

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Informasi genetika sangat diperlukan oleh hampir semua bidang keilmuan, terutama biologi dan kedokteran. Ditemukannya insulin hasil rekayasa genetika merupakan contoh paling shahih dari pemanfaatan informasi tersebut. Dalam biologi sendiri sistem taksonomi analisis filogenetika merupakan salah satu contoh pemanfaatan informasi genetika yang akhir – akhir ini terus berkembang. Para ahli biologi secara tradisional menggambarkan silsilah atau genealogi organisme sebagai pohon filogenetik (*phylogenetic tree*) [2].

Kemudahan dalam mengumpulkan data genetika menyebabkan basis data genetika semakin meningkat. Peningkatan jumlah data yang pesat ini menyebabkan analisa data secara manual tidak lagi efisien. Dibutuhkan komputasi untuk membantu analisa data sehingga dapat mengekstrak suatu informasi penting dari sekumpulan data genetika yang berjumlah besar. [1]

Dalam suatu makhluk hidup dikenal suatu substansi protein terpenting yang disebut dengan *Deoxyribonucleic Acid* (DNA) atau asam Deoksiribonukleat. Substansi penting ini memiliki gugus (bagian dari substansi) protein penyusun yaitu basa nitrogen yang terdiri dari *Adenine* (A), *Guanine* (G), *Chytosine* (C), dan *Thymine* (T). Dalam kenyataannya struktur penyusun DNA ini selalu muncul dalam triplet atau barisan dari tiga macam basa nitrogen penyusun DNA. Banyak eksperimen yang dilakukan untuk menentukan barisan basa nitrogen penyusun DNA yang dilakukan dalam banyak hal. Diantaranya untuk memodelkan struktur protein pada perkembangbiakan makhluk hidup, penemuan gen (faktor penurunan sifat makhluk hidup), analisis filogenetika, pemodelan daerah DNA, sub-family (bagian dari suatu makhluk hidup yang diduga merupakan satu jenis) protein dan lain-lain. [10]

Salah satu metode yang beberapa tahun belakangan ini banyak dikembangkan adalah Model Markov Tersembunyi (MMT). MMT merupakan aplikasi dari perkembangan rantai Markov (Andrei A. Markov, 1906) yaitu suatu

teknik peluang yang menganalisis pergerakan peluang dari suatu kondisi ke kondisi lainnya. MMT ini pada dasarnya sudah diperkenalkan pada akhir 1960-an (Baum dan Petrie 1966, Baum dan Eagon 1967, dan Baum dan Sell 1968). Pada awalnya MMT dipakai dalam bidang *Speech Recognition* (Gupta, 2004). Kemudian MMT mulai digunakan dalam bioinformatika sejak diperkenalkan oleh Krogh et al. (Gupta,2004). MMT adalah model statistika, dimana sistem yang dimodelkan diasumsikan sebagai proses Markov dengan parameter yang tidak diketahui (Koski,2001). Tantangan dalam metode ini adalah menentukan parameter – parameter tersembunyi yang tidak diketahui dari keluaran yang dihasilkan oleh model tersebut. Suatu proses disebut proses Markov apabila memiliki proverti Markov yaitu distribusi probabilitas *state* berikutnya ditentukan oleh *state* saat ini dimana probabilitas *state* saat ini ditentukan oleh *state* sebelumnya [1].

Berdasarkan penjelasan diatas maka penulis ingin mengangkat tema tersebut sebagai tugas akhir dengan judul “Penerapan Model Markov Tersembunyi untuk Mengetahui Persentase Kecocokan dari *Deoxyribonucleic Acid* pada Pohon Filogenetik *Ursidae* (Beruang)”.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka permasalahan yang akan dibahas dalam tugas akhir ini adalah :

1. Bagaimana penerapan MMT pada pohon filogenetik *ursidae* dalam menganalisis kecocokan DNA antar spesiesnya.
2. Bagaimana persentase kecocokan DNA antar spesies menggunakan NCBI dengan nilai kecocokan DNA antar spesies menggunakan MMT.

1.3. Batasan Masalah

Pada penulisan ini, penulis membatasi permasalahan dalam beberapa hal sebagai berikut :

1. Menganalisis DNA dari topologi pohon filogenetik *family ursidae* yaitu, *Ursus arctos mitochondrion genome* (beruang coklat), *Ursus martimus mitochondrion genome* (beruang Kutub), *Ursus thibetanus mitochondrion genome* (Beruang Hitam Asia) dan *Ursus Americanus mitochondrion*

genome, (beruang hitam Amerika) untuk diketahui hubungan kekerabatannya.

2. Tiga asumsi MMT terpenuhi pada kasus DNA, yaitu:
 - Asumsi Markov
 - Asumsi stasioneritas
 - Asumsi independensi
3. Pada penelitian tugas akhir ini hanya membandingkan tingkat kecocokan berdasarkan NCBI dengan tingkat kecocokan yang dihasilkan oleh metode MMT dari λ awal dan $\hat{\lambda}$ optimum yang diperoleh dari proses algoritma Baum-Welch.

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai dalam tugas akhir ini adalah :

1. Mengetahui bagaimana penerapan MMT pada pohon filogenetik *ursidae* dalam menganalisis kecocokan DNA antar spesiesnya, sehingga dapat diketahui hubungan kekerabatannya.
2. Mengetahui perbedaan persentase kecocokan DNA antar spesies menggunakan NCBI dengan persentase kecocokan DNA antar spesies menggunakan MMT.

Adapun manfaat yang ingin dicapai dalam tugas akhir ini adalah menambah wawasan terutama tentang penerapan model Markov tersembunyi pada bidang biomolekular.

1.5. Metodologi Penelitian

Tahapan yang digunakan dalam tugas akhir ini agar mencapai tujuan adalah sebagai berikut:

1. Mengetahui dan memahami tentang sistematika dan filogenetika molekuler.
2. Menganalisis 4 spesies yang berasal dari *family ursidae*, dengan menggunakan data DNA yang diperoleh dari program *The National Center for Biotechnology Information* (NCBI) . Kemudian diolah dengan menentukan model MMT dan menghitung parameter – parameter baru yang optimum dengan bantuan program MATLAB.

1.6. Sistematika Penulisan

Sistematika penulisan skripsi ini berguna untuk memudahkan pembaca dalam memahami skripsi ini. Adapun sistematikanya sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini dijelaskan latar belakang, rumusan masalah, batasan masalah, tujuan dan kegunaan penelitian, metode penelitian, dan sistematika penulisan

BAB II Landasan Teori

Pada bab ini dijelaskan teori tentang *Deoxyribonucleic Acid* (DNA), filogenetik dan sistematika, proses stokastik, rantai Markov, dan Model Markov Tersembunyi (MMT).

BAB III Kajian Utama Penelitian

Pada bab ini dijelaskan tahapan utama penelitian, parameter MMT, Uji Asumsi MMT pada Kasus DNA, Kompresi barisan nukleotida, Masalah – masalah dalam MMT.

BAB IV

Pada bab ini dijelaskan Kompresi barisan nukleotida, Parameter **A**, **B**, dan π Optimum, Barisan Keadaan Tersembunyi.

BAB V

Kesimpulan dan saran.