

Penerapan Model Markov Tersembunyi untuk Mengetahui Persentase Kecocokan dari Deoxyribonucleic Acid pada Pohon Filogenetik Ursidae (Beruang)

RINI CAHYANDARI¹, RENI NURSOLIAH²

^{1,2}Jurusan Matematika, Fakultas Sains dan Teknologi UIN Sunan Gunung Djati Bandung
Email: rcahyandari@yahoo.com, reninursoleha@gmail.com

ABSTRAK

Jumlah data biologi molekuler yang semakin meningkat pasca genome project membutuhkan pengelompokan data ke dalam suatu kelompok subfamili berdasarkan tingkat kesamaan data tersebut. Pengelompokan dan penentuan subfamili berdasarkan kumpulan sequence DNA merupakan salah satu cara yang penting dalam bidang ini. Cara ini hampir dikatakan tidak bisa dilakukan secara manual sehingga membutuhkan alat bantu komputasi, di mana Model Markov Tersembunyi (MMT) merupakan salah satu metode komputasi yang dapat membantu dalam menganalisis kemiripan dari topologi pohon filogenetik, salah satunya adalah spesies ursidae (Beruang) dengan membandingkan sequence antar DNA. Beberapa Algoritma dilibatkan untuk memecahkan permasalahan dalam MMT, yaitu Algoritma maju mundur, algoritma Baum – Welch, dan Algoritma Viterbi. Hasil proses pengujian MMT pada keempat DNA dari family ursidae (Beruang) dapat disimpulkan bahwa keempat spesies tersebut berkerabat dekat, karena memiliki persentase kecocokan yang cukup besar.

Kata Kunci: Deoxyribonucleic Acid (DNA), Pohon Filogenetik, Model Markov Tersembunyi (MMT), Spesies ursidae.

1. PENDAHULUAN

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*) (Herlina, 2006).

Dalam suatu makhluk hidup dikenal suatu substansi protein terpenting yang disebut dengan *Deoxyribonucleic Acid* (DNA) atau asam Deoksiribonukleat. Banyak eksperimen yang dilakukan untuk menentukan barisan basa nitrogen penyusun DNA, 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 (Rabbiner dan Lawrence, 1989).

Salah satu metode yang beberapa tahun belakangan ini banyak dikembangkan adalah Model Markov Tersembunyi (MMT). Model ini 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, di mana sistem yang dimodelkan diasumsikan sebagai proses Markov dengan parameter yang tidak diketahui (Koski, 2001). Suatu proses disebut proses Markov apabila memiliki sifat Markov yaitu distribusi probabilitas *state* berikutnya ditentukan oleh

state saat ini dimana probabilitas *state* saat ini ditentukan oleh *state* sebelumnya (Afihayati dan Mulyana, 2008).

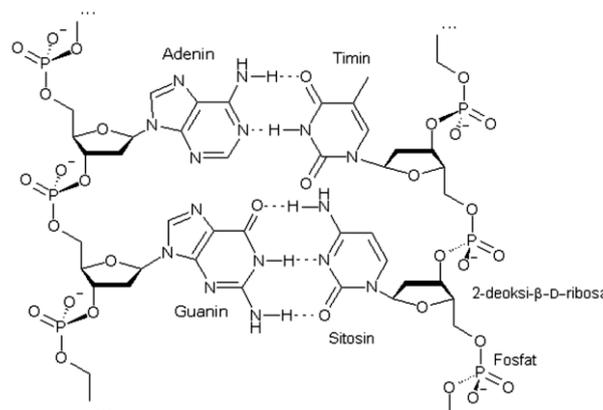
Pada makalah ini akan dilakukan kajian tentang penerapan model Markov tersembunyi (MMT) untuk mengetahui persentase kecocokan dari *deoxyribonucleic acid* pada pohon filogenetik *ursidae*(beruang).

2. DEOXYRIBONUCLEIC ACID (DNA)

Deoxyribonucleic Acid (DNA) atau asam deoksiribonukleat adalah jenis asam nukleat yang tergolong biomolekul utama penyusun berat kering (biomasa) suatu organisme. Struktur untai komplementer DNA menunjukkan pasangan basa (*adenine* dengan *thymine* dan *guanine* dengan *cytosine*) yang membentuk DNA beruntai ganda.

DNA merupakan polimer yang terdiri dari tiga komponen utama (Mukherjee, dan Mitra, 2005):

- 1) gugus fosfat
- 2) gula deoksiribosa
- 3) basa nitrogen, yang terdiri dari:
 - a. Purin : i. *Adenine* (A)
ii. *Guanine* (G)
 - b. Pirimidin : i. *Cytosine* (C)
ii. *Thymine* (T)



Gambar 1. Struktur dasar DNA

Pada penelitian DNA terdapat metode DNA *sequence alignment* atau penyejajaran DNA yang dilakukan untuk meneliti kecocokan DNA baru dengan DNA sebelumnya. Penyejajaran DNA ini juga dilakukan untuk melihat kecocokan DNA dari beberapa spesies yang berbed. Dalam penyejajaran DNA ini jika suatu basa nitrogen penyusun DNA suatu spesies sama dengan basa nitrogen penyusun DNA dari spesies lain maka dikatakan cocok, jika berbeda maka dikatakan sebagai insersi atau sisipan, dan jika terdapat suatu basa nitrogen yang hilang maka dikatakan sebagai delesi atau hapusan (Ibe, 2009).

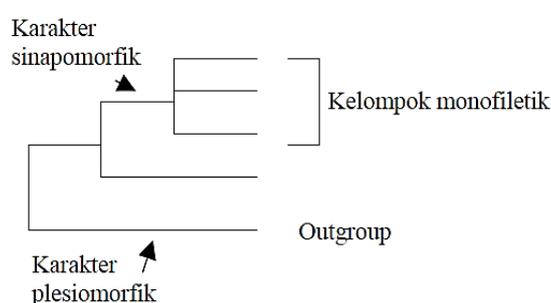
Salah satu peran matematika dalam proses rekayasa genetika ini adalah Model Markov Tersembunyi (MMT) yang digunakan dalam meneliti barisan basa nitrogen yang terdapat dalam DNA. Terutama dalam DNA *sequence alignment*, yaitu proses pencocokan DNA yang diteliti dengan DNA lain yang tersip sebelumnya (NCBI, 2005).

3. SISTEMATIKA DAN FILOGENETIKA MOLEKULER

Sistematika bertujuan untuk memahami dan mendeskripsikan keanekaragaman suatu organisme dan merekonstruksi hubungan kekerabatannya terhadap organisme lainnya, dan juga mendokumentasikan perubahan-perubahan yang terjadi selama evolusinya dan merubahnya ke dalam sebuah sistem klasifikasi yang mencerminkan evolusinya tersebut.

Sampai saat ini ada dua pendekatan untuk merekonstruksi hubungan evolusi dari sebuah kelompok organisme biologi, yaitu fenetik dan kladistik yang sering disebut atau ditulis di dalam literatur ilmiah sebagai filogenetika dan merupakan pendekatan yang umum digunakan di dalam banyak penelitian sistematika.

Pada pendekatan filogenetika, sebuah kelompok organisme dimana anggota-anggotanya memiliki banyak kesamaan karakter atau ciri dianggap memiliki hubungan yang sangat dekat dan diperkirakan diturunkan dari satu nenek moyang. Nenek moyang dan semua turunannya akan membentuk sebuah kelompok monofiletik. Dalam analisis filogenetika kelompok *outgroup* sangat dibutuhkan dan menyebabkan polarisasi karakter atau ciri, yaitu karakter apomorfik dan plesiomorfik. Karakter apomorfik adalah karakter yang berubah dan diturunkan dan terdapat pada *ingroup*, sedangkan karakter plesiomorfik merupakan karakter *primitive* yang terdapat pada *outgroup*. Karakter sinapomorfik adalah karakter yang diturunkan dan terdapat pada kelompok monofiletik.



Gambar 2. Pohon kekerabatan dan polarisasi dalam analisis filogenetik

Penggunaan sekuen DNA dalam penelitian filogenetika telah meningkat pesat dan telah dilakukan pada semua tingkatan taksonomi, misalnya famili, marga, dan spesies. Filogenetika molekuler mengkombinasikan teknik biologi molekuler dengan statistik untuk merekonstruksi hubungan filogenetika. Pemikiran dasar penggunaan sekuen DNA dalam studi filogenetika adalah bahwa terjadi perubahan basa nukleotida menurut waktu, sehingga akan dapat diperkirakan kecepatan evolusi yang terjadi dan akan dapat direkonstruksi hubungan evolusi antara satu kelompok organisme dengan yang lainnya.

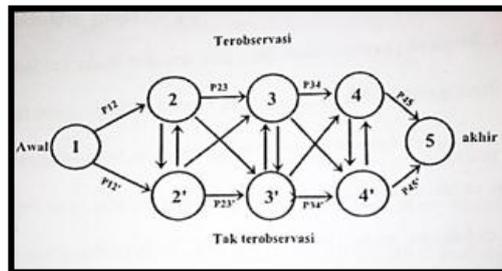
Tabel 1. Sistem gen dan genom yang telah digunakan dalam penelitian filogenetik

Gen atau genom	Metode
DNA Kloroplas	Analisis restriksi
Gen <i>rbcL</i> dari DNA kloroplas	Analisis <i>Sequence</i> DNA
Gen <i>matK</i> dari DNA kloroplas	Analisis <i>Sequence</i> DNA
DNA mitokondria	Analisis <i>Sequence</i> DNA
Gen RNA inti	Analisis <i>Sequence</i> DNA
Daerah ITS dari nrDNA	Analisis <i>Sequence</i> DNA
Kelompok gen repetitive:	
- Knob heterokromatin	Analisis <i>Sequence</i> DNA
- Gen CAB	Analisis <i>Sequence</i> DNA
- Gen <i>rbcS</i>	Analisis <i>Sequence</i> DNA
Gen kopi tunggal	Analisis <i>Sequence</i> DNA

4. MODEL MARKOV TERSEMBUNYI (MMT)

Pada tahun 1970 ahli matematika yaitu Baum dan Petrie memperkenalkan suatu teori pengembangan dari rantai Markov yang dipaparkan oleh Andrei A. Markov pada tahun 1856 yaitu Model Markov Tersembunyi (MMT) atau *Hidden Markov Model* (HMM) untuk menyelesaikan suatu kasus dimana suatu keadaan dalam kasus tersebut dipengaruhi oleh keadaan lain yang tidak terobservasi, selanjutnya keadaan yang tidak terobservasi ini disebut keadaan tersembunyi. MMT adalah gabungan dari dua proses stokastik yang salah satunya tidak dapat diobservasi secara langsung, namun tetap dapat diobservasi dengan cara menganalisis peluang dari salah satu proses stokastik yang dapat diobservasi (Rabbiner dan Lawrence, 1989).

Jika $X = \{X_1, X_2, \dots\}$ adalah sebuah rantai Markov, dan $O = \{O_1, O_2, \dots\}$ adalah sebuah fungsi dari X , dimana X dan O merupakan proses stokastik, maka X adalah sebuah MMT yang dapat diobservasi melalui O , atau dapat ditulis $O = f(X)$ untuk suatu fungsi f . parameter X menyatakan proses keadaan yang tersembunyi, sementara parameter O menyatakan proses keadaan yang dapat diobservasi.



Gambar 3. Ilustrasi MMT

Parameter – parameter pada MMT :

1. **N** menyatakan jumlah keadaan yang tersembunyi.
2. **M** menyatakan jumlah keadaan yang terobservasi.
3. Matriks peluang transisi **A**, berukuran $N \times N$, dimana elemen – elemen (a_{ij}) dari matriks ini menyatakan peluang transisi (pergerakan atau perpindahan) dari keadaan tersembunyi ke – i ke keadaan tersembunyi ke – j ($1 \leq i \leq N, 1 \leq j \leq N$), dimana untuk $\forall i$ dan $1 \leq i \leq N$ dipenuhi $\sum_{j=1}^N a_{ij} = 1$, yaitu sebagai berikut :

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix}$$

4. Matriks peluang emisi **B**, berukuran $N \times M$, dimana elemen – elemen (b_{ik}) dari matriks ini menyatakan peluang keadaan tersembunyi ke – i berada pada keadaan terobservasi ke – k ($1 \leq i \leq N, 1 \leq k \leq M$), dimana untuk $\forall i$ dan $1 \leq i \leq N$ dipenuhi $\sum_{k=1}^M b_{ik} = 1$, yaitu sebagai berikut :

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1M} \\ b_{21} & b_{22} & \dots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{N1} & b_{N2} & \dots & b_{NM} \end{bmatrix}$$

5. Matriks peluang keadaan awal / prior $\boldsymbol{\pi}$, berukuran $\mathbf{N} \times 1$, dimana elemen –elemen (π_i) menyatakan peluang awal dari keadaan tersembunyi ke i ($1 \leq i \leq \mathbf{N}$), dimana $\sum_{i=1}^{\mathbf{N}} \pi_i$ yaitu sebagai berikut:

$$\boldsymbol{\Pi} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix}$$

Masalah – masalah Utama dalam MMT dan Metode Penyelesaiannya:

1. Menghitung peluang observasi dengan algoritma maju mundur.

- A. Algoritma Maju didefinisikan $\alpha_t(i)$ sebagai variabel maju, dimana

$$\alpha_t(i) = P(O_1, O_2, \dots, O_t, X_t = i | \lambda)$$

dengan $\alpha_t(i)$ menyatakan total peluang observasi yang berakhir pada keadaan tersembunyi i pada saat t dimana $t = 1, 2, \dots, T$ jika diketahui suatu barisan observasi $\{O_1, O_2, \dots, O_t\}$. Menurut Rabiner (1989), secara umum algoritma maju terdiri atas tiga bagian, yaitu :

1. Tahap inisialisasi

$$\alpha_t(i) = \pi(i)b_t(O_i) \text{ dimana } 1 \leq i \leq \mathbf{N}$$

2. Tahap induksi

$$\alpha_{t+1}(j) = \left\{ \sum_{i=1}^{\mathbf{N}} \alpha_t(i)a_{ij} \right\} b_j(O_{t+1}) \quad j = 1, \dots, \mathbf{N}, t = 1, \dots, T - 1$$

3. Tahap terminasi

Pada tahap ini adalah menjumlahkan semua peluang gabungan dari observasi dan keadaan tersembunyi bila diketahui sebuah model sehingga diketahui peluang marginal dari observasi tersebut atau ditulis :

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

- B. Algoritma Mundur didefinisikan $\beta_t(i)$ sebagai variabel mundur, dimana

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

Tahap – tahap algoritma mundur dijelaskan sebagai berikut :

1. Tahap inisialisasi

$$\beta_t(i) = 1 \text{ untuk } i = 1, 2, \dots, \mathbf{N}$$

Pada tahap ini, dinyatakan $\beta_t(i) = 1$ karena diasumsikan i adalah keadaan terobservasi akhir, dan bernilai nol untuk i yang lainnya .

2. Tahap induksi

$$\beta_t(i) = \sum_{j=1}^{\mathbf{N}} b_j(O_{t+1}) \beta_{t+1}(j)a_{ij}$$

Untuk $t = T - 1, T - 2, \dots, 1$ dan $i = 1, 2, \dots, \mathbf{N}$

3. Tahap terminasi

$$P(O | \lambda) = \sum_{i=1}^{\mathbf{N}} b_i(O_1) \pi(i)\beta_1(i)$$

Algoritma maju maupun algoritma mundur akan menghasilkan peluang observasi yang bernilai sama.

2. Menentukan barisan keadaan tersembunyi dengan algoritma viterbi.

Algoritma Viterbi didefinisikan :

$$\delta_t(i) = \max_{x_1, x_2, \dots, x_{t-1}} P(O_1, O_2, \dots, O_t, X_1, X_2, \dots, X_{t-1}, X_t = i | \lambda)$$

$$\psi_t(j) = \arg \max_{i \leq 1 \leq \mathbf{N}} \{ \delta_{t-1}(i)a_{ij} \}$$

Variabel $\delta_t(i)$ menyatakan peluang terbesar sepanjang t observasi pertama dan berakhir pada keadaan tersembunyi i. Sementara $\psi_t(j)$ menyimpan keadaan tersembunyi sebelumnya yang akan membentuk barisan keadaan tersembunyi yang paling optimal.

Algoritma Viterbi terdiri atas empat tahap :

1. Tahap inisialisasi

Pada saat $t = 1$,

$$\delta_t(i) = P(X_1 = i, O_1) = P(O_1 | X_1 = i) P(X_1 = i)$$

dengan mensubstitusi asumsi awal pada MMT yaitu

$$b_i(k) = P(O_t = k | X_t = i) \text{ dan } \pi(i) = P(X_t = i)$$

diperoleh :

$$\delta_t(i) = b_i(O_1) \pi(i)$$

pada tahap ini $\psi_1(i) = 0$

2. Tahap rekursi

Pada tahap rekursi,

$$\delta_t(j) = b_j(O_t) \max_{1 \leq i \leq N} \{\delta_{t-1}(i) a_{ij}\}$$

$$\psi_1(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] , 1 \leq j \leq N$$

3. Tahap terminasi

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

$$X_T^* = \arg \max_{1 \leq i \leq N} \{\delta_T(i)\}$$

4. Tahap backtracking

$$X_t^* = \psi_{t+1}(X_{t+1}^*) , t = T - 1 , T - 2 , \dots , 1$$

Tahap backtracking memungkinkan barisan keadaan tersembunyi yang paling optimal ditemukan dari titik terakhir yang disimpan pada tahap rekursi.

3. Mengestimasi parameter dalam MMT dengan algoritma Baum - Welch.

Permasalahan ketiga dalam MMT berkaitan dengan mengestimasi 3 parameter pada MMT yaitu **A**, **B**, dan **π** sehingga terbentuk model baru $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$ dimana $P(O | \hat{\lambda}) \geq P(O | \lambda)$ (Nisa Pandu, 2011).

Dalam algoritma Baum - Welch didefinisikan variabel maju dan variabel mundur. Yang digunakan dalam algoritma maju dan algoritma mundur pada penyelesaian permasalahan MMT yang pertama yaitu menghitung peluang Observasi (Rabbiner dan Lawrence, 1989). Yaitu sebagai berikut

$$\begin{cases} \alpha_t(A) = P(O_1, O_2, \dots, O_t, X_t = i | \lambda) \\ \beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T, X_t = i | \lambda) \end{cases}$$

Selanjutnya didefinisikan sebuah variabel baru $\xi_t(i, j)$ dimana $\xi_t(i, j)$ adalah peluang proses berada pada keadaan tersembunyi j pada waktu t bila diketahui barisan observasi dari model

$$\xi_t(i, j) = P(X_t = i, X_{t+1} = j | O, \lambda)$$

Dengan menggunakan definisi variabel maju dan mundur, persamaan di atas dapat ditulis

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)}$$

Dengan diperoleh nilai $\xi_t(i, j)$, bisa dihitung peluang proses berada pada keadaan tersembunyi i pada waktu t , $\gamma_t(i)$ dengan menjumlahkan $\xi_t(i, j)$ atas j .

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (2.20)$$

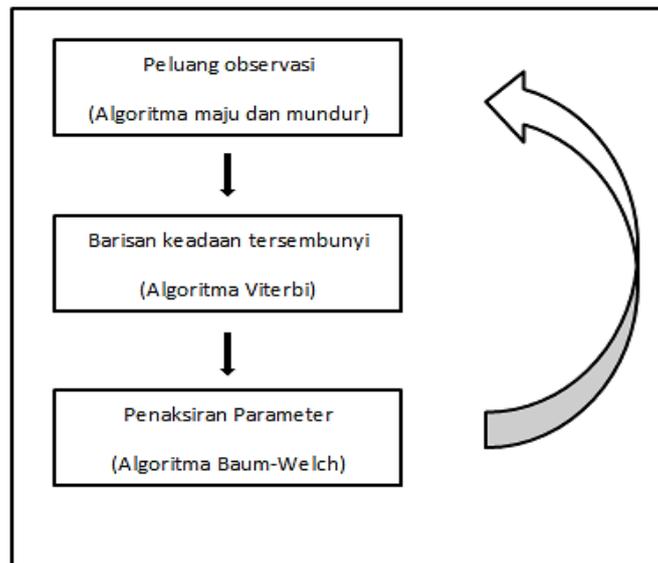
untuk melakukan estimasi terhadap nilai-nilai parameter MMT digunakan penaksir a_{ij} dan $b_i(j)$ di bawah ini:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

Penaksir tersebut diperoleh dengan membagi jumlah transisi dari keadaan tersembunyi i ke keadaan tersembunyi j dengan total seluruh transisi dari keadaan tersembunyi i .

$$\hat{b}_i(j) = \frac{\sum_{t=1, O_t=j}^T \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)}$$

diperoleh dengan membagi jumlah keadaan tersembunyi yang menghasilkan observasi j pada saat proses berada pada keadaan tersembunyi i dengan jumlah seluruh proses yang berada pada keadaan tersembunyi i .



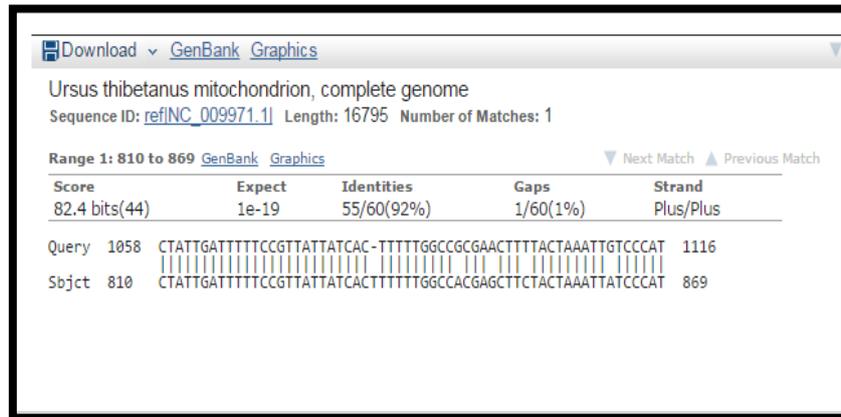
Gambar 4. Langkah – langkah algoritma Baum – Welch

5. STUDI KASUS DAN PEMBAHASAN

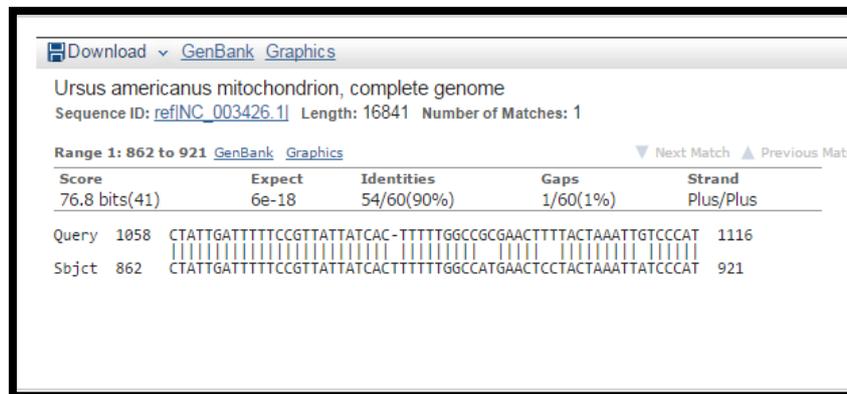
Penelitian yang dilakukan menggunakan pendekatan model rantai Markov tersembunyi untuk menganalisis *sequence alignment* pada DNA spesies *ursidae* (Beruang). Berdasarkan penelitian biomolekuler yang dilakukan oleh Syalini Venkataraman dan Vidhya Gunaseelan menyatakan bahwa barisan basa nitrogen penyusun DNA dapat diteliti menggunakan MMT.

a. Tahapan Utama

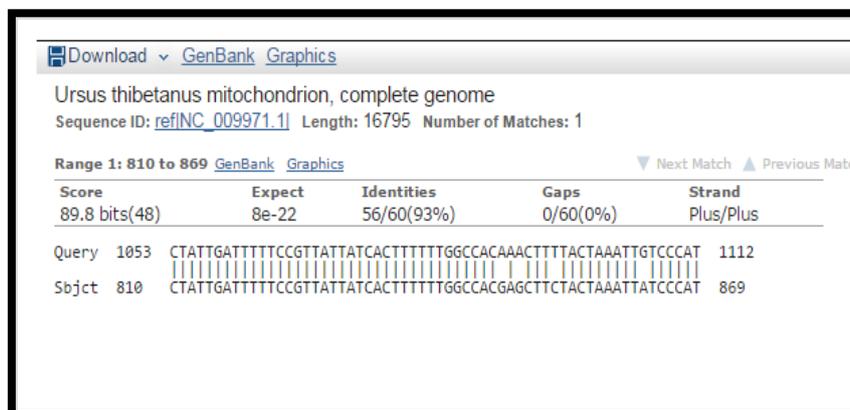
Terdapat beberapa tahapan utama yaitu penentuan spesies, teknik pengambilan data DNA, penentuan keadaan terobservasi dan keadaan tersembunyi.



Gambar 6. Hasil penyejajaran DNA Beruang coklat dan DNA Beruang hitam Asia



Gambar 7. Hasil penyejajaran DNA Beruang coklat dan DNA Beruang hitam Amerika



Gambar 8. Hasil penyejajaran DNA Beruang kutub dan DNA Beruang hitam Asia

1. Beruang coklat dan Beruang kutub

$$A = \begin{bmatrix} 53/56 & 2/56 & 1/56 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

2. Beruang coklat dan Beruang hitam asia

$$A = \begin{bmatrix} 49/54 & 4/54 & 1/54 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

3. Beruang coklat dan Beruang hitam amerika

$$A = \begin{bmatrix} 49/53 & 3/53 & 1/53 \\ 3/5 & 2/5 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

4. Beruang kutub dan Beruang hitam asia

$$A = \begin{bmatrix} 51/55 & 4/55 & 0 \\ 4/4 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

5. Beruang kutub dan Beruang hitam amerika

$$A = \begin{bmatrix} 51/54 & 3/54 & 0 \\ 3/5 & 2/5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

6. Beruang hitam asia dan Beruang hitam amerika

$$A = \begin{bmatrix} 53/56 & 3/56 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Matriks Emisi (B)

Matriks emisi untuk penelitian ini berordo 3×4. Matriks emisidiperoleh dengan menghitung jumlah basa nitrogen A, C, G, dan T berasal dari basa nitrogen yang match, insersi, atau delesi.

1. Beruang coklat dan Beruang kutub

$$B = \begin{bmatrix} 12/57 & 5/57 & 27/57 & 13/57 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

2. Beruang coklat dan Beruang hitam asia

$$B = \begin{bmatrix} 11/55 & 5/55 & 26/55 & 13/55 \\ 1/4 & 2/4 & 1/4 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

3. Beruang coklat dan Beruang hitam amerika

$$B = \begin{bmatrix} 12/54 & 5/54 & 25/54 & 12/54 \\ 0 & 2/5 & 2/5 & 1/5 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

4. Beruang kutub dan Beruang hitam asia

$$B = \begin{bmatrix} 12/56 & 4/56 & 27/56 & 13/56 \\ 2/4 & 1/4 & 1/4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

5. Beruang kutub dan Beruang hitam amerika

$$\mathbf{B} = \begin{bmatrix} 13/55 & 4/55 & 26/55 & 12/55 \\ 1/5 & 2/5 & 1/5 & 1/5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

6. Beruang hitam asia dan Beruang hitam amerika

$$\mathbf{B} = \begin{bmatrix} 13/57 & 5/57 & 26/57 & 13/57 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Matriks Prior

Kedaaan tersembunyi pada kasus ini adalah sebanyak 3, sehingga matriks prior pada penelitian ini berordo 3×1 . Matriks prior diperoleh dengan menghitung jumlah keadaan tersembunyi yaitu basa nitrogen yang *match*, *insersi*, dan *delesi*.

1. Beruang coklat dan Beruang kutub

$$\boldsymbol{\pi} = \begin{bmatrix} 57/60 \\ 2/60 \\ 1/60 \end{bmatrix}$$

2. Beruang coklat dan Beruang hitam asia

$$\boldsymbol{\pi} = \begin{bmatrix} 55/60 \\ 4/60 \\ 1/60 \end{bmatrix}$$

3. Beruang coklat dan Beruang hitam amerika

$$\boldsymbol{\pi} = \begin{bmatrix} 54/60 \\ 5/60 \\ 1/60 \end{bmatrix}$$

4. Beruang kutub dan Beruang hitam asia

$$\boldsymbol{\pi} = \begin{bmatrix} 56/60 \\ 4/60 \\ 0 \end{bmatrix}$$

5. Beruang kutub dan Beruang hitam amerika

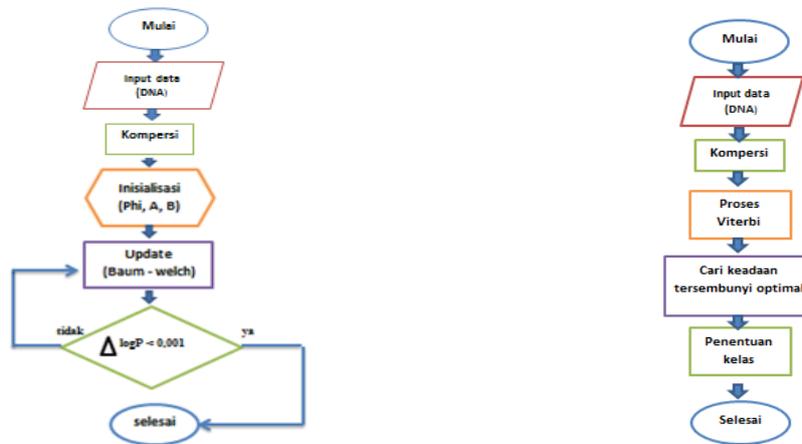
$$\boldsymbol{\pi} = \begin{bmatrix} 55/60 \\ 5/60 \\ 0 \end{bmatrix}$$

6. Beruang hitam asia dan Beruang hitam amerika

$$\boldsymbol{\pi} = \begin{bmatrix} 57/60 \\ 3/60 \\ 0 \end{bmatrix}$$

- c. Masalah – masalah dalam MMT

Setelah didapatkan nilai $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$, tahap selanjutnya adalah mengolah data dengan bantuan perangkat lunak MATLAB untuk mengetahui persentase kecocokan dari DNA keempat spesies beruang tersebut.



Gambar 11. (a) Alur proses untuk mendapatkan \mathbf{A} , \mathbf{B} , dan $\boldsymbol{\pi}$ Optimum (b) Alur proses untuk mendapatkan keadaan tersembunyi optimum

Persoalan pertama adalah mencari $\hat{\lambda} = (\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\boldsymbol{\pi}})$ optimum dengan menggunakan algoritma Baum Welch. Kemudian persoalan kedua mencari keadaan tersembunyi yang optimum dengan algoritma Viterbi menggunakan data \mathbf{A} , \mathbf{B} , dan $\boldsymbol{\pi}$ awal dan $\hat{\lambda} = (\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\boldsymbol{\pi}})$ untuk dibandingkan persentase kecocokan DNA dari keempat spesies yang diteliti dengan hasil yang didapatkan dari NCBI dan pohon filogenetik *family Ursidae* (Beruang).

Tabel 3 Parameter $\hat{\lambda} = (\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\boldsymbol{\pi}})$ optimum dari DNA

Spesies \ λ	A	B	$\boldsymbol{\pi}$
B.coklat vs B.Kutub	$\begin{bmatrix} 0.89 & 1 & 1 \\ 0.09 & 0 & 0 \\ 0.02 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.23 & 0.15 & 0.49 & 0.14 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 2.19 \\ 0.99 \\ 0 \end{bmatrix}$
B.coklat vs B.H.Asia	$\begin{bmatrix} 0.83 & 1 & 1 \\ 0.16 & 0 & 0 \\ 0.02 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.22 & 0.15 & 0.49 & 0.14 \\ 0.13 & 0.7 & 0.17 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1.85 \\ 0.99 \\ 0 \end{bmatrix}$
B.coklat vs B.H.Amerika	$\begin{bmatrix} 0.83 & 0.39 & 1 \\ 0.16 & 0.61 & 0 \\ 0.02 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.28 & 0.07 & 0.62 & 0.04 \\ 0 & 0.65 & 0.01 & 0.35 \\ 0 & 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1.95 \\ 1 \\ 0 \end{bmatrix}$
B.kutub vs B.H.Asia	$\begin{bmatrix} 0.83 & 1 & 0 \\ 0.17 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.20 & 0.16 & 0.50 & 0.14 \\ 0.18 & 0.64 & 0.19 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.95 \\ 0.10 \\ 0 \end{bmatrix}$
B.kutub vs B.H.Amerika	$\begin{bmatrix} 0.78 & 0.35 & 0 \\ 0.23 & 0.66 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.22 & 0.001 & 0.74 & 0.05 \\ 0.18 & 0.58 & 0.001 & 0.25 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 3.05 \\ 1 \\ 0 \end{bmatrix}$
B.H.Asia vs B.H.Amerika	$\begin{bmatrix} 0.87 & 1 & 0 \\ 0.14 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.23 & 0.15 & 0.49 & 0.13 \\ 0 & 0.76 & 0.20 & 0.04 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 3.05 \\ 0.10 \\ 0 \end{bmatrix}$

Tabel 4 Persentase hasil NCBI dan MMT

Spesies	Penelitian	NCBI	MMT (λ)	MMT ($\hat{\lambda}$)
B.coklat vs B.Kutub		95%	96,7%	98%
B.coklat vs B.H.Asia		92%	81,7%	81,6%
B.coklat vs B.H.Amerika		90%	80%	80%
B.kutub vs B.H.Asia		93%	83,3%	100%
B.kutub vs B.H.Amerika		92%	65%	65%
B.H.Asia vs B.H. Amerika		95%	98,3%	100%

6. KESIMPULAN

Model Makov Tersembunyi (MMT) dapat digunakan dalam menganalisis kecocokan DNA suatu spesies dari topologi pohon filogenetik. Hasil proses pengujian MMT dan NCBI pada keempat DNA dari *family ursidae* (Beruang) dapat disimpulkan bahwa keempat spesies tersebut berkerabat dekat, karena memiliki persentase kecocokan yang cukup besar.

DAFTAR PUSTAKA

- Afihayati, Mulyana S. Multiple Sequence Alignment menggunakan Hidden Markov Model, Semnas IF. UGM. 2008.
- Herlina I. Biologi untuk SMA kelas XII. PT. Remaja Rosda Karya. Jakarta. 2006
- Hidayat T, Pancoro A. Sistematika dan Filogenetika Molekuler. 2006
- Ibe O C. Markov Processes for Stochastic Modelling, London. Elsevier Inc. 2009
- Jamasoka, S. (2011) Modifikasi String dan Pattern untuk Mempercepat Pencocokan Rantai Asam Amino pada Rantai DNA, ITB.
- Mukherjee, S., Mitra, S. (2005). Hidden Markov Models, Grammars and Biology. A Tutorial. Journal of Bioinformatics and Computational Biology, Vol. 3, No. 2 : 491 – 526.
- Mulyana, S., Afihayati, Surya, A. W. (2008). Penerapan Hidden Markov Model dalam Clustering Sequence Protein Globin, UGM, Yogyakarta.
- NCBI, Mapview and Blast. Tersedia : www.ncbi.nlm.nih.gov [Januari 2015]
- Pandu, N. (2011). Aplikasi Algoritma Viterbi untuk Menentukan Barisan Keadaan tersembunyi pada Hidden Markov Models. Tugas Akhir Program Studi Matematika Universitas Pendidikan Indonesia. Tidak diterbitkan.
- Rabbiner, Lawrence, R. A. (1989). Tutorial on Hidden Markov Models and selected Applications in speech Recognition, Proceedings of the IEEE. Vol. 77, No. 2, pp. 257-286.
- Ross, S. M. (1996). Stochastic Processes. United States Of America. John Wiley and Sons, Inc.
- Siepel, A., Haussler, D. (2004). Combining Phylogenetic and Hidden Markov Models in Biosequence Analysis, Vol.11, No 2-3; 413-428.