | |
| 2 | 0 | 0.4 | 0.2 | 0.6 | 0.2 | 0.7 | 0 | 0.1 | 1 | 0.4488 | 0.4522 | 36 | | | | | | |
| 3 | 0 | 0.6 | 0.2 | 0.6 | 0.2 | 0.7 | 0 | 0.1 | 1 | 0.4664 | 0.4824 | 32 | | | | | | |
| 4 | 0 | 0.4 | 0.2 | 0.65 | 0.2 | 0.7 | 0 | 0.1 | 1 | 0.5937 | 0.5943 | 66 | | | | | | |
| 5 | 0 | 0.8 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 1 | 0.6088 | 0.6233 | 20 | | | | | | |
| 6 | 0 | 0.3 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 1 | 0.6713 | 0.6643 | 21 | | | | | | |
| 7 | 0 | 0.5 | 0.1 | 0.9 | 0.1 | 0.8 | 0.1 | 0.1 | 1 | 0.6731 | 0.6643 | 18 | | | | | | |
| 8 | 0 | 0.9 | 0.5 | 0.9 | 0.5 | 0.85 | 0.1 | 0.05 | 1 | 0.6733 | 0.6737 | 16 | | | | | | |
| 9 | 0 | 0.9 | 0.3 | 0.9 | 0.3 | 0.85 | 0.1 | 0.05 | 1 | 0.6763 | 0.6758 | 25 | | | | | | |
| 10 | 0 | 0.9 | 0.1 | 0.9 | 0.1 | 0.85 | 0.1 | 0.05 | 1 | 0.6778 | 0.6743 | 21 | | | | | | |
| 11 | 0 | 0.9 | 0.1 | 0.9 | 0.1 | 0.75 | 0.1 | 0.15 | 1 | 0.6800 | 0.6730 | 18 | | | | | | |
| 12 | 0 | 0.9 | 0.5 | 0.9 | 0.5 | 0.75 | 0.1 | 0.15 | 1 | 0.6821 | 0.6828 | 19 | | | | | | |
| 13 | 0 | 0.9 | 0.4 | 0.9 | 0.4 | 0.8 | 0.1 | 0.1 | 1 | 0.6829 | 0.6853 | 20 | | | | | | |
| 14 | 0 | 0.9 | 0.6 | 0.9 | 0.6 | 0.75 | 0.1 | 0.15 | 1 | 0.6851 | 0.6936 | 20 | | | | | | |
| 15 | 0.1 | 0.1 | 0.2 | 0.8 | 0.2 | 0.8 | 0 | 0 | 1 | 0.7016 | 0.6774 | 26 | | | | | | |

Tabel 5.4 Nilai CC rata-rata Struktur Dasar Model 9 state

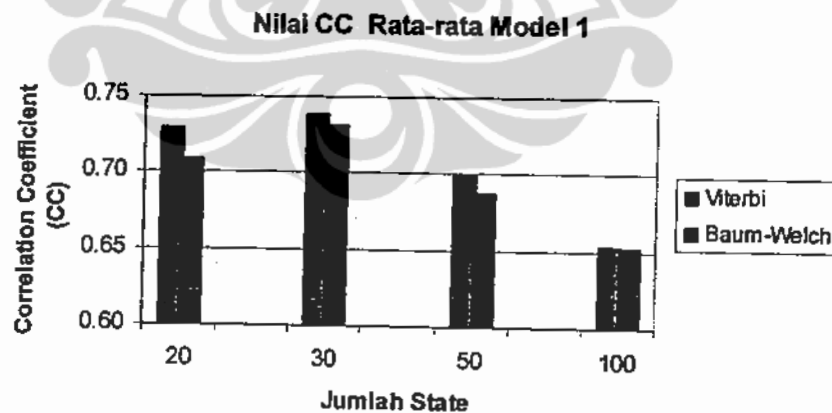
| Ujicoba | Transition Matrix | | | | | | | | | | | | | | | Corr. Coeff. (CC) | | Iterations |
|---------|-------------------|------|-----|-----|------|------|------|------|-----|-----|------|--------|--------|-----|-----|-------------------|---------|------------|
| | t11 | t12 | t13 | t14 | t15 | t16 | t17 | t18 | t19 | t20 | t21 | t22 | t23 | t24 | t25 | Viterbi | B-Weich | |
| 1 | 0 | 0.8 | 0.5 | 0.7 | 0.5 | 0.5 | 0.5 | 0.8 | 1 | 0.1 | 0.15 | 0.3557 | 0.3805 | 27 | | | | |
| 2 | 0 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.89 | 1 | 0 | 0.1 | 0.3872 | 0.3987 | 40 | | | | |
| 3 | 0 | 0.81 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 0.7 | 1 | 0 | 0.1 | 0.6744 | 0.6756 | 12 | | | | |
| 4 | 0 | 0.8 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 0.7 | 1 | 0 | 0.1 | 0.6787 | 0.6755 | 21 | | | | |
| 5 | 0 | 0.79 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 0.7 | 1 | 0 | 0.1 | 0.6791 | 0.6679 | 15 | | | | |
| 6 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.89 | 1 | 0 | 0.1 | 0.6887 | 0.6637 | 20 | | | | |
| 7 | 0 | 0.9 | 0.1 | 0.9 | 0.1 | 0.1 | 0.8 | 0.8 | 1 | 0.1 | 0.1 | 0.6958 | 0.6984 | 15 | | | | |
| 8 | 0 | 0.9 | 0.1 | 0.9 | 0.1 | 0.1 | 0.85 | 0.85 | 1 | 0.1 | 0.1 | 0.6993 | 0.7001 | 23 | | | | |
| 9 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.85 | 1 | 0 | 0.1 | 0.7055 | 0.6975 | 16 | | | | |
| 10 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.85 | 1 | 0.1 | 0.1 | 0.7193 | 0.7161 | 14 | | | | |
| 11 | 0 | 0.3 | 0.1 | 0.1 | 0.78 | 0.09 | 0.7 | 0.7 | 1 | 0 | 0.1 | 0.7242 | 0.7124 | 20 | | | | |
| 12 | 0 | 0.3 | 0.1 | 0.1 | 0.8 | 0.15 | 0.74 | 0.74 | 1 | 0 | 0.1 | 0.7258 | 0.7120 | 22 | | | | |
| 13 | 0 | 0.3 | 0.1 | 0.1 | 0.79 | 0.1 | 0.7 | 0.7 | 1 | 0.1 | 0.1 | 0.7242 | 0.7124 | 21 | | | | |
| 14 | 0 | 0.3 | 0.1 | 0.1 | 0.8 | 0.2 | 0.8 | 0.8 | 1 | 0.1 | 0.1 | 0.7273 | 0.7134 | 17 | | | | |
| 15 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.8 | 1 | 0.1 | 0.1 | 0.7289 | 0.7166 | 16 | | | | |

5.2.5 Pengujian dan Analisis Pengembangan Struktur Dasar Model

Hasil uji coba struktur model 9 *state* seperti pada Gambar 5.8 dan grafik pada Gambar 5.9 menghasilkan nilai CC rata-rata paling baik maka untuk pengembangan model-model berikut ini menggunakan dasar model seperti struktur model ini. Selanjutnya uji coba pengembangan struktur model menggunakan 4 macam model seperti telah dijelaskan pada BAB 4.1.4 dan jumlah *state* ditentukan secara acak mulai dari 20, 30, 50 dan 100.

A.1. Pengujian Struktur Model 1

Struktur model ini mempunyai bentuk struktur umum seperti Gambar 4.10 pada BAB 4.1.4 dan bagian intron selalu terdiri dari 5 *state* dan ekson terakhir 2 *state*, sehingga sisa jumlah *state* untuk bagian ekson pertama. Uji coba dilakukan dengan 4 macam jumlah *state* dan hasil simulasi tersebut terdapat pada Tabel 5.6 dengan nilai CC rata-rata, nilai CC minimum, nilai CC maksimum dan nilai CC median, sedangkan grafik nilai CC rata-rata model ini terdapat pada Gambar 5.10.



Gambar 5.10 Grafik Nilai CC rata-rata Model 1

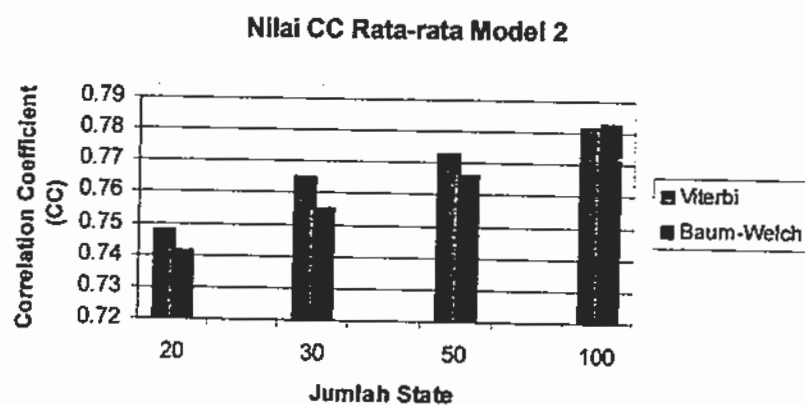
Dari hasil simulasi diperoleh pula nilai CC rata-rata bagi sekuen-sekuen dengan jumlah ekson tertentu seperti 2, 3 sampai 10 lokasi ekson sesuai dengan data sekuen pada masukan saat proses *training* seperti yang terdapat pada Tabel 5.7.

A.2. Analisis Pengujian Struktur Model 1

Ke empat jenis model ini hanya menambah *state* model pada bagian lokasi ekson pertama saja sedangkan untuk lokasi intron terdiri dari 5 *state* dan lokasi ekson terakhir 2 *state*. Dari hasil simulasi pada Gambar 5.6 nilai CC rata-rata model dengan jumlah *state* 30 lebih baik dari pada model-model lainnya. Hal ini menunjukkan jumlah basa pada lokasi ekson pertama dari masukan sekuen DNA lebih banyak yang berhasil di prediksi karena nilai-nilai transisi *state* dan emisi *state* masih dapat terjadi pada sebagian besar sekuen dibandingkan dengan struktur model dengan jumlah *state* 50 dan 100.

B.1. Pengujian Struktur Model 2

Struktur model ini mempunyai bentuk struktur umum seperti Gambar 4.11 pada BAB 4.1.4 dan hanya bagian ekson pertama dan intron yang dilakukan penambahan *state* sedangkan ekson terakhir tetap 2 *state*. Jumlah *state* pada bagian ekson dan intron tersebut bervariasi banyaknya disesuaikan dengan jumlah *state* pada model. Hasil simulasi untuk beberapa model dengan jumlah *state* tersebut terdapat pada Tabel 5.8 dengan nilai CC rata-rata, nilai CC minimum, nilai CC maksimum dan nilai CC median, sedangkan grafik nilai CC rata-rata untuk model ini terdapat pada Gambar 5.11.



Gambar 5.11 Grafik Nilai CC rata-rata Model 2

Dari hasil simulasi diperoleh pula nilai CC rata-rata bagi sekuen-sekuen dengan jumlah ekson tertentu seperti 2, 3 sampai 10 lokasi ekson sesuai dengan data sekuen pada masukan saat proses *training* seperti yang terdapat pada Tabel 5.9.

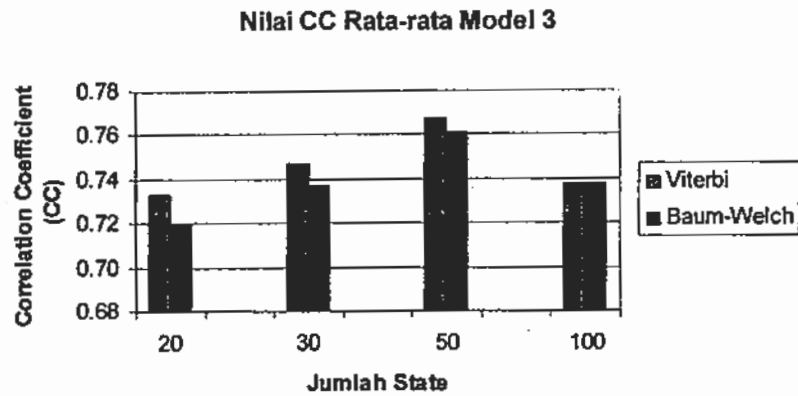
B.2. Analisis Pengujian Struktur Model 2

Penambahan jumlah *state* struktur model ini dilakukan pada bagian lokasi ekson dan intron kecuali bagian lokasi ekson terakhir. Pada model ini proses transisi lebih banyak terjadi disebabkan oleh sekuen yang memiliki jumlah ekson lebih dari dua yang memungkinkan terjadi arah *backward* ke lokasi intron. Selain itu penentuan besar nilai komponen matriks transisi sangat mempengaruhi hasil simulasi. Nilai CC rata-rata untuk semua struktur model ini semakin baik dengan bertambahnya jumlah *state* yang dijelaskan oleh grafik nilai CC rata-rata model pada Gambar 5.7.

C. 1. Pengujian Struktur Model 3

Struktur model ini mempunyai bentuk struktur umum seperti Gambar 4.12 pada BAB 4.1.4 dan semua bagian ekson dan intron dilakukan penambahan *state*. Jumlah *state* pada bagian ekson dan intron tersebut bervariasi banyaknya disesuaikan dengan jumlah *state* pada model. Hasil simulasi untuk beberapa model dengan jumlah *state* tersebut terdapat pada Tabel 5.10 dengan nilai CC rata-rata, nilai CC minimum, nilai CC maksimum dan nilai CC median, sedangkan grafik nilai CC rata-rata untuk model ini terdapat pada Gambar 5.12.

Dari hasil simulasi diperoleh pula nilai CC rata-rata bagi sekuen-sekuen dengan jumlah ekson tertentu seperti 2, 3 sampai 10 lokasi ekson sesuai dengan data sekuen pada masukan saat proses *training* seperti yang terdapat pada Tabel 5.11.



Gambar 5.12 Grafik Nilai CC rata-rata Model 3

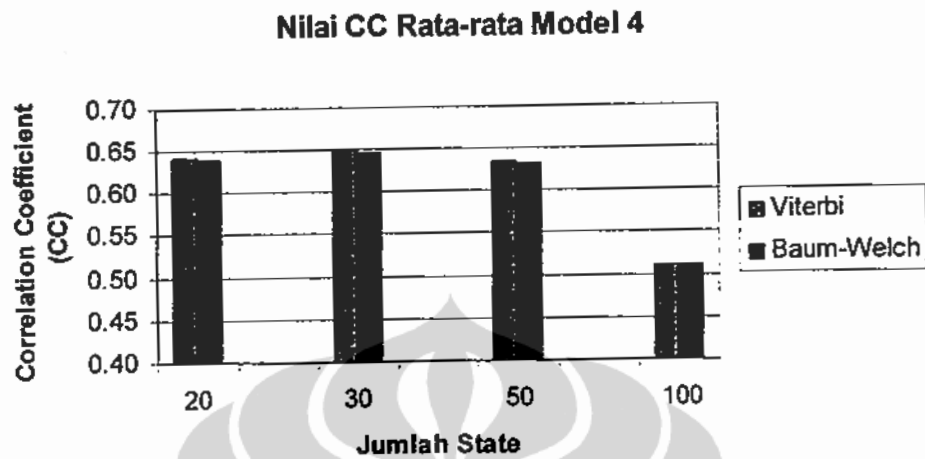
C. 2. Analisis Pengujian Struktur Model 3

Penambahan *state* dilakukan pada semua lokasi ekson dan intron yang terdapat pada model dan disesuaikan dengan rancangan model. Untuk menentukan berapa banyak jumlah *state* pada lokasi ekson maupun intron dilakukan secara acak sehingga mendapatkan nilai CC rata-rata model yang baik. Hasil simulasi pada Gambar 5.8 menunjukkan nilai CC rata-rata untuk model 50 *state* lebih baik dibandingkan dengan struktur model dengan jumlah *state* 20, 30 dan 100.

D.1. Pengujian Struktur Model 4

Struktur model ini mempunyai bentuk struktur umum seperti Gambar 4.13 pada BAB 4.1.4 dengan jumlah *state* pada model dirancang untuk 20, 30, 50 dan 100 *state*. Pada semua struktur model ini dilakukan penambahan *state* pada semua bagian ekson dan intron, pada lokasi intron pertama dari tiap sekuen dipisahkan dari bagian intron yang berulang dengan jumlah *state* pada bagian ekson dan intron tersebut bervariasi banyaknya disesuaikan dengan jumlah *state* pada model. Hasil simulasi untuk beberapa model dengan jumlah *state* tersebut terdapat pada Tabel 5.12 dengan nilai CC rata-rata,

nilai CC minimum, nilai CC maksimum dan nilai CC median, sedangkan grafik nilai CC rata-rata untuk model ini terdapat pada Gambar 5.13.



Gambar 5.13 Grafik Nilai CC rata-rata Model 4

Dari hasil simulasi diperoleh pula nilai CC rata-rata bagi sekuen-sekuen dengan jumlah ekson tertentu seperti 2, 3 sampai 10 lokasi ekson sesuai dengan data sekuen pada masukan saat proses *training* seperti yang terdapat pada Tabel 5.13.

D.2. Analisis Pengujian Struktur Model 4

Struktur model ini berbeda dengan ketiga struktur diatas dimana lokasi ekson pertama dan kedua serta intron pertama dari tiap sekuen dipisah dari bagian ekson dan intron lainnya. Sehingga untuk sekuen yang hanya memiliki 2 lokasi ekson maka dari posisi *state* terakhir pada ekson kedua harus dihubungkan dengan arah *forward* dengan bagian *state* stop. Sedangkan proses transisi banyak terjadi pada bagian yang berulang dari ekson menuju arah *backward* ke intron kedua pada tiap sekuen agar sesuai dengan rantai Markov. Hasil simulasi menunjukkan nilai CC rata-rata seperti yang terdapat pada Gambar 5.9 dan lebih rendah bila dibandingkan dengan Model 1, Model 2 dan Model 3.

5.2.6 Pengujian Waktu Proses Simulasi

Waktu proses simulasi model *hidden* Markov dibagi dua tahap yaitu waktu proses *training* dan waktu proses *testing* terhadap 152 sekuen DNA yang sama. Pada proses *training* hanya menggunakan algoritma Viterbi yang dilakukan uji coba terhadap model sedangkan untuk proses *testing* menggunakan algoritma Viterbi dan Baum-Welch. Dari hasil simulasi untuk ke empat macam bentuk struktur pengembangan model (Model 1 sampai dengan Model 4) dengan jumlah *state* 20, 30, 50 dan 100 membutuhkan waktu komputasi saat proses *training* dan *testing* terdapat pada Tabel 5.14.

5.2.7 Analisis Waktu Proses Simulasi

Waktu proses *training* dan waktu proses *testing* seperti pada Tabel 5.14 menunjukkan waktu rata-rata saat proses *training* dan *testing* tersebut. Untuk jumlah *state* 100 pada semua model membutuhkan waktu *training* yang jauh lebih lama bila dibandingkan dengan jumlah *state* yang lain, sedangkan untuk waktu *testing* dengan menggunakan algoritma Baum-Welch rata-rata membutuhkan waktu hampir dua kali lebih lama bila dibandingkan dengan menggunakan algoritma Viterbi.

Tabel 5.6 Nilai CC rata-rata Model 1

| Jumlah State | CC (Viterbi) | | | CC (Baum-Welch) | | | Jumlah Iterasi | |
|--------------|--------------|---------|----------|-----------------|-----------|---------|----------------|----------|
| | Rata-rata | Minimum | Maksimum | Median | Rata-rata | Minimum | | Maksimum |
| 20 | 0.7288 | -0.1366 | 1.0000 | 0.4317 | 0.7087 | -0.1351 | 0.9929 | 20 |
| 30 | 0.7379 | -0.1672 | 1.0000 | 0.4164 | 0.7309 | -0.1717 | 1.0000 | 26 |
| 50 | 0.6992 | -0.1238 | 0.9975 | 0.4368 | 0.6872 | -0.1238 | 0.9973 | 33 |
| 100 | 0.6537 | -0.2611 | 0.9782 | 0.3585 | 0.6518 | -0.2667 | 0.9801 | 83 |

Tabel 5.7 Nilai CC rata-rata Model 1 untuk jumlah ekson tertentu

| Jlh Ekson | 20 | | | 30 | | | 50 | | | 100 | | |
|-----------|----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|--|
| | CC | Viterbi | B-Welch | Viterbi | B-Welch | Viterbi | B-Welch | Viterbi | B-Welch | Viterbi | B-Welch | |
| 2 | | 0.7215 | 0.6998 | 0.7493 | 0.7416 | 0.7008 | 0.6947 | 0.6767 | 0.6574 | | | |
| 3 | | 0.7669 | 0.7451 | 0.7897 | 0.7741 | 0.7240 | 0.7059 | 0.7163 | 0.7212 | | | |
| 4 | | 0.7153 | 0.6796 | 0.6700 | 0.6512 | 0.6892 | 0.6404 | 0.5903 | 0.5863 | | | |
| 5 | | 0.7511 | 0.7332 | 0.7616 | 0.7526 | 0.7242 | 0.7116 | 0.6612 | 0.6639 | | | |
| 6 | | 0.6471 | 0.6288 | 0.5949 | 0.6300 | 0.5888 | 0.5774 | 0.5240 | 0.5203 | | | |
| 7 | | 0.7744 | 0.7687 | 0.7761 | 0.7529 | 0.7512 | 0.7426 | 0.6769 | 0.6946 | | | |
| 8 | | 0.5880 | 0.6173 | 0.5565 | 0.5502 | 0.5909 | 0.5935 | 0.5355 | 0.5234 | | | |
| 9 | | 0.8134 | 0.7850 | 0.7941 | 0.7831 | 0.7382 | 0.7408 | 0.6654 | 0.6611 | | | |
| 10 | | 0.8645 | 0.8724 | 0.7991 | 0.8345 | 0.8532 | 0.8386 | 0.7491 | 0.7717 | | | |

Tabel 5.8 Nilai CC rata-rata Model 2

| Jumlah State | CC (Viterbi) | | | CC (Baum-Welch) | | | Jumlah Iterasi | |
|--------------|--------------|---------|----------|-----------------|-----------|---------|----------------|----------|
| | Rata-rata | Minimum | Maksimum | Median | Rata-rata | Minimum | | Maksimum |
| 20 | 0.7481 | -0.0930 | 1.0000 | 0.4535 | 0.7414 | -0.1184 | 1.0000 | 21 |
| 30 | 0.7651 | -0.0848 | 1.0000 | 0.4576 | 0.7551 | -0.1108 | 1.0000 | 22 |
| 50 | 0.7727 | -0.0852 | 1.0000 | 0.4574 | 0.7661 | -0.0874 | 1.0000 | 19 |
| 100 | 0.7820 | -0.0789 | 1.0000 | 0.4605 | 0.7827 | -0.0789 | 1.0000 | 27 |

Tabel 5.9 Nilai CC rata-rata Model 2 untuk jumlah ekson tertentu

| Jlh. Ekson | 20 | | | 30 | | | 50 | | | 100 | | |
|------------|------------|----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | Jlh. State | CC | Viterbi | B-Weich | Viterbi | B-Weich | Viterbi | B-Weich | Viterbi | B-Weich | Viterbi | B-Weich |
| 2 | | | 0.7504 | 0.7439 | 0.7764 | 0.7709 | 0.7850 | 0.7804 | 0.7727 | 0.7725 | 0.7864 | 0.7842 |
| 3 | | | 0.7882 | 0.7923 | 0.7979 | 0.7862 | 0.7925 | 0.7944 | 0.7922 | 0.7864 | 0.7842 | 0.8254 |
| 4 | | | 0.6705 | 0.6511 | 0.6936 | 0.6862 | 0.7244 | 0.7111 | 0.7742 | 0.8259 | 0.7560 | 0.7649 |
| 5 | | | 0.7896 | 0.7665 | 0.7841 | 0.7841 | 0.8236 | 0.7999 | 0.8259 | 0.8254 | 0.7560 | 0.7649 |
| 6 | | | 0.6474 | 0.6436 | 0.6453 | 0.6357 | 0.6454 | 0.6353 | 0.7592 | 0.7560 | 0.7649 | 0.7336 |
| 7 | | | 0.8286 | 0.8051 | 0.8367 | 0.8288 | 0.8613 | 0.8459 | 0.7659 | 0.7336 | 0.8613 | 0.6573 |
| 8 | | | 0.6301 | 0.6372 | 0.6495 | 0.6472 | 0.5351 | 0.5398 | 0.7202 | 0.8613 | 0.6573 | 0.6573 |
| 9 | | | 0.7785 | 0.7650 | 0.8137 | 0.8160 | 0.8138 | 0.8076 | 0.8560 | 0.8613 | 0.6573 | 0.6573 |
| 10 | | | 0.8721 | 0.8795 | 0.8370 | 0.8471 | 0.8672 | 0.8580 | 0.5828 | 0.6573 | 0.6573 | 0.6573 |

Tabel 5.10 Nilai CC rata-rata Model 3

| Jumlah State | CC (Viterbi) | | | CC (Baun-Weich) | | | Jumlah Iterasi |
|--------------|--------------|---------|----------|-----------------|-----------|---------|----------------|
| | Rata-rata | Minimum | Maksimum | Median | Rata-rata | Minimum | |
| 20 | 0.7330 | -0.0863 | 1.0000 | 0.4568 | 0.7195 | -0.0925 | 1.0000 |
| 30 | 0.7469 | -0.0878 | 1.0000 | 0.4561 | 0.7367 | -0.1146 | 1.0000 |
| 50 | 0.7671 | -0.1433 | 1.0000 | 0.4283 | 0.7611 | -0.0885 | 1.0000 |
| 100 | 0.7378 | -0.037 | 1.0000 | 0.4815 | 0.7375 | -0.0397 | 1.0000 |

Tabel 5.11 Nilai CC rata-rata Model 3 untuk jumlah ekson tertentu

| Jlh. Ekson | 20 | | | 30 | | | 50 | | | 100 | | |
|------------|------------|----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | Jlh. State | CC | Viterbi | B-Weich | Viterbi | B-Weich | Viterbi | B-Weich | Viterbi | B-Weich | Viterbi | B-Weich |
| 2 | | | 0.7601 | 0.7321 | 0.7561 | 0.7499 | 0.7667 | 0.7684 | 0.7409 | 0.7300 | 0.7254 | 0.7047 |
| 3 | | | 0.7617 | 0.7621 | 0.7785 | 0.7686 | 0.7873 | 0.7700 | 0.7268 | 0.7254 | 0.7047 | 0.7676 |
| 4 | | | 0.6401 | 0.6155 | 0.6952 | 0.6870 | 0.7341 | 0.7143 | 0.7002 | 0.6661 | 0.7321 | 0.6800 |
| 5 | | | 0.7642 | 0.7383 | 0.7740 | 0.7624 | 0.8363 | 0.8191 | 0.7789 | 0.7676 | 0.6661 | 0.6800 |
| 6 | | | 0.6391 | 0.6401 | 0.6374 | 0.6250 | 0.6670 | 0.6665 | 0.6618 | 0.6661 | 0.7321 | 0.6800 |
| 7 | | | 0.6894 | 0.7410 | 0.8084 | 0.7804 | 0.7943 | 0.7764 | 0.7364 | 0.6596 | 0.6800 | 0.7850 |
| 8 | | | 0.5446 | 0.5839 | 0.5588 | 0.5345 | 0.6687 | 0.6475 | 0.6596 | 0.6800 | 0.7850 | 0.6217 |
| 9 | | | 0.7053 | 0.7254 | 0.7770 | 0.7773 | 0.7965 | 0.8082 | 0.7629 | 0.7850 | 0.6217 | 0.6217 |
| 10 | | | 0.9062 | 0.8947 | 0.8928 | 0.8911 | 0.8087 | 0.8160 | 0.5416 | 0.6217 | 0.6217 | 0.6217 |

Tabel 5.12 Nilai CC rata-rata Model 4

| Jumlah State | CC (Viterbi) | | | | CC (Baum-Welch) | | | | Jumlah Iterasi |
|--------------|--------------|---------|----------|--------|-----------------|---------|----------|--------|----------------|
| | Rata-rata | Minimum | Maksimum | Median | Rata-rata | Minimum | Maksimum | Median | |
| 20 | 0.6406 | -0.0188 | 1.0000 | 0.4060 | 0.6390 | -0.2088 | 1.0000 | 0.3956 | 21 |
| 30 | 0.6504 | -0.2219 | 1.0000 | 0.3891 | 0.6456 | -0.3149 | 1.0000 | 0.3426 | 44 |
| 50 | 0.6341 | -0.2371 | 1.0000 | 0.3815 | 0.6326 | -0.2367 | 1.0000 | 0.3817 | 41 |
| 100 | 0.5109 | -0.3779 | 0.9503 | 0.2862 | 0.5101 | -0.4379 | 0.9503 | 0.2562 | 43 |

Tabel 5.13 Nilai CC rata-rata Model 4 untuk jumlah ekson tertentu

| Jlh Ekson | 20 | | | | 30 | | | | 50 | | | | 100 | | | |
|-----------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|
| | Viterbi | | B-Welch | | Viterbi | | B-Welch | | Viterbi | | B-Welch | | Viterbi | | B-Welch | |
| 2 | 0.6343 | 0.6343 | 0.6623 | 0.6361 | 0.6447 | 0.6361 | 0.6308 | 0.6223 | 0.4797 | 0.4775 | 0.5682 | 0.5682 | 0.4264 | 0.4264 | 0.5529 | 0.5529 |
| 3 | 0.6721 | 0.6721 | 0.5689 | 0.5391 | 0.6786 | 0.6814 | 0.6583 | 0.6552 | 0.5728 | 0.5728 | 0.4282 | 0.4282 | 0.5507 | 0.5507 | 0.5211 | 0.5211 |
| 4 | 0.5729 | 0.5729 | 0.6937 | 0.6532 | 0.5504 | 0.5391 | 0.5262 | 0.5267 | 0.4282 | 0.4282 | 0.5507 | 0.5507 | 0.5211 | 0.5211 | 0.5460 | 0.5460 |
| 5 | 0.6927 | 0.6927 | 0.5523 | 0.6829 | 0.6532 | 0.6537 | 0.6529 | 0.6368 | 0.5507 | 0.5507 | 0.5211 | 0.5211 | 0.5460 | 0.5460 | 0.4564 | 0.4564 |
| 6 | 0.5502 | 0.5502 | 0.6900 | 0.6724 | 0.6829 | 0.6724 | 0.5785 | 0.5810 | 0.5211 | 0.5211 | 0.5460 | 0.5460 | 0.6176 | 0.6176 | 0.7129 | 0.7129 |
| 7 | 0.6841 | 0.6841 | 0.5612 | 0.7219 | 0.7116 | 0.7219 | 0.7189 | 0.6946 | 0.5364 | 0.5364 | 0.4534 | 0.4534 | 0.6176 | 0.6176 | 0.7129 | 0.7129 |
| 8 | 0.5724 | 0.5724 | 0.6741 | 0.6499 | 0.5800 | 0.5946 | 0.5157 | 0.5337 | 0.4534 | 0.4534 | 0.6176 | 0.6176 | 0.7129 | 0.7129 | 0.7129 | 0.7129 |
| 9 | 0.6678 | 0.6678 | 0.8691 | 0.8265 | 0.6504 | 0.6499 | 0.7035 | 0.6955 | 0.6165 | 0.6165 | 0.6242 | 0.6242 | 0.7129 | 0.7129 | 0.7129 | 0.7129 |
| 10 | 0.8932 | 0.8932 | 0.8691 | 0.8265 | 0.8517 | 0.8265 | 0.9225 | 0.9225 | 0.6242 | 0.6242 | 0.7129 | 0.7129 | 0.7129 | 0.7129 | 0.7129 | 0.7129 |

Tabel 5.14 Waktu Proses training dan testing Model 1 ~ Model 4

| Jumlah State | Training (Viterbi) [jam] | | | | Tes (Viterbi) [jam] | | | | Tes (Baum-Welch) [jam] | | | |
|--------------|--------------------------|-------|-------|-------|---------------------|-------|-------|-------|------------------------|-------|-------|-------|
| | Mod.1 | Mod.2 | Mod.3 | Mod.4 | Mod.1 | Mod.2 | Mod.3 | Mod.4 | Mod.1 | Mod.2 | Mod.3 | Mod.4 |
| 20 | 0.33 | 0.43 | 0.63 | 0.42 | 0.02 | 0.02 | 0.02 | 0.02 | 0.03 | 0.03 | 0.03 | 0.03 |
| 30 | 0.74 | 0.81 | 0.93 | 1.46 | 0.03 | 0.03 | 0.03 | 0.03 | 0.07 | 0.07 | 0.07 | 0.07 |
| 50 | 2.11 | 2.00 | 3.56 | 4.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.18 | 0.18 | 0.18 | 0.18 |
| 100 | 27.01 | 16.50 | 14.00 | 15.77 | 0.36 | 0.36 | 0.36 | 0.36 | 0.66 | 0.66 | 0.66 | 0.66 |

BAB 6

KESIMPULAN

6.1 Kesimpulan

Dari hasil penelitian diperoleh bahwa struktur HMM berbasis struktur ekson pada CDS dapat dipergunakan untuk memprediksi ekson DNA *Plasmodium falciparum*. Beberapa hal yang dapat disimpulkan dari hasil penelitian ini adalah sebagai berikut:

- Pada struktur model, penambahan jumlah *state* menghasilkan nilai CC rata-rata semakin kecil. Oleh karena itu dilakukan perubahan struktur model agar nilai CC rata-rata dapat lebih baik.
- Pada struktur dasar model yang sesuai dengan struktur ekson dan intron pada CDS adalah model yang mempunyai jumlah *state* 5, 7 dan 9 sehingga menghasilkan nilai CC rata-rata lebih baik dibandingkan dengan struktur model.
- Struktur Model 2 menghasilkan nilai CC rata-rata lebih baik dari tiga model lain yang ada pada pengembangan struktur dasar model.
- Nilai transisi *state* yang ditentukan secara acak sangat mempengaruhi kinerja model yang dihasilkan sedangkan untuk nilai emisi *state* merupakan distribusi basa DNA tiap *state* sesuai dengan struktur model.
- Proses *testing* dengan algoritma Baum-Welch rata-rata membutuhkan waktu hampir dua kali lebih lama dibandingkan dengan algoritma Viterbi karena algoritma Baum-Welch menghasilkan banyak jalur sedangkan algoritma Viterbi hanya satu jalur probabilitas sekuen untuk mendapatkan nilai maksimum.

6.2 Penelitian Lanjut

Struktur model yang telah di uji coba pada penelitian ini dapat di lanjutkan dengan melakukan uji coba pada algoritma atau struktur model sistem HMM yang berbeda seperti berikut ini:

- Struktur model yang telah di uji coba pada penelitian ini hendaknya dapat dilanjutkan dengan beberapa bagian dari sistem VEIL HMM tetapi tetap mengikuti pengembangan model yang dilakukan yaitu mengikuti struktur ekson pada CDS.
- Perangkat keras yang digunakan untuk simulasi model dapat dirancang khusus hanya untuk memprediksi ekson dengan adanya *Embedded PC*.
- Melakukan uji coba transformasi dari *domain* waktu ke *domain* frekuensi terhadap hasil-hasil yang telah diperoleh dari rancangan model untuk prediksi ekson DNA *Plasmodium falciparum*.

REFERENSI

- [1]. Steven M. Carr, *Gel image from an automated DNA sequencer*, Images & text material, 2005. http://www.mun.ca/biology/scarr/377_gel_file.htm, 25 Februari 2008.
- [2]. Xiaohua (Edward) Li, *Channel-Independent Viterbi Algorithm (CIVA) for DNA Sequencing*, Department of Electrical and Computer Engineering, State University of New York at Binghamton, 2003.
<http://www.ucesp.ws.binghamton.edu/~xli/old/papers/ciss2003.ppt>, Desember 2007.
- [3]. C.Q. Chang, *Genomic Signal Processing*, Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, P.R. China, 2006.
www.hku.hk/facmed/images/document/04research/database/presentations/cq_chang.ppt, 8 Desember 2007.
- [4]. Dimitris Anastassiou, *Genomic signal processing*, IEEE Signal Processing Magazine, Vol. 18, No.4, pp 8-20, Juli 2001.
- [5]. P. D. Cristea, *Conversion of nucleotides sequences into genomic signals*, Bio-Medical Engineering Center, "Politehnica" University of Bucharest, Romania, 29 April, 2002.
- [6]. Suhartati Agoes, Dadang Gunawan, Sar Sardy, *Proses Kode Genetik Pada Bidang Kompleks Menggunakan Filter Digital Finite Impulse Response*, Proceeding of The 6th Quality in Research, Fakultas Teknik Universitas Indonesia, Oktober 2003, ISSN: 1411-1284.
- [7]. Lawrence R. Rabiner, *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proceedings of The IEEE, Vol 77, No 2, pp 257-286, Februari 1989.
- [8]. Tetsushi Yada, Masato Ishikawa, Hidetoshi Tanaka and Kiyoshi Asai, *Extraction of Hidden Markov Model Representations of Signal Patterns in DNA Sequences*. The Japan Information Center of Science and Technology, 2005.
<http://www.psb.stanford.edu/psb-online/proceedings/psb96/yada.pdf>, 27 November 2007.

- [9]. Tetsushi Yada, Makoto Hirosawa, *Gene recognition in cyanobacterium genomic sequence data using the hidden Markov model*, Proceeding International Conference Intell. Syst. Mol. Biol, Vol 4, pp 252-260, 1996.
- [10]. Tetsushi Yada, Makoto Hirosawa, *Detection of short protein coding regions within the cyanobacterium genome: application of the hidden Markov model*, DNA. Res. Vol 31, Issue 6, pp 355-361, 31 Desember, 1996.
- [11]. John Henderson, Steven Salzberg, Kenneth H Fasman, *Finding gene in DNA with a Hidden Markov Model*, Journal Computational Biology, Vol. 4, Issue 2, pp 127-141, 1997.
- [12]. Daniel Nicorici, Jaakko Astola, Ioan Tobus , *Computational identification of exons in DNA with a Hidden Markov Model*, Tampere International Center for Signal Processing, Tampere University of Tecnology, 2002. <http://www.gensips.gatech.edu/proceedings/Contributed/CP2-06.pdf>, Juli 2003.
- [13]. Michael Brudno, *Gene Finding*. Scribed by Hui Lan, 17 Februari, 2006. <http://www.cs.utoronto.ca/~brudno/csc2427/Lec12Notes.pdf>
- [14]. Nagiza F Samatova, *Computational gene finding using HMMs*, Computational Biology Institute Oak Ridge National Laboratory, 2003. www.cs.utk.edu/~langston/courses/cs594-fall2003/GeneFinding-Nov12-2003.notes.ppt, Juli 2005.
- [15]. Suhartati Agoes, Dadang Gunawan, Sar Sardy, *Correlation Coefficient Analysis of Exon Region in DNA of Genes Plasmodium falciparum Using Hidden Markov Model*, Proceeding of BME DAYS 2006, Institut Teknologi Bandung, November 2006.
- [16]. Suhartati Agoes, Dadang Gunawan, Sar Sardy, *Exon Prediction on DNA-Genes Plasmodium falciparum Based on Coding Sequence Structure by Using Hidden Markov Model*, Jurnal Universa Medicina, Jurnal Kedokteran Trisakti, Vol. 26 No 4, (Oktober-Desember 2007), No. ISSN: 1411- 7169, Akreditasi Dirjen Dikti No. SK39/ Dikti/Kep /2004.
- [17]. Suhartati Agoes, Dadang Gunawan, Sar Sardy, *Exon Prediction on DNA-Gen Plasmodium falciparum by Using Hidden Markov Model*, Proceeding of 4th International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2007, Orlando, Florida, USA, Juli 2007

- [18]. Suhartati Agoes, Dadang Gunawan, Sar Sardy, *The Implementation of Hidden Markov Model for Exon Prediction on DNA-gene Plasmodium falciparum*, Jurnal Kedokteran Yarsi, No. ISSN: 0854 - 1159, November 2007 (Sedang dalam proses).
- [19]. Luke Alphey, *DNA Sequencing. From Experimental Methods to Bioinformatics*, Bios Scientific Publishers Limited, 1997.
- [20]. Amit Bahl, Brian Brunk, Ross L. Coppel, Jonathan Crabtree, Sharon J. Diskin, Martn J. Fraunholz *et al*, PlasmoDB: *the Plasmodium Genome Resource*. Nucleic Acids Res. 30(1): 87-90, 1 Januari 2002.
- [21]. Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, James D Watson, *Biologi molekuler sel: Mengenal sel*, Edisi kedua, PT Gramedia Pustaka Utama, pp 145-165, 1994.
- [22]. Jan Koolman, Klaus-Heinrich Rohm, *Atlas Berwarna & Teks : Biokimia*. Cetakan 1. Hipokrates. 2001.
- [23]. Image of *Plasmodium falciparum*.
http://www.sanger.ac.uk/Info/Press/gfx/nature_mos.gif, 12 Februari 2008.
- [23]. Thomas E. Wellems, Su Xin-zhuan, Michael Timothy Ferdig & David A Fidock, *Genome projects, genetic analysis, and the changing landscape of malaria research*. Current Opinions in Microbiology 2, 415-9 (1999).
- [25]. Malcolm J. Gardner, *The genome of the malaria parasite*. Current Opinion in Genetics and Development 9,704-708 (1999).
- [26]. Dominique Soldati, *The apicoplast as a potential therapeutic target in Toxoplasma and other apicomplexan parasites*. Parasitol Today 15, 5-7 (1999).
- [27]. Malcolm J. Gardner *et al*, *Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum*. Science 282, 1126-1132 (1998).
- [28]. Tapas Kanungo, *Hidden Markov Model*, Center for Automation Research, University of Maryland, 2003. <http://www.kanungo.com/software/hmmtut.pdf>, November 2005.
- [29]. Tulyakov, *Introduction to Hidden Markov models*, 2003.
- [30]. Andrew W. Moore, *Hidden Markov Models*. School of Computer Science, Carnegie Mellon University. 15 November, 2005.
- [31]. Jose Renau, *Hidden Markov Models: Fundamentals and applications to bioinformatics*. <http://web.mst.edu/~shoukat/IPDPS2007-BoA.pdf>, Juli 2005

LAMPIRAN 1
Data Sekuen DNA Plasmodium Falciparum

| No | Nama Sekuen | Panjang(bp) | | Jumlah | No | Nama Sekuen | Panjang(bp) | | Jumlah |
|----|--------------|-------------|--------|--------|----|--------------|-------------|--------|--------|
| | | Ekson | Intron | | | | Ekson | Intron | |
| 1 | NC_000521.1 | 6630 | 888 | 2 | 31 | NC_000521.31 | 2850 | 351 | 3 |
| 2 | NC_000521.2 | 966 | 246 | 2 | 32 | NC_000521.32 | 486 | 630 | 3 |
| 3 | NC_000521.3 | 681 | 224 | 2 | 33 | NC_000521.33 | 1074 | 377 | 4 |
| 4 | NC_000521.4 | 771 | 224 | 3 | 34 | NC_000521.34 | 930 | 357 | 4 |
| 5 | NC_000521.5 | 4251 | 1027 | 9 | 35 | NC_000521.35 | 471 | 312 | 3 |
| 6 | NC_000521.6 | 4254 | 989 | 9 | 36 | NC_000521.36 | 9402 | 693 | 4 |
| 7 | NC_000521.7 | 3516 | 1267 | 5 | 37 | NC_000521.37 | 666 | 731 | 6 |
| 8 | NC_000521.8 | 1851 | 857 | 8 | 38 | NC_000521.38 | 1302 | 224 | 2 |
| 9 | NC_000521.9 | 4233 | 58 | 2 | 39 | NC_000521.39 | 477 | 606 | 3 |
| 10 | NC_000521.10 | 1050 | 556 | 5 | 40 | NC_000521.40 | 987 | 365 | 5 |
| 11 | NC_000521.11 | 438 | 716 | 3 | 41 | NC_000521.41 | 2208 | 305 | 2 |
| 12 | NC_000521.12 | 1479 | 153 | 2 | 42 | NC_000521.42 | 1050 | 98 | 2 |
| 13 | NC_000521.13 | 1728 | 777 | 6 | 43 | NC_000521.43 | 954 | 146 | 2 |
| 14 | NC_000521.14 | 1833 | 675 | 2 | 44 | NC_000521.44 | 906 | 105 | 2 |
| 15 | NC_000521.15 | 843 | 831 | 7 | 45 | NC_000521.45 | 1116 | 243 | 2 |
| 16 | NC_000521.16 | 1689 | 613 | 5 | 46 | NC_000521.45 | 1065 | 215 | 2 |
| 17 | NC_000521.17 | 438 | 511 | 2 | 47 | NC_000910.1 | 2910 | 670 | 4 |
| 18 | NC_000521.18 | 4104 | 195 | 2 | 48 | NC_000910.3 | 414 | 379 | 3 |
| 19 | NC_000521.19 | 2109 | 643 | 5 | 49 | NC_000910.4 | 723 | 156 | 2 |
| 20 | NC_000521.20 | 1830 | 362 | 3 | 50 | NC_000910.5 | 987 | 645 | 3 |
| 21 | NC_000521.21 | 381 | 619 | 3 | 51 | NC_000910.6 | 1032 | 965 | 4 |
| 22 | NC_000521.22 | 822 | 410 | 4 | 52 | NC_000910.7 | 4095 | 738 | 3 |
| 23 | NC_000521.23 | 2424 | 644 | 4 | 53 | NC_000910.8 | 1575 | 650 | 5 |
| 24 | NC_000521.24 | 1908 | 713 | 5 | 54 | NC_000910.9 | 4677 | 158 | 2 |
| 25 | NC_000521.25 | 2721 | 613 | 3 | 55 | NC_000910.10 | 4323 | 951 | 9 |
| 26 | NC_000521.26 | 3846 | 178 | 2 | 56 | NC_000910.11 | 912 | 238 | 2 |
| 27 | NC_000521.27 | 1458 | 419 | 3 | 57 | NC_000910.12 | 951 | 135 | 2 |
| 28 | NC_000521.28 | 2811 | 213 | 2 | 58 | NC_000910.13 | 1107 | 244 | 2 |
| 29 | NC_000521.29 | 1143 | 481 | 3 | 59 | NC_000910.14 | 1047 | 148 | 2 |
| 30 | NC_000521.30 | 666 | 513 | 2 | 60 | NC_000910.15 | 894 | 101 | 2 |

| No | Nama Sekuen | Panjang Sekuen(bp) | Panjang(bp) | | Jumlah |
|----|--------------|--------------------|-------------|--------|--------|
| | | | Ekson | Intron | |
| 61 | NC_000910.16 | 1430 | 960 | 470 | 2 1 |
| 62 | NC_000910.17 | 1184 | 1038 | 146 | 2 1 |
| 63 | NC_000910.18 | 1705 | 1296 | 409 | 2 1 |
| 64 | NC_000910.19 | 1214 | 984 | 230 | 2 1 |
| 65 | NC_004318.1 | 1304 | 1149 | 155 | 2 1 |
| 66 | NC_004318.2 | 2912 | 1740 | 1172 | 9 8 |
| 67 | NC_004318.3 | 9726 | 9411 | 315 | 2 1 |
| 68 | NC_004318.4 | 1476 | 819 | 657 | 6 5 |
| 69 | NC_004318.5 | 1732 | 819 | 913 | 7 6 |
| 70 | NC_004318.6 | 1279 | 1041 | 238 | 2 1 |
| 71 | NC_004318.7 | 4159 | 4020 | 139 | 2 1 |
| 72 | NC_004318.8 | 862 | 705 | 157 | 2 1 |
| 73 | NC_004318.9 | 1262 | 579 | 683 | 3 2 |
| 74 | NC_004318.10 | 1137 | 843 | 294 | 2 1 |
| 75 | NC_004318.11 | 1187 | 1041 | 146 | 2 1 |
| 76 | NC_004318.12 | 1539 | 987 | 552 | 5 4 |
| 77 | NC_004318.13 | 1680 | 1353 | 327 | 3 2 |
| 78 | NC_004318.14 | 4698 | 4167 | 531 | 6 5 |
| 79 | NC_004318.15 | 2234 | 1869 | 365 | 3 2 |
| 80 | NC_004318.16 | 3985 | 3669 | 316 | 3 2 |
| 81 | NC_004318.17 | 1019 | 945 | 74 | 2 1 |
| 82 | NC_004325.1 | 7617 | 6492 | 1125 | 2 1 |
| 83 | NC_004325.2 | 1274 | 1119 | 155 | 2 1 |
| 84 | NC_004325.3 | 790 | 627 | 163 | 2 1 |
| 85 | NC_004325.4 | 3464 | 3258 | 206 | 2 1 |
| 86 | NC_004325.5 | 3614 | 3321 | 293 | 3 2 |
| 87 | NC_004325.6 | 1771 | 1386 | 385 | 2 1 |
| 88 | NC_004325.7 | 1968 | 645 | 1323 | 9 8 |
| 89 | NC_004325.8 | 1327 | 507 | 820 | 6 5 |
| 90 | NC_004325.9 | 4786 | 4605 | 181 | 2 1 |

| No | Nama Sekuen | Panjang Sekuen(bp) | Panjang(bp) | | Jumlah |
|-----|--------------|--------------------|-------------|--------|--------|
| | | | Ekson | Intron | |
| 91 | NC_004325.10 | 3701 | 3357 | 344 | 2 1 |
| 92 | NC_004325.11 | 6431 | 5133 | 1298 | 7 6 |
| 93 | NC_004325.12 | 1687 | 1104 | 583 | 3 2 |
| 94 | NC_004325.13 | 1773 | 1209 | 564 | 4 3 |
| 95 | NC_004325.14 | 4504 | 4356 | 148 | 2 1 |
| 96 | NC_004325.15 | 1339 | 1107 | 232 | 2 1 |
| 97 | NC_004325.16 | 1008 | 906 | 102 | 2 1 |
| 98 | NC_004326.1 | 7528 | 6624 | 904 | 2 1 |
| 99 | NC_004326.2 | 1184 | 1014 | 170 | 2 1 |
| 100 | NC_004326.3 | 5334 | 5163 | 171 | 2 1 |
| 101 | NC_004326.4 | 1339 | 909 | 490 | 6 5 |
| 102 | NC_004326.5 | 7344 | 7182 | 162 | 2 1 |
| 103 | NC_004326.6 | 1854 | 1509 | 345 | 3 2 |
| 104 | NC_004326.7 | 1814 | 1083 | 731 | 4 3 |
| 105 | NC_004326.8 | 1550 | 1017 | 533 | 5 4 |
| 106 | NC_004326.9 | 2301 | 2169 | 132 | 2 1 |
| 107 | NC_004326.10 | 5041 | 4680 | 361 | 4 3 |
| 108 | NC_004326.11 | 2662 | 1542 | 1120 | 10 9 |
| 109 | NC_004326.12 | 3004 | 2850 | 154 | 2 1 |
| 110 | NC_004326.13 | 1454 | 603 | 851 | 6 5 |
| 111 | NC_004326.14 | 7376 | 7203 | 173 | 2 1 |
| 112 | NC_004326.15 | 2152 | 1323 | 829 | 6 5 |
| 113 | NC_004326.16 | 4287 | 3999 | 288 | 3 2 |
| 114 | NC_004326.17 | 3756 | 3426 | 330 | 3 2 |
| 115 | NC_004326.18 | 931 | 570 | 361 | 3 2 |
| 116 | NC_004326.19 | 2571 | 2094 | 477 | 5 4 |
| 117 | NC_004326.20 | 1695 | 1464 | 231 | 2 1 |
| 118 | NC_004326.21 | 2519 | 1839 | 680 | 6 5 |
| 119 | NC_004326.22 | 1999 | 888 | 1111 | 8 7 |
| 120 | NC_004326.23 | 1742 | 1377 | 365 | 2 1 |

| No | Nama Sekuen | Panjang Sekuen(bp) | | Panjang(bp) | | Jumlah | |
|-----|--------------|--------------------|--------|-------------|--------|--------|--------|
| | | Ekson | Intron | Ekson | Intron | Ekson | Intron |
| 137 | NC_004330.11 | 1027 | 538 | 489 | 6 | 5 | |
| 138 | NC_004330.12 | 2057 | 944 | 1113 | 5 | 4 | |
| 139 | NC_004330.13 | 1932 | 384 | 1548 | 3 | 2 | |
| 140 | NC_004330.14 | 1120 | 544 | 576 | 5 | 4 | |
| 141 | NC_004330.15 | 1771 | 328 | 1443 | 2 | 1 | |
| 142 | NC_004330.16 | 4890 | 261 | 4629 | 3 | 2 | |
| 143 | NC_004330.17 | 3338 | 398 | 2940 | 4 | 3 | |
| 144 | NC_004330.18 | 861 | 507 | 354 | 5 | 4 | |
| 145 | NC_004330.19 | 4253 | 419 | 3834 | 4 | 3 | |
| 146 | NC_004330.20 | 1757 | 878 | 879 | 8 | 7 | |
| 147 | NC_004330.21 | 1070 | 248 | 822 | 2 | 1 | |
| 148 | NC_004330.22 | 1068 | 495 | 573 | 4 | 3 | |
| 149 | NC_004330.23 | 1914 | 885 | 1029 | 5 | 4 | |
| 150 | NC_004330.24 | 5915 | 1892 | 4023 | 9 | 8 | |
| 151 | NC_004330.25 | 1274 | 158 | 1116 | 2 | 1 | |
| 152 | NC_004330.26 | 4118 | 170 | 3948 | 2 | 1 | |

| No | Nama Sekuen | Panjang Sekuen(bp) | | Panjang(bp) | | Jumlah | |
|-----|--------------|--------------------|--------|-------------|--------|--------|--------|
| | | Ekson | Intron | Ekson | Intron | Ekson | Intron |
| 121 | NC_004326.24 | 2232 | 1011 | 1221 | 8 | 7 | |
| 122 | NC_004326.25 | 1935 | 990 | 945 | 9 | 8 | |
| 123 | NC_004326.26 | 2360 | 713 | 1647 | 6 | 5 | |
| 124 | NC_004326.27 | 1207 | 667 | 540 | 6 | 5 | |
| 125 | NC_004326.28 | 1684 | 97 | 1587 | 2 | 1 | |
| 126 | NC_004326.29 | 1185 | 228 | 957 | 2 | 1 | |
| 127 | NC_004330.1 | 7806 | 993 | 6813 | 2 | 1 | |
| 128 | NC_004330.2 | 1187 | 155 | 1032 | 2 | 1 | |
| 129 | NC_004330.3 | 1014 | 102 | 912 | 2 | 1 | |
| 130 | NC_004330.4 | 1901 | 539 | 1362 | 3 | 2 | |
| 131 | NC_004330.5 | 1886 | 431 | 1455 | 2 | 1 | |
| 132 | NC_004330.6 | 1557 | 969 | 588 | 7 | 6 | |
| 133 | NC_004330.7 | 1974 | 201 | 1773 | 2 | 1 | |
| 134 | NC_004330.8 | 1399 | 799 | 600 | 7 | 6 | |
| 135 | NC_004330.9 | 1358 | 341 | 1017 | 2 | 1 | |
| 136 | NC_004330.10 | 2292 | 333 | 1959 | 2 | 1 | |

| No | Nama Sekuen | Panjang tiap lokasi ekson/intron (bp) | | | | | | | | | | | | | | | | | | |
|----|--------------|---------------------------------------|----------|---------|---|---|---|---|---|---|----|--|--|--|--|--|--|--|--|--|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | | | | | | | | |
| 1 | NC_000521.1 | 5319/888 | 1311 | | | | | | | | | | | | | | | | | |
| 2 | NC_000521.2 | 54/246 | 912 | | | | | | | | | | | | | | | | | |
| 3 | NC_000521.3 | 69/104 | 578/120 | 34 | | | | | | | | | | | | | | | | |
| 4 | NC_000521.4 | 69/210 | 702 | | | | | | | | | | | | | | | | | |
| 5 | NC_000521.5 | 862/91 | 576/111 | 54/109 | | | | | | | | | | | | | | | | |
| 6 | NC_000521.6 | 856/96 | 576/116 | 54/71 | | | | | | | | | | | | | | | | |
| 7 | NC_000521.7 | 29/242 | 3229/113 | 101/185 | | | | | | | | | | | | | | | | |
| 8 | NC_000521.8 | 1308/91 | 74/142 | 41/140 | | | | | | | | | | | | | | | | |
| 9 | NC_000521.9 | 3238/58 | 995 | | | | | | | | | | | | | | | | | |
| 10 | NC_000521.10 | 38/109 | 118/114 | 147/128 | | | | | | | | | | | | | | | | |
| 11 | NC_000521.11 | 10/203 | 179/513 | 249 | | | | | | | | | | | | | | | | |
| 12 | NC_000521.12 | 1284/153 | 195 | | | | | | | | | | | | | | | | | |
| 13 | NC_000521.13 | 1311/240 | 119/119 | 76/83 | | | | | | | | | | | | | | | | |
| 14 | NC_000521.14 | 80/675 | 1753 | | | | | | | | | | | | | | | | | |
| 15 | NC_000521.15 | 55/126 | 128/92 | | | | | | | | | | | | | | | | | |
| 16 | NC_000521.16 | 978/218 | 492/175 | 77/148 | | | | | | | | | | | | | | | | |
| 17 | NC_000521.17 | 51/511 | 387 | 75/119 | | | | | | | | | | | | | | | | |
| 18 | NC_000521.18 | 913/195 | 3191 | | | | | | | | | | | | | | | | | |
| 19 | NC_000521.19 | 303/162 | 1397/166 | 104/116 | | | | | | | | | | | | | | | | |
| 20 | NC_000521.20 | 26/204 | 219/158 | 1585 | | | | | | | | | | | | | | | | |
| 21 | NC_000521.21 | 16/410 | 113/209 | 252 | | | | | | | | | | | | | | | | |
| 22 | NC_000521.22 | 3/154 | 48/109 | 608/147 | | | | | | | | | | | | | | | | |
| 23 | NC_000521.23 | 1536/376 | 376/116 | 337/152 | | | | | | | | | | | | | | | | |
| 24 | NC_000521.24 | 282/222 | 198/173 | 105/156 | | | | | | | | | | | | | | | | |
| 25 | NC_000521.25 | 457/252 | 227/361 | 2037 | | | | | | | | | | | | | | | | |
| 26 | NC_000521.26 | 3003/178 | 843 | | | | | | | | | | | | | | | | | |
| 27 | NC_000521.27 | 475/312 | 177/107 | 806 | | | | | | | | | | | | | | | | |
| 28 | NC_000521.28 | 2728/213 | 83 | | | | | | | | | | | | | | | | | |
| 29 | NC_000521.29 | 663/239 | 63/242 | 417 | | | | | | | | | | | | | | | | |
| 30 | NC_000521.30 | 304/513 | 362 | | | | | | | | | | | | | | | | | |

| No | Nama Sekuen | Panjang tiap lokasi ekson/intron (bp) | | | | | | | | | | | | |
|----|--------------|---------------------------------------|----------|----------|---------|---------|---------|--------|---------|------|----|--|--|--|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | | |
| 31 | NC_000521.31 | 1152/291 | 108/60 | 1590 | | | | | | | | | | |
| 32 | NC_000521.32 | 24/448 | 371/182 | 91 | | | | | | | | | | |
| 33 | NC_000521.33 | 762/108 | 121/167 | 95/102 | 96 | | | | | | | | | |
| 34 | NC_000521.34 | 121/111 | 99/135 | 155/111 | 555 | | | | | | | | | |
| 35 | NC_000521.35 | 128/155 | 151/157 | 192 | | | | | | | | | | |
| 36 | NC_000521.36 | 142/306 | 103/226 | 9044/161 | 113 | | | | | | | | | |
| 37 | NC_000521.37 | 167/186 | 107/195 | 87/102 | 147/129 | 44/119 | 114 | | | | | | | |
| 38 | NC_000521.38 | 391/224 | 911 | | | | | | | | | | | |
| 39 | NC_000521.39 | 94/395 | 252/211 | 131 | | | | | | | | | | |
| 40 | NC_000521.40 | 305/89 | 539/95 | 49/98 | 27/83 | 67 | | | | | | | | |
| 41 | NC_000521.41 | 1855/305 | 353 | | | | | | | | | | | |
| 42 | NC_000521.42 | 72/98 | 978 | | | | | | | | | | | |
| 43 | NC_000521.43 | 69/146 | 885 | | | | | | | | | | | |
| 44 | NC_000521.44 | 69/105 | 837 | | | | | | | | | | | |
| 45 | NC_000521.45 | 69/243 | 1047 | | | | | | | | | | | |
| 46 | NC_000910.1 | 118/215 | 947 | | | | | | | | | | | |
| 47 | NC_000910.2 | 21/277 | 2685/262 | 124/131 | 80 | | | | | | | | | |
| 48 | NC_000910.3 | 46/204 | 221/175 | 147 | | | | | | | | | | |
| 49 | NC_000910.4 | 634/156 | 89 | | | | | | | | | | | |
| 50 | NC_000910.5 | 9/488 | 406/157 | 572 | | | | | | | | | | |
| 51 | NC_000910.6 | 109/505 | 509/167 | 330/293 | 84 | | | | | | | | | |
| 52 | NC_000910.7 | 52/134 | 3965/604 | 78 | | | | | | | | | | |
| 53 | NC_000910.8 | 1218/240 | 100/149 | 119/132 | 78/129 | 60 | | | | | | | | |
| 54 | NC_000910.9 | 198/158 | 4479 | | | | | | | | | | | |
| 55 | NC_000910.10 | 922/159 | 570/112 | 54/97 | 63/145 | 132/105 | 186/112 | 159/87 | 767/134 | 1470 | | | | |
| 56 | NC_000910.11 | 57/238 | 855 | | | | | | | | | | | |
| 57 | NC_000910.12 | 69/135 | 882 | | | | | | | | | | | |
| 58 | NC_000910.13 | 54/244 | 1053 | | | | | | | | | | | |
| 59 | NC_000910.14 | 69/148 | 978 | | | | | | | | | | | |
| 60 | NC_000910.15 | 48/101 | 846 | | | | | | | | | | | |

| No | Nama Sekuen | Panjang tiap lokasi ekson/intron (bp) | | | | | | | | | | | | | | | | | | |
|----|--------------|---------------------------------------|----------|----------|--------|---------|---------|---------|--------|-----|----|--|--|--|--|--|--|--|--|--|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | | | | | | | | |
| 61 | NC_000910.16 | 21/470 | 939 | | | | | | | | | | | | | | | | | |
| 62 | NC_000910.17 | 69/146 | 969 | | | | | | | | | | | | | | | | | |
| 63 | NC_000910.18 | 72/409 | 1224 | | | | | | | | | | | | | | | | | |
| 64 | NC_000910.19 | 54/230 | 930 | | | | | | | | | | | | | | | | | |
| 65 | NC_004318.1 | 54/155 | 1095 | | | | | | | | | | | | | | | | | |
| 66 | NC_004318.2 | 130/98 | 141/178 | 187/101 | 59/94 | 164/243 | 124/138 | 260/223 | 88/97 | 587 | | | | | | | | | | |
| 67 | NC_004318.3 | 51/315 | 9360 | | | | | | | | | | | | | | | | | |
| 68 | NC_004318.4 | 37/109 | 320/185 | 118/219 | 142/94 | 43/50 | 159 | | | | | | | | | | | | | |
| 69 | NC_004318.5 | 33/170 | 65/176 | 46/82 | 76/127 | 114/163 | 265/195 | 220 | | | | | | | | | | | | |
| 70 | NC_004318.6 | 75/238 | 966 | | | | | | | | | | | | | | | | | |
| 71 | NC_004318.7 | 3013/139 | 1007 | | | | | | | | | | | | | | | | | |
| 72 | NC_004318.8 | 156/157 | 549 | | | | | | | | | | | | | | | | | |
| 73 | NC_004318.9 | 171/174 | 62/509 | 500 | | | | | | | | | | | | | | | | |
| 74 | NC_004318.10 | 69/294 | 774 | | | | | | | | | | | | | | | | | |
| 75 | NC_004318.11 | 69/146 | 972 | | | | | | | | | | | | | | | | | |
| 76 | NC_004318.12 | 123/192 | 333/115 | 91/150 | 131/95 | 309 | | | | | | | | | | | | | | |
| 77 | NC_004318.13 | 62/208 | 12/119 | 1170 | | | | | | | | | | | | | | | | |
| 78 | NC_004318.14 | 179/32 | 1047/97 | 2666/169 | 79/140 | 134/93 | 62 | | | | | | | | | | | | | |
| 79 | NC_004318.15 | 102/221 | 1692/144 | 75 | | | | | | | | | | | | | | | | |
| 80 | NC_004318.16 | 201/165 | 3387/151 | 81 | | | | | | | | | | | | | | | | |
| 81 | NC_004318.17 | 72/74 | 873 | | | | | | | | | | | | | | | | | |
| 82 | NC_004325.1 | 5253/1125 | 1239 | | | | | | | | | | | | | | | | | |
| 83 | NC_004325.2 | 54/155 | 1065 | | | | | | | | | | | | | | | | | |
| 84 | NC_004325.3 | 27/163 | 600 | | | | | | | | | | | | | | | | | |
| 85 | NC_004325.4 | 195/206 | 3063 | | | | | | | | | | | | | | | | | |
| 86 | NC_004325.5 | 79/64 | 2902/229 | 340 | | | | | | | | | | | | | | | | |
| 87 | NC_004325.6 | 49/385 | 1337 | | | | | | | | | | | | | | | | | |
| 88 | NC_004325.7 | 100/218 | 84/208 | 70/148 | 39/154 | 113/129 | 53/255 | 94/96 | 53/115 | 39 | | | | | | | | | | |
| 89 | NC_004325.8 | 124/194 | 68/222 | 87/96 | 81/196 | 60/112 | 87 | | | | | | | | | | | | | |
| 90 | NC_004325.9 | 3083/181 | 1522 | | | | | | | | | | | | | | | | | |

| No | Nama Sekuen | Panjang tiap lokasi ekson/intron (bp) | | | | | | | | | | | | |
|-----|--------------|---------------------------------------|---------|---------|---------|---------|---------|---------|--------|-------|----|--|--|--|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | | |
| 91 | NC_004325.10 | 1645/344 | 1712 | 398/350 | 80/141 | 39/322 | 50/199 | 38 | | | | | | |
| 92 | NC_004325.11 | 4395/186 | 133/100 | 573 | 566 | | | | | | | | | |
| 93 | NC_004325.12 | 73/386 | 458/197 | 400/218 | | | | | | | | | | |
| 94 | NC_004325.13 | 111/228 | 132/118 | | | | | | | | | | | |
| 95 | NC_004325.14 | 126/148 | 4230 | | | | | | | | | | | |
| 96 | NC_004325.15 | 54/232 | 1053 | | | | | | | | | | | |
| 97 | NC_004325.16 | 69/102 | 837 | | | | | | | | | | | |
| 98 | NC_004326.1 | 5382/904 | 1242 | | | | | | | | | | | |
| 99 | NC_004326.2 | 723/170 | 291 | | | | | | | | | | | |
| 100 | NC_004326.3 | 300/171 | 4863 | | | | | | | | | | | |
| 101 | NC_004326.4 | 102/93 | 39/67 | 52/75 | 191/95 | 122/160 | 403 | | | | | | | |
| 102 | NC_004326.5 | 6848/162 | 334 | | | | | | | | | | | |
| 103 | NC_004326.6 | 596/199 | 726/146 | 187 | | | | | | | | | | |
| 104 | NC_004326.7 | 53/198 | 175/161 | 81/372 | 774 | | | | | | | | | |
| 105 | NC_004326.8 | 58/127 | 172/150 | 163/161 | 433/95 | 191 | | | | | | | | |
| 106 | NC_004326.9 | 301/132 | 1868 | | | | | | | | | | | |
| 107 | NC_004326.10 | 4176/96 | 316/150 | 89/115 | 99 | | | | | | | | | |
| 108 | NC_004326.11 | 69/164 | 152/144 | 86/123 | 98/121 | 640/149 | 161/96 | 90/107 | 88/140 | 82/76 | 76 | | | |
| 109 | NC_004326.12 | 1092/154 | 1758 | | | | | | | | | | | |
| 110 | NC_004326.13 | 14/237 | 73/262 | 96/90 | 64/131 | 76/131 | 280 | | | | | | | |
| 111 | NC_004326.14 | 6865/173 | 338 | | | | | | | | | | | |
| 112 | NC_004326.15 | 100/212 | 65/169 | 104/100 | 61/150 | 860/198 | 133 | | | | | | | |
| 113 | NC_004326.16 | 125/156 | 406/132 | 3468 | | | | | | | | | | |
| 114 | NC_004326.17 | 3207/218 | 139/112 | 80 | | | | | | | | | | |
| 115 | NC_004326.18 | 63/216 | 68/145 | 439 | | | | | | | | | | |
| 116 | NC_004326.19 | 1536/168 | 126/95 | 64/102 | 269/112 | 99 | | | | | | | | |
| 117 | NC_004326.20 | 329/231 | 1135 | | | | | | | | | | | |
| 118 | NC_004326.21 | 283/211 | 1216/86 | 72/107 | 111/120 | 90/156 | 67 | | | | | | | |
| 119 | NC_004326.22 | 141/192 | 60/237 | 97/140 | 135/117 | 143/135 | 135/170 | 120/120 | 57 | | | | | |
| 120 | NC_004326.23 | 19/365 | 1358 | | | | | | | | | | | |

| No | Nama Sekuen | Panjang tiap lokasi ekson/intron (bp) | | | | | | | | | | | | |
|-----|--------------|---------------------------------------|---------|---------|---------|---------|---------|---------|---------|------|----|--|--|--|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | | |
| 121 | NC_004326.24 | 117/166 | 276/129 | 65/166 | 415/228 | 95/89 | 73/130 | 79/103 | 101 | | | | | |
| 122 | NC_004326.25 | 88/105 | 113/100 | 87/131 | 62/92 | 213/152 | 67/196 | 93/93 | 94/121 | 128 | | | | |
| 123 | NC_004326.26 | 157/114 | 153/153 | 144/112 | 176/165 | 903/169 | 114 | | | | | | | |
| 124 | NC_004326.27 | 78/149 | 80/167 | 98/93 | 134/99 | 68/159 | 82 | | | | | | | |
| 125 | NC_004326.28 | 198/97 | 1389 | | | | | | | | | | | |
| 126 | NC_004326.29 | 72/228 | 885 | | | | | | | | | | | |
| 127 | NC_004330.1 | 5454/993 | 1359 | | | | | | | | | | | |
| 128 | NC_004330.2 | 54/155 | 978 | | | | | | | | | | | |
| 129 | NC_004330.3 | 69/102 | 843 | | | | | | | | | | | |
| 130 | NC_004330.4 | 62/344 | 121/195 | 1179 | | | | | | | | | | |
| 131 | NC_004330.5 | 14/431 | 1441 | | | | | | | | | | | |
| 132 | NC_004330.6 | 63/125 | 92/134 | 115/134 | 114/113 | 69/173 | 84/290 | 51 | | | | | | |
| 133 | NC_004330.7 | 168/1201 | 92 | | | | | | | | | | | |
| 134 | NC_004330.8 | 77/146 | 70/134 | 113/80 | 130/160 | 57/155 | 85/124 | 68 | | | | | | |
| 135 | NC_004330.9 | 520/341 | 497 | | | | | | | | | | | |
| 136 | NC_004330.10 | 76/333 | 1883 | | | | | | | | | | | |
| 137 | NC_004330.11 | 9/173 | 85/86 | 86/34 | 44/38 | 149/207 | 116 | | | | | | | |
| 138 | NC_004330.12 | 91/555 | 155/100 | 118/131 | 119/158 | 630 | | | | | | | | |
| 139 | NC_004330.13 | 927/186 | 514/198 | 107 | | | | | | | | | | |
| 140 | NC_004330.14 | 163/137 | 77/87 | 76/194 | 148/126 | 112 | | | | | | | | |
| 141 | NC_004330.15 | 746/328 | 697 | | | | | | | | | | | |
| 142 | NC_004330.16 | 4245/172 | 75/89 | 309 | | | | | | | | | | |
| 143 | NC_004330.17 | 2678/140 | 112/174 | 101/84 | 49 | | | | | | | | | |
| 144 | NC_004330.18 | 78/224 | 75/79 | 67/96 | 81/108 | 53 | | | | | | | | |
| 145 | NC_004330.19 | 108/145 | 104/106 | 68/168 | 3554 | | | | | | | | | |
| 146 | NC_004330.20 | 122/116 | 80/133 | 87/99 | 73/85 | 85/130 | 246/139 | 80/176 | 106 | | | | | |
| 147 | NC_004330.21 | 375/248 | 447 | | | | | | | | | | | |
| 148 | NC_004330.22 | 45/211 | 277/154 | 125/130 | 126 | | | | | | | | | |
| 149 | NC_004330.23 | 127/184 | 158/240 | 185/137 | 160/324 | 399 | | | | | | | | |
| 150 | NC_004330.24 | 775/350 | 603/108 | 54/144 | 63/156 | 132/148 | 186/363 | 159/411 | 767/212 | 1284 | | | | |
| 151 | NC_004330.25 | 57/158 | 1059 | | | | | | | | | | | |
| 152 | NC_004330.26 | 2712/170 | 1236 | | | | | | | | | | | |

LAMPIRAN 2

Pengujian Data Sekuen

```
clear
close all
clc
diary dataexon.txt
diary off
diary dataintron.txt
diary off
diary datasequence.txt
diary off
diary dataNameSeq.txt
diary off
diary dataestimatedStates.txt
diary off
diary dataexontes.txt
diary off
diary datapemisah.txt
diary off
delete('datapemisah.txt')
delete('dataexontes.txt')
delete('dataestimatedStates.txt')
delete('dataNameSeq.txt')
delete('dataexon.txt')
delete('dataintron.txt')
delete('datasequence.txt')
% Pembacaan nama sekuen
% =====
get = {'NC_004325.1.txt','NC_004325.2.txt','NC_004325.3.txt'};
experiment = sek;
test = sek;
file(experiment,test,'all sequence');
function [Sn,Sp,CC,AC ] = file(experiment,test,fuih)

% Memasukkan banyaknya jumlah file yang akan di proses
% =====
sek = experiment;
jumlahfile = length(sek);

% Memasukkan file yang akan di proses
% =====
for t = 1:jumlahfile;
    masukkan = char(sek(t));
    s = genbankread(masukkan);
    [exon,intron,sequence,Jexon,Jintron,NameSeq] = genom(masukkan);
    diary dataNameSeq.txt;
    disp(NameSeq);
```

```

diary off;
diary dataexon.txt;
disp(exon);
diary off;
diary dataintron.txt;
disp(intron);
diary off;
diary datasequence.txt;
disp(sequence);
% diary datatr.doc;
disp(Jexon);
disp(Jintron);
diary off;
end

```

Hasil Pengujian Data Sekuen

Nama Sekuen:
NC_004318.1.txt

Total Ekson:

```

Atgaaaatccattataactaataatattattgtttcctctaaaattaaatatatttggtaaataccacaaaaaa
ccatccattacaccacgtcatatacaaaactaacagattattatgtgagtgcaattatatgcaccacaaaat
tatgataatgacccggaatgaaaagggtcatgcaacaatttcatgatcgtacaacacaaagggttcaagaa
tacgatgaaagggtacaagaaaaacgacaagatgtaaagatacatgtgataaggaaaatccaaaaattatt
ttaaagataaattagaaaaagaattaatggacaaatttgccacactacaaacagataatacaaagtgatgct
attccaacatgtatttgcgaaaagtcgtagcagataaaagtggaaaaaggatgttgagatgtgcacaaaat
ttgggagggattgttgcaccctctcaggagtagcaggaattgctgaagggtgcgctatatgcgtggaaa
cctaaggcacttgaggcctctattaaaacagctatagctgagggtaactgctaataatttggctgcagggtgt
gaagcaggtgaggttacaggaaggaattagttattgcaggattaaaaaaaatgggtatatcaactctagat
aataagtcattggaatcctattttgctacaacatctataaaaatatcacaacattgctcaagctgttcaa
aaactatatttggagacatgtgcatgtgattcttcagggaagtggtctacctttacggtgacgctaaccgt
catattcctatttgcattcgggtgtggaaccaaactccagctgtattaacaacaaaaaagggtatttcaacc
aaagaagttatagaaaaaactgtacaaactatgggtgacagacgccgaagggtgctgccgacgctagtgtc
gctaaaagtgccgagtttaactgctgcaataaaaagctcgagaggccgaagtgataaatactatatttatgagt
aagcaaaactgctattattgcttctgtttgcaatattaattatagtttaattatggtgataatatttg
attttacggttatagacgaaaaaaaatgaagaaaaaactacaatacataaaattattaaatcaataa

```

Total Intron:

```

Gtaacattatgtgataaataaataaaatattcataaatataattattatatctatatgacttatatatatag
tttaaatatacatattaattacaaatatataaaatcataaactacaaataaacattacatatatatgtcacc
atttttatag

```

Sekuen Asli:

atgaaaatccattataactaatatattattgtttcctctaaaattaatataattggtaacattatgtgataaa
taaataaaaatattcataaatataattattatatatctatatgacttatatatatagtttaaatatacatattaa
ttacaaatatataaaaatcataaaactacaaataaacattacatatatatatgtcaccattttttataggtaaata
cccacaaaaaacatccattacaccacgctcatatacaaaactaacagattattatgtgagtgcaattatg
caccacaaaattatgataatgaccggaaatgaaaagggcatgcaacaatttcatgatcgtacaacacaaa
ggtttcaagaatac gatgaaaggttacaagaaaaacgacaagtatgtaagatacatgtgataaggaaatcc
aaaaaattattttaaaagataaattagaaaaagaattaatggacaaatttgccacactacaaacagatatac
aaagtgatgctattccaacatgtatttgcgaaaaagtcgttagcagataaaagtggaaaaaggatgttgagat
gtgcacaaaatttgggagggattgttgcaccctcttcaggagatttagcaggaaattgctgaagggtgcgctat
atgcggtggaacctaaggcacttgaggcctctattaaaacagctatagctgagggtactgctaattttgg
ctgcaggtgttgaagcaggtgaggttacaggttaaggaattagttattgcaggattaaaaaaaatgggtatat
caactctagataataagtcatgtggaatcctattttgctacaacatctataaaaaatcacaaacattgctc
aagctgttcaaaaactatattttgagacatgtgcatgtgattcttcagggaaagtgggtctaccttacgggtg
acgctaaccgtcatattcctatttgcattcggtgtggaacaaactccagctgtattaacaacaaaaaagg
gtatttcaaccaaagaagttatagaaaaactgtacaaactatgggtgtcagacgccgaaggggttgcgccc
acgctagtgtgctgtaaaagtgcggagttaactgctgcaataaaagctcgagaggccgaagtataaatacta
tatttatgagtaagcaaaactgctattattgcttctgttgtgcaatattaattatagttttaattatgttga
taatataatttgattttacgttatagacgaaaaaaaatgaagaaaaactacaatacataaaattattaa
atcaataa

Jumlah Lokasi Ekson:

2

Jumlah Lokasi Intron:

1

LAMPIRAN 3

Hasil Ujicoba Struktur Dasar Model 5 State.

| Ujicoba | Transition Matrix | | | | | | | Corr.Coeff.(CC) | | Iterations |
|---------|-------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----------------|---------|------------|
| | tr ₁₁ | tr ₂₂ | tr ₃₃ | tr ₄₄ | tr ₅₅ | tr ₁₃ | tr ₄₃ | Viterbi | B-Welch | |
| 1 | 0 | 0.9 | 0.9 | 0.88 | 1 | 0 | 0.1 | 0.6091 | 0.6225 | 29 |
| 2 | 0 | 0.7 | 0.9 | 0.88 | 1 | 0 | 0.1 | 0.6988 | 0.6648 | 26 |
| 3 | 0 | 0.6 | 0.9 | 0.88 | 1 | 0 | 0.1 | 0.6988 | 0.6648 | 26 |
| 4 | 0 | 0.5 | 0.9 | 0.88 | 1 | 0 | 0.1 | 0.7042 | 0.6724 | 22 |
| 5 | 0 | 0.5 | 0.8 | 0.88 | 1 | 0 | 0.1 | 0.0262 | 0.0251 | 15 |
| 6 | 0 | 0.5 | 0.5 | 0.88 | 1 | 0 | 0.1 | -0.0131 | -0.0135 | 16 |
| 7 | 0 | 0.5 | 0.4 | 0.88 | 1 | 0 | 0.1 | -0.0131 | -0.0135 | 16 |
| 8 | 0 | 0.5 | 0.3 | 0.88 | 1 | 0 | 0.1 | -0.0131 | -0.0135 | 16 |
| 9 | 0 | 0.5 | 0.2 | 0.88 | 1 | 0 | 0.1 | 0.0216 | 0.0117 | 13 |
| 10 | 0 | 0.5 | 0.1 | 0.88 | 1 | 0 | 0.1 | 0.0216 | 0.0117 | 13 |
| 11 | 0 | 0.5 | 0.9 | 0.78 | 1 | 0 | 0.2 | -0.0862 | -0.092 | 18 |
| 12 | 0 | 0.5 | 0.9 | 0.68 | 1 | 0 | 0.3 | -0.0848 | -0.0861 | 20 |
| 13 | 0 | 0.5 | 0.9 | 0.58 | 1 | 0 | 0.4 | -0.0848 | -0.0861 | 23 |
| 14 | 0 | 0.5 | 0.9 | 0.48 | 1 | 0 | 0.5 | -0.0821 | -0.0838 | 22 |
| 15 | 0 | 0.5 | 0.9 | 0.38 | 1 | 0 | 0.6 | -0.0821 | -0.0838 | 26 |
| 16 | 0 | 0.5 | 0.9 | 0.28 | 1 | 0 | 0.7 | -0.0821 | -0.0838 | 25 |
| 17 | 0 | 0.5 | 0.9 | 0.18 | 1 | 0 | 0.8 | -0.0821 | -0.0838 | 25 |
| 18 | 0 | 0.5 | 0.9 | 0.08 | 1 | 0 | 0.9 | -0.0821 | -0.0838 | 25 |
| 19 | 0 | 0.5 | 0.9 | 0.78 | 1 | 0 | 0.1 | -0.0824 | -0.0805 | 17 |
| 20 | 0 | 0.5 | 0.9 | 0.68 | 1 | 0 | 0.1 | -0.0848 | -0.0861 | 22 |
| 21 | 0 | 0.5 | 0.9 | 0.58 | 1 | 0 | 0.1 | -0.0848 | -0.0861 | 23 |
| 22 | 0 | 0.5 | 0.9 | 0.48 | 1 | 0 | 0.1 | -0.0848 | -0.0861 | 23 |
| 23 | 0 | 0.5 | 0.9 | 0.38 | 1 | 0 | 0.1 | -0.0848 | -0.0861 | 21 |
| 24 | 0 | 0.5 | 0.9 | 0.28 | 1 | 0 | 0.1 | -0.0848 | -0.0861 | 21 |
| 25 | 0 | 0.5 | 0.9 | 0.18 | 1 | 0 | 0.1 | -0.0848 | -0.0861 | 24 |
| 26 | 0 | 0.5 | 0.9 | 0.08 | 1 | 0 | 0.1 | -0.0848 | -0.0861 | 24 |
| 27 | 0 | 0.5 | 0.9 | 0.90 | 1 | 0 | 0.05 | 0.0011 | 0.0013 | 14 |
| 28 | 0 | 0.5 | 0.9 | 0.90 | 1 | 0 | 0.1 | 0.0011 | 0.0013 | 18 |
| 29 | 0 | 0.5 | 0.9 | 0.80 | 1 | 0 | 0.05 | -0.0741 | -0.0761 | 16 |
| 30 | 0 | 0.5 | 0.9 | 0.80 | 1 | 0 | 0.1 | -0.0743 | -0.0763 | 17 |
| 31 | 0 | 0.5 | 0.9 | 0.80 | 1 | 0 | 0.15 | -0.0741 | -0.0761 | 16 |
| 32 | 0 | 0.5 | 0.9 | 0.80 | 1 | 0 | 0.2 | 0.7054 | 0.6750 | 29 |
| 33 | 0 | 0.5 | 0.9 | 0.80 | 1 | 0 | 0 | -0.0824 | -0.0805 | 16 |
| 34 | 0 | 0.5 | 0.9 | 0.70 | 1 | 0 | 0 | -0.0848 | -0.0861 | 21 |
| 35 | 0 | 0.5 | 0.9 | 0.70 | 1 | 0 | 0.05 | -0.0848 | -0.0861 | 20 |
| 36 | 0 | 0.5 | 0.9 | 0.70 | 1 | 0 | 0.1 | -0.0848 | -0.0861 | 21 |
| 37 | 0 | 0.5 | 0.9 | 0.70 | 1 | 0 | 0.15 | -0.0848 | -0.0861 | 20 |
| 38 | 0 | 0.5 | 0.9 | 0.70 | 1 | 0 | 0.2 | -0.0848 | -0.0861 | 18 |
| 39 | 0 | 0.5 | 0.9 | 0.70 | 1 | 0 | 0.25 | -0.0824 | -0.0805 | 18 |
| 40 | 0 | 0.5 | 0.9 | 0.80 | 1 | 0 | 0.19 | -0.0741 | -0.0761 | 16 |
| 41 | 0 | 0.5 | 0.9 | 0.81 | 1 | 0 | 0.19 | 0.7061 | 0.6759 | 25 |
| 42 | 0 | 0.5 | 0.9 | 0.82 | 1 | 0 | 0.18 | 0.7054 | 0.6750 | 28 |
| 43 | 0 | 0.5 | 0.9 | 0.83 | 1 | 0 | 0.17 | 0.7054 | 0.6750 | 27 |
| 44 | 0 | 0.5 | 0.9 | 0.84 | 1 | 0 | 0.16 | 0.7045 | 0.6741 | 20 |
| 45 | 0 | 0.5 | 0.9 | 0.85 | 1 | 0 | 0.15 | 0.7054 | 0.6750 | 25 |

| Ujicoba | Transition Matrix | | | | | | | Corr.Coeff.(CC) | | Iterations |
|---------|-------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----------------|---------|------------|
| | tr ₁₁ | tr ₂₂ | tr ₃₃ | tr ₄₄ | tr ₅₅ | tr ₁₃ | tr ₄₃ | Viterbi | B-Weich | |
| 46 | 0 | 0.5 | 0.9 | 0.86 | 1 | 0 | 0.14 | 0.7062 | 0.6744 | 27 |
| 47 | 0 | 0.5 | 0.9 | 0.87 | 1 | 0 | 0.13 | 0.0012 | 0.0015 | 21 |
| 48 | 0 | 0.5 | 0.9 | 0.88 | 1 | 0 | 0.12 | 0.0011 | 0.0013 | 20 |
| 49 | 0 | 0.5 | 0.9 | 0.89 | 1 | 0 | 0.11 | 0.0011 | 0.0013 | 19 |
| 50 | 0 | 0.5 | 0.9 | 0.9 | 1 | 0 | 0.1 | 0.0011 | 0.0013 | 18 |
| 51 | 0 | 0.5 | 0.9 | 0.6 | 1 | 0 | 0 | -0.0848 | -0.0861 | 23 |
| 52 | 0 | 0.5 | 0.9 | 0.6 | 1 | 0 | 0.05 | -0.0848 | -0.0861 | 23 |
| 53 | 0 | 0.5 | 0.95 | 0.81 | 1 | 0.05 | 0.19 | 0.7061 | 0.6759 | 25 |
| 54 | 0 | 0.5 | 0.9 | 0.86 | 1 | 0.1 | 0.14 | 0.7044 | 0.6741 | 20 |
| 55 | 0 | 0.5 | 0.9 | 0.86 | 1 | 0.05 | 0.14 | 0.7062 | 0.6744 | 27 |
| 56 | 0 | 0.9 | 0.9 | 0.8 | 1 | 0.1 | 0.1 | 0.5688 | 0.5968 | 12 |
| 57 | 0 | 0.9 | 0.9 | 0.85 | 1 | 0.1 | 0.1 | 0.6091 | 0.6225 | 23 |
| 58 | 0 | 0.9 | 0.9 | 0.86 | 1 | 0.05 | 0.14 | 0.5708 | 0.5992 | 36 |
| 59 | 0 | 0.9 | 0.1 | 0.9 | 1 | 0.1 | 0.1 | 0.0380 | -0.0049 | 15 |
| 60 | 0 | 0.9 | 0.1 | 0.9 | 1 | 0 | 0.1 | 0.0376 | -0.0052 | 28 |
| 61 | 0 | 0.9 | 0.1 | 0.9 | 1 | 0.05 | 0.1 | 0.0380 | -0.0049 | 15 |
| 62 | 0 | 0.8 | 0.9 | 0.9 | 1 | 0 | 0.1 | 0.0010 | 0.0014 | 16 |
| 63 | 0 | 0.7 | 0.9 | 0.8 | 1 | 0.1 | 0.1 | -0.0741 | 0.0763 | 16 |
| 64 | 0 | 0.7 | 0.9 | 0.88 | 1 | 0.1 | 0.1 | 0.6988 | 0.6648 | 25 |
| 65 | 0 | 0.5 | 0.9 | 0.88 | 1 | 0.1 | 0.1 | 0.7042 | 0.6724 | 22 |
| 66 | 0 | 0.5 | 0.9 | 0.88 | 1 | 0.1 | 0.12 | 0.0011 | 0.0013 | 20 |
| 67 | 0 | 0.5 | 0.9 | 0.88 | 1 | 0.1 | 0.1 | 0.0011 | 0.0013 | 18 |
| 68 | 0 | 0.1 | 0.9 | 0.9 | 1 | 0.1 | 0.05 | 0.0011 | 0.0013 | 18 |
| 69 | 0 | 0.1 | 0.9 | 0.85 | 1 | 0.1 | 0.05 | 0.7018 | 0.6710 | 18 |
| 70 | 0 | 0.1 | 0.9 | 0.8 | 1 | 0.1 | -0.1 | -0.0743 | -0.0763 | 17 |
| 71 | 0 | 0.1 | 0.9 | 0.75 | 1 | 0.1 | 0.15 | -0.0824 | -0.0805 | 17 |
| 72 | 0 | 0.1 | 0.8 | 0.8 | 1 | 0.1 | 0.1 | 0.0011 | 0.0013 | 15 |
| 73 | 0 | 0.1 | 0.8 | 0.75 | 1 | 0.1 | 0.15 | 0.7018 | 0.6710 | 20 |
| 74 | 0 | 0.1 | 0.8 | 0.7 | 1 | 0.1 | 0.2 | -0.0824 | -0.0805 | 15 |
| 75 | 0 | 0.1 | 0.7 | 0.9 | 1 | 0.1 | 0.05 | -0.0131 | -0.0135 | 16 |
| 76 | 0 | 0.1 | 0.7 | 0.85 | 1 | 0.1 | 0.05 | -0.0131 | -0.0135 | 16 |
| 77 | 0 | 0.1 | 0.7 | 0.7 | 1 | 0.1 | 0.2 | 0.0011 | 0.0013 | 16 |
| 78 | 0 | 0.2 | 0.9 | 0.9 | 1 | 0.1 | 0.05 | 0.0011 | 0.0013 | 14 |
| 79 | 0 | 0.2 | 0.9 | 0.85 | 1 | 0.1 | 0.05 | 0.7018 | 0.6710 | 18 |
| 80 | 0 | 0.2 | 0.9 | 0.8 | 1 | 0.1 | 0.1 | -0.0743 | -0.0763 | 17 |
| 81 | 0 | 0.2 | 0.9 | 0.75 | 1 | 0.1 | 0.15 | -0.0824 | -0.0805 | 17 |
| 82 | 0 | 0.2 | 0.8 | 0.8 | 1 | 0.1 | 0.1 | 0.0011 | 0.0013 | 14 |
| 83 | 0 | 0.2 | 0.8 | 0.75 | 1 | 0.1 | 0.15 | 0.7018 | 0.6710 | 20 |
| 84 | 0 | 0.2 | 0.8 | 0.7 | 1 | 0.1 | 0.2 | -0.0824 | -0.0805 | 16 |
| 85 | 0 | 0.2 | 0.7 | 0.9 | 1 | 0.1 | 0.05 | -0.0131 | -0.0135 | 16 |
| 86 | 0 | 0.2 | 0.7 | 0.85 | 1 | 0.1 | 0.05 | -0.0131 | -0.0135 | 16 |
| 87 | 0 | 0.2 | 0.7 | 0.7 | 1 | 0.1 | 0.2 | 0.0011 | 0.0013 | 16 |
| 88 | 0 | 0.2 | 0.7 | 0.6 | 1 | 0.1 | 0.3 | -0.0824 | -0.0805 | 20 |
| 89 | 0 | 0.2 | 0.6 | 0.9 | 1 | 0.1 | 0.05 | -0.0131 | -0.0135 | 17 |
| 90 | 0 | 0.3 | 0.9 | 0.9 | 1 | 0.1 | 0.05 | 0.0011 | 0.0013 | 14 |
| 91 | 0 | 0.3 | 0.9 | 0.85 | 1 | 0.1 | 0.05 | 0.7018 | 0.6710 | 18 |
| 92 | 0 | 0.3 | 0.9 | 0.8 | 1 | 0.1 | 0.1 | -0.0743 | -0.0763 | 17 |
| 93 | 0 | 0.3 | 0.9 | 0.75 | 1 | 0.1 | 0.15 | -0.0824 | -0.0805 | 17 |

| Ujicoba | Transition Matrix | | | | | | | Corr. Coeff. (CC) | | Iterations |
|---------|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-------------------|---------|------------|
| | tr_{11} | tr_{22} | tr_{33} | tr_{44} | tr_{55} | tr_{13} | tr_{43} | Viterbi | B-Welch | |
| 94 | 0 | 0.3 | 0.9 | 0.7 | 1 | 0.1 | 0.2 | -0.0848 | -0.0861 | 18 |
| 95 | 0 | 0.3 | 0.9 | 0.6 | 1 | 0.1 | 0.3 | -0.0848 | -0.0861 | 23 |
| 96 | 0 | 0.4 | 0.9 | 0.9 | 1 | 0.1 | 0.05 | 0.0011 | 0.0013 | 14 |
| 97 | 0 | 0.4 | 0.9 | 0.85 | 1 | 0.1 | 0.05 | 0.7018 | 0.6710 | 18 |
| 98 | 0 | 0.4 | 0.9 | 0.8 | 1 | 0.1 | 0.1 | -0.0743 | -0.0763 | 17 |
| 99 | 0 | 0.4 | 0.9 | 0.75 | 1 | 0.1 | 0.15 | -0.0824 | -0.0805 | 17 |
| 100 | 0 | 0.4 | 0.9 | 0.7 | 1 | 0.1 | 0.2 | -0.0848 | -0.0861 | 18 |
| 101 | 0 | 0.4 | 0.9 | 0.6 | 1 | 0.1 | 0.3 | -0.0848 | -0.0861 | 23 |



LAMPIRAN 4

Hasil Uji coba Struktur Dasar Model 7 State.

| Uji coba | Transition Matrix | | | | | | | | Corr. Coef. (CC) | | Iterations |
|----------|-------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|---------|------------|
| | tr ₁₁ | tr ₂₂ | tr ₃₃ | tr ₄₄ | tr ₅₅ | tr ₆₆ | tr ₁₃ | tr ₆₃ | Viterbi | B-Welch | |
| 1 | 0 | 0.3 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 2 | 0 | 0.4 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 3 | 0 | 0.5 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 4 | 0 | 0.6 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 5 | 0 | 0.8 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6068 | 0.6233 | 20 |
| 6 | 0 | 0.9 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.0836 | 0.071 | 39 |
| 7 | 0 | 0.3 | 0.1 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 25 |
| 8 | 0 | 0.3 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 9 | 0 | 0.3 | 0.5 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0718 | -0.0777 | 18 |
| 10 | 0 | 0.3 | 0.6 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0701 | -0.0770 | 16 |
| 11 | 0 | 0.3 | 0.7 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0633 | -0.0700 | 15 |
| 12 | 0 | 0.3 | 0.8 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0638 | -0.0697 | 16 |
| 13 | 0 | 0.3 | 0.9 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0638 | -0.0697 | 16 |
| 14 | 0 | 0.4 | 0.1 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 24 |
| 15 | 0 | 0.4 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 16 | 0 | 0.4 | 0.5 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0638 | -0.0697 | 17 |
| 17 | 0 | 0.4 | 0.6 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0701 | -0.0770 | 16 |
| 18 | 0 | 0.4 | 0.7 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0638 | -0.0697 | 15 |
| 19 | 0 | 0.4 | 0.8 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0638 | -0.0697 | 16 |
| 20 | 0 | 0.4 | 0.9 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0638 | -0.0697 | 15 |
| 21 | 0 | 0.5 | 0.1 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 24 |
| 22 | 0 | 0.5 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 23 | 0 | 0.5 | 0.5 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0638 | -0.0697 | 17 |
| 24 | 0 | 0.5 | 0.6 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0701 | -0.0770 | 16 |
| 25 | 0 | 0.5 | 0.7 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0633 | -0.0700 | 15 |
| 26 | 0 | 0.5 | 0.8 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0638 | -0.0697 | 15 |
| 27 | 0 | 0.5 | 0.9 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0638 | -0.0697 | 16 |
| 28 | 0 | 0.6 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 29 | 0 | 0.6 | 0.6 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0641 | -0.0709 | 17 |
| 30 | 0 | 0.6 | 0.7 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | -0.0638 | -0.0697 | 17 |
| 31 | 0 | 0.3 | 0.2 | 0.2 | 0.2 | 0.7 | 0 | 0.1 | -0.0580 | -0.0577 | 12 |
| 32 | 0 | 0.3 | 0.2 | 0.4 | 0.2 | 0.7 | 0 | 0.1 | -0.0266 | -0.0309 | 23 |
| 33 | 0 | 0.3 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 34 | 0 | 0.3 | 0.2 | 0.9 | 0.2 | 0.7 | 0 | 0.1 | -0.0802 | -0.0816 | 16 |
| 35 | 0 | 0.4 | 0.2 | 0.5 | 0.2 | 0.7 | 0 | 0.1 | -0.0262 | -0.0297 | 27 |
| 36 | 0 | 0.4 | 0.2 | 0.6 | 0.2 | 0.7 | 0 | 0.1 | 0.4488 | 0.4522 | 36 |
| 37 | 0 | 0.4 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 38 | 0 | 0.4 | 0.2 | 0.9 | 0.2 | 0.7 | 0 | 0.1 | -0.0724 | -0.0826 | 17 |
| 39 | 0 | 0.4 | 0.2 | 0.65 | 0.2 | 0.7 | 0 | 0.1 | 0.5937 | 0.5943 | 66 |
| 40 | 0 | 0.4 | 0.2 | 0.77 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 25 |
| 41 | 0 | 0.4 | 0.2 | 0.78 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 24 |
| 42 | 0 | 0.4 | 0.2 | 0.79 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 23 |
| 43 | 0 | 0.5 | 0.2 | 0.5 | 0.2 | 0.7 | 0 | 0.1 | -0.0262 | -0.0297 | 27 |
| 44 | 0 | 0.5 | 0.2 | 0.6 | 0.2 | 0.7 | 0 | 0.1 | 0.4488 | 0.4522 | 37 |
| 45 | 0 | 0.5 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 46 | 0 | 0.5 | 0.2 | 0.9 | 0.2 | 0.7 | 0 | 0.1 | -0.0792 | -0.0865 | 16 |
| 47 | 0 | 0.5 | 0.2 | 0.65 | 0.2 | 0.7 | 0 | 0.1 | 0.5937 | 0.5943 | 71 |

| Uji coba | Transition Matrix | | | | | | | | Corr. Coef. (CC) | | Iterations |
|----------|-------------------|------------|------------|------------|------------|------------|------------|------------|------------------|---------|------------|
| | π_{11} | π_{22} | π_{33} | π_{44} | π_{55} | π_{66} | π_{13} | π_{63} | Viterbi | B-Weich | |
| 48 | 0 | 0.5 | 0.2 | 0.77 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 25 |
| 49 | 0 | 0.5 | 0.2 | 0.79 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 23 |
| 50 | 0 | 0.6 | 0.2 | 0.5 | 0.2 | 0.7 | 0 | 0.1 | -0.0262 | -0.0297 | 27 |
| 51 | 0 | 0.6 | 0.2 | 0.6 | 0.2 | 0.7 | 0 | 0.1 | 0.4664 | 0.4624 | 32 |
| 52 | 0 | 0.6 | 0.2 | 0.8 | 0.2 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 21 |
| 53 | 0 | 0.6 | 0.2 | 0.9 | 0.2 | 0.7 | 0 | 0.1 | -0.0724 | -0.0826 | 17 |
| 54 | 0 | 0.6 | 0.2 | 0.65 | 0.2 | 0.7 | 0 | 0.1 | 0.3740 | 0.3608 | 22 |
| 55 | 0 | 0.4 | 0.2 | 0.8 | 0.1 | 0.7 | 0 | 0.1 | 0.6713 | 0.6643 | 23 |
| 56 | 0 | 0.4 | 0.2 | 0.8 | 0.3 | 0.7 | 0 | 0.1 | 0.6733 | 0.6660 | 19 |
| 57 | 0 | 0.4 | 0.2 | 0.8 | 0.4 | 0.7 | 0 | 0.1 | 0.6731 | 0.6660 | 22 |
| 58 | 0 | 0.4 | 0.2 | 0.8 | 0.5 | 0.7 | 0 | 0.1 | 0.6735 | 0.6661 | 29 |
| 59 | 0 | 0.4 | 0.2 | 0.8 | 0.6 | 0.7 | 0 | 0.1 | -0.0642 | -0.0729 | 20 |
| 60 | 0 | 0.4 | 0.2 | 0.8 | 0.55 | 0.7 | 0 | 0.1 | 0.6735 | 0.6661 | 31 |
| 61 | 0 | 0.4 | 0.2 | 0.8 | 0.56 | 0.7 | 0 | 0.1 | 0.6735 | 0.6661 | 31 |
| 62 | 0 | 0.4 | 0.2 | 0.8 | 0.57 | 0.7 | 0 | 0.1 | 0.6735 | 0.6661 | 31 |
| 63 | 0 | 0.4 | 0.2 | 0.8 | 0.58 | 0.7 | 0 | 0.1 | 0.6735 | 0.6661 | 31 |
| 64 | 0 | 0.4 | 0.2 | 0.8 | 0.59 | 0.7 | 0 | 0.1 | -0.0568 | -0.0639 | 20 |
| 65 | 0 | 0.2 | 0.1 | 0.9 | 0.1 | 0.8 | 0.1 | 0.1 | 0.6731 | 0.6712 | 18 |
| 66 | 0 | 0.3 | 0.1 | 0.9 | 0.1 | 0.8 | 0.1 | 0.1 | 0.6731 | 0.6643 | 18 |
| 67 | 0 | 0.4 | 0.1 | 0.9 | 0.1 | 0.8 | 0.1 | 0.1 | 0.6731 | 0.6643 | 18 |
| 68 | 0 | 0.5 | 0.1 | 0.9 | 0.1 | 0.8 | 0.1 | 0.1 | 0.6731 | 0.6643 | 18 |
| 69 | 0 | 0.9 | 0.1 | 0.9 | 0.1 | 0.8 | 0.1 | 0.1 | 0.6823 | 0.6762 | 22 |
| 70 | 0 | 0.9 | 0.1 | 0.9 | 0.1 | 0.75 | 0.1 | 0.15 | 0.6800 | 0.6730 | 18 |
| 71 | 0 | 0.9 | 0.1 | 0.9 | 0.1 | 0.85 | 0.1 | 0.05 | 0.6778 | 0.6743 | 21 |
| 72 | 0 | 0.9 | 0.2 | 0.9 | 0.2 | 0.8 | 0.1 | 0.1 | 0.6854 | 0.6841 | 33 |
| 73 | 0 | 0.9 | 0.3 | 0.9 | 0.3 | 0.8 | 0.1 | 0.1 | 0.6854 | 0.6841 | 34 |
| 74 | 0 | 0.9 | 0.4 | 0.9 | 0.4 | 0.8 | 0.1 | 0.1 | 0.6829 | 0.6853 | 20 |
| 75 | 0 | 0.9 | 0.5 | 0.9 | 0.5 | 0.8 | 0.1 | 0.1 | 0.6829 | 0.6853 | 22 |
| 76 | 0 | 0.9 | 0.6 | 0.9 | 0.6 | 0.8 | 0.1 | 0.1 | 0.6821 | 0.6828 | 19 |
| 77 | 0 | 0.9 | 0.7 | 0.9 | 0.7 | 0.8 | 0.1 | 0.1 | 0.6851 | 0.6936 | 19 |
| 78 | 0 | 0.9 | 0.8 | 0.9 | 0.8 | 0.8 | 0.1 | 0.1 | 0.6850 | 0.6921 | 19 |
| 79 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.8 | 0.1 | 0.1 | 0.6850 | 0.6921 | 19 |
| 80 | 0 | 0.9 | 0.2 | 0.9 | 0.2 | 0.75 | 0.1 | 0.15 | 0.6854 | 0.6841 | 33 |
| 81 | 0 | 0.9 | 0.2 | 0.9 | 0.2 | 0.85 | 0.1 | 0.05 | 0.6764 | 0.6748 | 25 |
| 82 | 0 | 0.9 | 0.3 | 0.9 | 0.3 | 0.75 | 0.1 | 0.15 | 0.6854 | 0.6841 | 32 |
| 83 | 0 | 0.9 | 0.3 | 0.9 | 0.3 | 0.85 | 0.1 | 0.05 | 0.6763 | 0.6758 | 25 |
| 84 | 0 | 0.9 | 0.4 | 0.9 | 0.4 | 0.75 | 0.1 | 0.15 | 0.6828 | 0.6851 | 17 |
| 85 | 0 | 0.9 | 0.4 | 0.9 | 0.4 | 0.85 | 0.1 | 0.05 | 0.6707 | 0.6716 | 15 |
| 86 | 0 | 0.9 | 0.5 | 0.9 | 0.5 | 0.75 | 0.1 | 0.15 | 0.6821 | 0.6828 | 19 |
| 87 | 0 | 0.9 | 0.5 | 0.9 | 0.5 | 0.85 | 0.1 | 0.05 | 0.6733 | 0.6737 | 16 |
| 88 | 0 | 0.9 | 0.6 | 0.9 | 0.6 | 0.75 | 0.1 | 0.15 | 0.6851 | 0.6936 | 20 |
| 89 | 0 | 0.9 | 0.6 | 0.9 | 0.6 | 0.85 | 0.1 | 0.05 | 0.6733 | 0.6736 | 16 |
| 90 | 0 | 0.9 | 0.7 | 0.9 | 0.7 | 0.75 | 0.1 | 0.15 | 0.6850 | 0.6921 | 20 |
| 91 | 0 | 0.9 | 0.7 | 0.9 | 0.7 | 0.85 | 0.1 | 0.05 | 0.6870 | 0.6905 | 21 |
| 92 | 0 | 0.9 | 0.8 | 0.9 | 0.8 | 0.75 | 0.1 | 0.15 | 0.6850 | 0.6921 | 19 |
| 93 | 0 | 0.9 | 0.8 | 0.9 | 0.8 | 0.85 | 0.1 | 0.05 | 0.6850 | 0.6921 | 18 |
| 94 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.75 | 0.1 | 0.15 | 0.6866 | 0.6871 | 19 |
| 95 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.85 | 0.1 | 0.05 | 0.6850 | 0.6921 | 16 |

LAMPIRAN 5 Hasil Ujicoba Struktur Dasar Model 9 State.

| Uji coba | Transition Matrix | | | | | | | | | | Contr. Coeff.(CC) | | | Iterations | |
|----------|-------------------|------------------|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|-------------------|------------------|---------|------------|---------|
| | tr ₁₁ | tr ₁₂ | T ₁₃ | tr ₄₄ | tr ₅₅ | tr ₆₆ | tr ₇₇ | tr ₈₈ | tr ₉₉ | tr ₁₀ | tr ₁₃ | tr ₁₃ | Viterbi | | B-Welch |
| 1 | 0 | 0.3 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.74 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 2 | 0 | 0.3 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 3 | 0 | 0.9 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.0451 | 0.0437 | 27 |
| 4 | 0 | 0.8 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.6787 | 0.6755 | 21 |
| 5 | 0 | 0.34 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 6 | 0 | 0.33 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 7 | 0 | 0.32 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 8 | 0 | 0.31 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 9 | 0 | 0.29 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 10 | 0 | 0.3 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.73 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 11 | 0 | 0.3 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.72 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 22 |
| 12 | 0 | 0.3 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.71 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 22 |
| 13 | 0 | 0.81 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.6744 | 0.6756 | 12 |
| 14 | 0 | 0.79 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.6791 | 0.6679 | 15 |
| 15 | 0 | 0.78 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.6791 | 0.6679 | 16 |
| 16 | 0 | 0.77 | 0.2 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.6791 | 0.6679 | 19 |
| 17 | 0 | 0.3 | 0.22 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 18 | 0 | 0.3 | 0.21 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 19 | 0 | 0.3 | 0.19 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 20 | 0 | 0.3 | 0.18 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 21 | 0 | 0.3 | 0.17 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 22 | 0 | 0.3 | 0.16 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 23 | 0 | 0.3 | 0.15 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 24 | 0 | 0.3 | 0.1 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 25 | 0 | 0.3 | 0.1 | 0.2 | 0.8 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 26 | 0 | 0.3 | 0.1 | 0.1 | 0.79 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 27 | 0 | 0.3 | 0.1 | 0.1 | 0.78 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 20 |
| 28 | 0 | 0.3 | 0.1 | 0.1 | 0.77 | 0.1 | 0.1 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 21 |
| 29 | 0 | 0.3 | 0.1 | 0.1 | 0.78 | 0.09 | 0.09 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7242 | 0.7124 | 20 |
| 30 | 0 | 0.3 | 0.1 | 0.1 | 0.78 | 0.2 | 0.2 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7258 | 0.7120 | 24 |
| 31 | 0 | 0.3 | 0.1 | 0.1 | 0.8 | 0.15 | 0.15 | 0.7 | 1 | 0 | 0 | 0.1 | 0.7258 | 0.7120 | 25 |
| 32 | 0 | 0.3 | 0.1 | 0.1 | 0.8 | 0.15 | 0.15 | 0.74 | 1 | 0 | 0 | 0.1 | 0.7258 | 0.7120 | 22 |
| 33 | 0 | 0.3 | 0.1 | 0.1 | 0.8 | 0.15 | 0.15 | 0.8 | 1 | 0 | 0 | 0.1 | 0.7258 | 0.7120 | 21 |

| Uji coba | Transition Matrix | | | | | | | | | | Corr. Coeff.(CC) | | | Iterations |
|----------|-------------------|------|------|------|------|------|------|------|------|------|------------------|----------|---------|------------|
| | tr11 | tr12 | tr13 | tr14 | tr15 | tr16 | tr17 | tr18 | tr19 | tr13 | tr13 | V(terbl) | B-Weich | |
| 34 | 0 | 0.3 | 0.1 | 0.1 | 0.8 | 0.2 | 0.2 | 0.8 | 1 | 0 | 0.1 | 0.7258 | 0.712 | 21 |
| 35 | 0 | 0.3 | 0.1 | 0.1 | 0.8 | 0.2 | 0.2 | 0.85 | 1 | 0 | 0.1 | 0.7251 | 0.7114 | 20 |
| 36 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.85 | 1 | 0 | 0.1 | 0.7055 | 0.6975 | 16 |
| 37 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.89 | 1 | 0 | 0.1 | 0.6887 | 0.6637 | 20 |
| 38 | 0 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.89 | 1 | 0 | 0.1 | 0.3872 | 0.3987 | 40 |
| 39 | 0 | 0.3 | 0.1 | 0.1 | 0.8 | 0.2 | 0.2 | 0.8 | 1 | 0.1 | 0.1 | 0.7273 | 0.7134 | 17 |
| 40 | 0 | 0.3 | 0.1 | 0.1 | 0.8 | 0.2 | 0.2 | 0.85 | 1 | 0.1 | 0.1 | 0.7273 | 0.7134 | 20 |
| 41 | 0 | 0.3 | 0.1 | 0.1 | 0.79 | 0.1 | 0.1 | 0.7 | 1 | 0.1 | 0.1 | 0.7242 | 0.7124 | 21 |
| 42 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.8 | 1 | 0.1 | 0.1 | 0.7289 | 0.7166 | 16 |
| 43 | 0 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.85 | 1 | 0.1 | 0.1 | 0.7193 | 0.7161 | 14 |
| 44 | 0 | 0.8 | 0.5 | 0.5 | 0.7 | 0.5 | 0.5 | 0.8 | 1 | 0.1 | 0.15 | 0.3557 | 0.3805 | 27 |
| 45 | 0 | 0.9 | 0.1 | 0.1 | 0.9 | 0.1 | 0.1 | 0.8 | 1 | 0.1 | 0.1 | 0.6958 | 0.6984 | 15 |
| 46 | 0 | 0.9 | 0.1 | 0.1 | 0.9 | 0.1 | 0.1 | 0.85 | 1 | 0.1 | 0.1 | 0.6993 | 0.7001 | 23 |



UCAPAN TERIMA KASIH

Dengan mengucapkan syukur Alhamdulillah atas karunia dan rahmat yang diberikan oleh Allah SWT kepada saya dan keluarga sehingga saya dapat menyelesaikan program Doktor pada Program Studi Teknik Elektro Fakultas Teknik Universitas Indonesia.

Sebagai rasa terima kasih dan hormat saya sampaikan kepada :

1. Bapak Prof. Dr. Ir. Sar Sardy, selaku Promotor Doktor yang telah memberikan banyak waktunya untuk memberikan petunjuk dan pengarahan serta saran dalam penulisan penelitian ini.
2. Bapak Prof. Dr. Ir. Dadang Gunawan M. Eng, selaku Ko Promotor yang telah banyak memberi sumbangan pikiran dan bahan-bahan untuk keperluan tugas penelitian ini.
3. Bapak Prof. Dr. Ir. Eko Tjipto Raharjo, selaku Ketua Departemen Teknik Elektro Universitas Indonesia, yang telah memberi masukan dan bantuan dalam bidang akademik.
4. dr. Herawati Sudoyo, PhD selaku Manager Eksekutif, dr. Helena MS, selaku Peneliti Senior dan Hidayat Trimarsanto, B.Sc selaku Asisten Peneliti Bioinformatik dari Lembaga Biologi Molekuler Eijkman dalam lingkup Menristek Jakarta yang telah bersedia memberikan izin dan konsultasi selama penelitian dilakukan.
5. Dekan Fakultas Teknologi Industri Usakti dan Kajar Teknik Elektro FTI Usakti yang telah memberikan kesempatan kepada Penulis untuk menempuh studi S3 di Universitas Indonesia.
6. Bapak Dr. Ir. Tjandra Susila M Eng Sc, selaku Ketua Konsentrasi Telekomunikasi Jurusan Teknik Elektro, Fakultas Teknologi Industri Universitas Trisakti.
7. Bapak Ir. Tony Winata, selaku Kepala Laboratorium Telekomunikasi, Jurusan Teknik Elektro Universitas Trisakti.
8. dr. Djaja S. Atmadja, Sp.F. Ph D. SH. DFM, selaku Staf Pengajar, ahli DNA Bagian Kedokteran Forensik, Fakultas Kedokteran Universitas Indonesia.
9. Dosen-dosen staf pengajar Departemen Teknik Elektro, Fakultas Teknik Universitas Indonesia atas dukungannya.

10. Fendy Sutandio, S. Kom, S Si, Alumni Fakultas Matematika dan Ilmu Pengetahuan Alam, Fakultas Ilmu Komputer, Universitas Bina Nusantara.
11. dr. Sheirly Novan Indra, Alumni Fakultas Kedokteran Universitas Indonesia.
12. Bapak Purwowibowo, yang telah memberi semangat dan dukungan serta bantuan selama penelitian dan penulisan disertasi ini.
13. Semua Karyawan dan Karyawati Fakultas Teknik Universitas Indonesia yang telah membantu dalam penyelesaian Disertasi ini.
14. Rekan-rekan kerja di Jurusan Teknik Elektro FTI Usakti khususnya di Laboratorium Telekomunikasi Jurusan Teknik Elektro FTI Usakti atas bantuan dan dukungan pada pelaksanaan penelitian ini.
15. Secara khusus saya sampaikan terima kasih yang tak terhingga kepada Ibunda tercinta yang telah membesarkan, mendidik dan mengajarkan saya hingga saya bisa mencapai pendidikan ini. Begitu pula kepada Kakak-Kakak, Adik-Adik tercinta dan seluruh Keluarga yang dengan penuh perhatian dan pengertian telah membantu saya dalam menyelesaikan disertasi ini.
16. Ucapan terima kasih dan sayang yang tak terhingga saya sampaikan kepada Suami tercinta, Anak-anak saya Dony Agoesfiansyah dan Nila Fitrianty, serta Reddy Ariyuansyah atas semua bantuan, dukungan, pengertian, sehingga semua tugas dan kewajiban saya di rumah dapat diselesaikan selama saya menyelesaikan disertasi ini.

Akhir kata saya mengucapkan banyak terima kasih kepada semua pihak yang telah memberikan dukungan dan bantuannya sehingga tulisan disertasi ini dapat diselesaikan. Semoga Allah SWT membalas semua kebajikannya. Amin.

Depok, Maret 2008

Suhartati Agoes

BIODATA

Nama : Suhartati Agoes
Tempat/Tanggal Lahir : Jakarta/12-02-1953
Kebangsaan : Indonesia
Alamat : - Kantor :
Lab. Telekomunikasi Jur. Teknik Elektro FTI Usakti
Jl. Kyai Tapa Grogol Jakarta 11440
Phone: (+62)(21)5663232 Pes.429
Fax : (+62)(21)5605831
Email: suhartati_agoes@yahoo.com
- Rumah :
Pondok Kelapa Indah Blok A/18, Kalimalang
Jakarta 13450

PENDIDIKAN

1998 S2 : Magister Bidang Ilmu Teknik
Departemen Teknik Elektro
Universitas Indonesia
1979 S1 : Jurusan Teknik Elektro
Fakultas Teknik
Universitas Trisakti

RISET/PENELITIAN

- Suhartati Agoes, et al, Penyesuaian berbagai saluran transmisi frekuensi ultra tinggi. Laporan riset Jurusan Teknik Elektro, Fakultas Teknologi Industri Universitas Trisakti, 1988.
- Suhartati Agoes, et al, Data Paket Radio. Laporan riset Jurusan Teknik Elektro, Fakultas Teknologi Industri Universitas Trisakti, 1990.
- Suhartati Agoes, et al, Pengaruh cuaca terhadap gelombang mikro (Propagation Experiment-via Terrestrial). Laporan riset Jurusan Teknik Elektro FTI Usakti bekerja sama dengan LAPAN, 1985.
- Suhartati Agoes, et al, Pengaruh cuaca terhadap gelombang mikro (Propagation Experiment-via Sattelite). Laporan riset Laboratorium Telekomunikasi Jurusan Teknik Elektro FTI Usakti bekerja sama dengan TU Eindhoven, 2000.

PUBLIKASI

Nasional

- Suhartati Agoes, Dadang Gunawan, Sar Sardy, "Proses Kode Genetik Pada Bidang Kompleks Menggunakan Filter Digital Finite Impulse Response", *Proceeding of The 6th Quality in Research*, Fakultas Teknik Universitas Indonesia, October 2003, ISSN : 1411-1284
- Suhartati Agoes, Dadang Gunawan, Sar Sardy and Hoedjo, "Exon Prediction on DNA-Genes *Plasmodium falciparum* Based on Coding Sequence Structure by Using Hidden Markov Model", *Jurnal Universa Medicina*, Jurnal Kedokteran Trisakti, Vol. 26 No 3, Juli-September 2007, No. ISSN : 1907- 3062, Akreditasi Dirjen Dikti No. SK 39 / Dikti / Kep / 2004.
- Suhartati Agoes, Dadang Gunawan, Sar Sardy, "The implementation of hidden Markov model for exon prediction on DNA-gen *Plasmodium falciparum* ", *Jurnal Kedokteran Yarsi*, Jakarta November 2007, Terakreditasi Dirjen Dikti. (*Accepted untuk di publikasikan*)

Internasional

- Suhartati Agoes, Dadang Gunawan, Sar Sardy, "Correlation Coefficient Analysis of Exon region in DNA of Genes *Plasmodium falciparum* Using Hidden Markov Model", *Proceeding of BME DAYS 2006*, Institut Teknologi Bandung, November 2006.
- Suhartati Agoes, Dadang Gunawan, Sar Sardy, "Exon Prediction on DNA-Gen *Plasmodium falciparum* by Using Hidden Markov Model", *Proceeding of 4th International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2007*, Orlando, Florida, USA, Juli 2007.