BAB III

METODE JUKES-CANTOR DAN METODE KIMURA

Pada Bab ini akan dibahas mengenai metode Jukes-Cantor dan metode Kimura dalam membentuk suatu *evolutionary model*. Pada dasarnya kedua metode menggunakan ide yang sama dalam menggambarkan substitusi pada *single site*, yaitu bahwa penggambaran substitusi nukleotida pada *single site* ini berjalan sesuai dengan suatu *regular continuous-time Markov chain* [5]. Perbedaan utama antara keduanya terletak pada pembentukan matriks probabilitas transisi $P(t)$ dari *Markov chain* yang nantinya digunakan untuk menentukan jarak antara dua barisan. Berikut dijelaskan lebih lanjut mengenai kedua metode.

3.1 METODE JUKES-CANTOR

Metode Jukes-Cantor menghasilkan suatu model yang biasa disebut sebagai model satu parameter. Model ini mengasumsikan tingkat substitusi yang sama antara transisi dengan transversi. Atau dengan kata lain, tingkat substitusi suatu *site* berubah dari suatu *state* menjadi *state* lain dalam $S = \{A, C, G, T\}$ adalah sama. Sehingga untuk sembarang *state* dalam S, probabilitas transisi antar *state-state* tersebut juga sama, misal sama dengan

$f(t)$. Dengan demikian, dengan diasumsikan berjalan sesuai dengan suatu *regular continuous-time Markov chain*, maka pada Doron-Faigenboim, Adi [2] ditentukan matriks probabilitas transisi dari Jukes-Cantor sebagai berikut

$$\begin{array}{cc} & \begin{array}{cccc} \quad A & \qquad\quad C & \qquad\quad G & \qquad\quad T \end{array} \\ P(t)= & \begin{bmatrix} 1-3f(t) & f(t) & f(t) & f(t) \\ f(t) & 1-3f(t) & f(t) & f(t) \\ f(t) & f(t) & 1-3f(t) & f(t) \\ f(t) & f(t) & f(t) & 1-3f(t) \end{bmatrix} \begin{array}{c} A \\ C \\ G \\ T \end{array} \end{array}, \tag{3.1}$$

dengan $f(t)$ merupakan fungsi bernilai non negatif yang diferensiabel untuk setiap $t \geq 0$.

Pada subbab ini akan dijelaskan bentuk elemen-elemen dari matriks $P(t)$ (3.1) yang pembahasannya diberikan sebagai berikut.

Karena Jukes-Cantor diasumsikan berjalan sesuai dengan *regular Markov chain* maka seluruh elemen $P(t)$ dari Jukes-Cantor diferensiabel untuk setiap $t \geq 0$ sehingga elemen-elemen matriks $P(t)$ dari Jukes-Cantor tersebut dapat diturunkan terhadap t sebagai berikut

$$P'(t)=\begin{bmatrix} -3f'(t) & f'(t) & f'(t) & f'(t) \\ f'(t) & -3f'(t) & f'(t) & f'(t) \\ f'(t) & f'(t) & -3'f(t) & f'(t) \\ f'(t) & f'(t) & f'(t) & -3f'(t) \end{bmatrix}.$$

Saat $t=0$,

$$P'(0)=\begin{bmatrix} -3f'(0) & f'(0) & f'(0) & f'(0) \\ f'(0) & -3f'(0) & f'(0) & f'(0) \\ f'(0) & f'(0) & -3'f(0) & f'(0) \\ f'(0) & f'(0) & f'(0) & -3f'(0) \end{bmatrix}.$$

Misal didefinisikan suatu *matrix of instantaneous change* sebagai berikut

$$Q=P'(0)=\begin{bmatrix} -3\alpha & \alpha & \alpha & \alpha \\ \alpha & -3\alpha & \alpha & \alpha \\ \alpha & \alpha & -3\alpha & \alpha \\ \alpha & \alpha & \alpha & -3\alpha \end{bmatrix},$$

dimana $\alpha = f'(0)$ merupakan suatu konstanta tingkat substitusi yang bernilai positif.

Berdasarkan lemma pada subbab 2.4.1.2, untuk $t \geq 0$ berlaku

$$P'(t)=P(t)P'(0).$$

Maka

$$\begin{bmatrix} -3f'(t) & f'(t) & f'(t) & f'(t) \\ f'(t) & -3f'(t) & f'(t) & f'(t) \\ f'(t) & f'(t) & -3'f(t) & f'(t) \\ f'(t) & f'(t) & f'(t) & -3f'(t) \end{bmatrix}$$

$$=\begin{bmatrix} 1-3f(t) & f(t) & f(t) & f(t) \\ f(t) & 1-3f(t) & f(t) & f(t) \\ f(t) & f(t) & 1-3f(t) & f(t) \\ f(t) & f(t) & f(t) & 1-3f(t) \end{bmatrix} . \begin{bmatrix} -3\alpha & \alpha & \alpha & \alpha \\ \alpha & -3\alpha & \alpha & \alpha \\ \alpha & \alpha & -3\alpha & \alpha \\ \alpha & \alpha & \alpha & -3\alpha \end{bmatrix}.$$

Pandang elemen pada baris ke satu kolom ke dua dari perkalian matriks tersebut,

$$f'(t)=\left(1-3f(t)\right)\alpha+f(t)(-3\alpha)+f(t)\cdot\alpha+f(t)\cdot\alpha$$

$$f'(t) = \alpha - 3\alpha f(t) - 3\alpha f(t) + \alpha f(t) + \alpha f(t)$$

$$\frac{d(f(t))}{dt} = \alpha - 4\alpha f(t)$$

$$\frac{d(f(t))}{\alpha - 4\alpha f(t)} = dt\,.$$

Integralkan kedua ruas maka

$$\int \frac{d(f(t))}{\alpha - 4\alpha f(t)} = \int dt$$

$$\frac{\ln(\alpha - 4\alpha f(t))}{-4\alpha} = t + c_1$$

$$\ln(\alpha - 4\alpha f(t)) = -4\alpha t - 4\alpha c_1$$

$$\alpha - 4\alpha f(t) = e^{-4\alpha t - 4\alpha c_1}$$

$$= e^{-4\alpha t} e^{-4\alpha c_1}$$

$$= ce^{-4\alpha t}$$

$$-4\alpha f(t) = ce^{-4\alpha t} - \alpha$$

$$f(t) = \frac{1}{4} - \frac{c}{4\alpha} e^{-4\alpha t}\,. \tag{3.2}$$

Anggap nilai awal $f(0) = 0$. Sehingga untuk $t = 0$,

$$f(0) = \frac{1}{4} - \frac{c}{4\alpha} e^{-4\alpha \cdot 0} = 0\,,$$

diperoleh nilai $c = \alpha$. Substitusikan nilai c ke (3.2) diperoleh

$$f(t) = \frac{1}{4} - \frac{1}{4} e^{-4\alpha t} \tag{3.3}$$

Dengan mensubstitusikan (3.3) ke (3.1) maka diperoleh elemen-elemen matriks probabilitas transisi $P(t)$ dari Jukes-Cantor sebagai berikut

$$p_{ij}(t)=\begin{cases}\frac{1}{4}+\frac{3}{4}e^{-4\alpha t}, & i=j\\ \frac{1}{4}-\frac{1}{4}e^{-4\alpha t}, & i\neq j\end{cases} \tag{3.4}$$

untuk setiap $i, j \in S$, dan $t \geq 0$.

Berikutnya akan ditunjukkan dua kondisi yang harus dipenuhi oleh Jukes-Cantor dalam menggambarkan substitusi nukleotida.

1) *Stationary probability distribution* φ yang bersifat unik

Untuk mencari *stationary probability distribution* $\varphi=(\varphi_A,\varphi_C,\varphi_G,\varphi_T)$ dari Jukes-Cantor, perhatikan sistem persamaan linier yang terbentuk dari

$$\varphi Q=\begin{bmatrix}\varphi_A & \varphi_C & \varphi_G & \varphi_T\end{bmatrix}\begin{bmatrix}-3\alpha & \alpha & \alpha & \alpha\\ \alpha & -3\alpha & \alpha & \alpha\\ \alpha & \alpha & -3\alpha & \alpha\\ \alpha & \alpha & \alpha & -3\alpha\end{bmatrix}=\begin{bmatrix}0 & 0 & 0 & 0\end{bmatrix}=\mathbf{0}$$

berikut:

(1) $-3\varphi_A+\varphi_C+\varphi_G+\varphi_T=0$

(2) $\varphi_A-3\varphi_C+\varphi_G+\varphi_T=0$

(3) $\varphi_A+\varphi_C-3\varphi_G+\varphi_T=0$

(4) $\varphi_A+\varphi_C+\varphi_G-3\varphi_T=0$

Solusi untuk sistem persamaan linier tersebut adalah

$$\varphi_A=\varphi_C=\varphi_G=\varphi_T\ .$$

Akan dicari $\varphi_A = \varphi_C = \varphi_G = \varphi_T$ yang memenuhi kondisi

$$\sum_{i \in S} \varphi_i = \varphi_A + \varphi_C + \varphi_G + \varphi_T = 1 \text{ dengan } \varphi_i \geq 0 \text{ untuk setiap } i \in S.$$

Substitusikan $\varphi_A = \varphi_C = \varphi_G = \varphi_T$ ke $\sum_{i \in S} \varphi_i = \varphi_A + \varphi_C + \varphi_G + \varphi_T = 1$ diperoleh

$$\varphi_A = \varphi_C = \varphi_G = \varphi_T = \frac{1}{4}.$$

Berdasarkan Definisi 1 pada subbab 2.4.1.2 maka *stationary probability distribution* dari Jukes-Cantor adalah

$$\varphi = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right). \tag{3.5}$$

Ini merupakan satu-satunya *stationary probability distribution* dari Jukes-Cantor sehingga Jukes-Cantor memiliki *stationary probability distribution* φ yang bersifat unik.

2) Model yang bersifat *time-reversible*

Akan dibuktikan bahwa Jukes-Cantor merupakan model yang bersifat *time-reversible*.

Bukti.

Misalkan $\pi = (\pi_A, \pi_C, \pi_G, \pi_T)$ merupakan suatu vektor probabilitas awal dari Jukes-Cantor dengan $\pi = \varphi = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right)$. Berdasarkan Definisi 2 pada subbab 2.4.1.2, maka reversed *Markov chain* dari Jukes-Cantor didefinisikan oleh

$$p_{ij}*(t)=\frac{\varphi_j p_{ji}(t)}{\varphi_i},$$

untuk setiap $i,j\in S$, dan $t\geq 0$.

Karena $\varphi_i=\varphi_j=\frac{1}{4}$ untuk setiap $i,j\in S$ maka

$$p_{ij}*(t)=p_{ji}(t),$$

untuk setiap $i,j\in S$, dan $t\geq 0$.

Berdasarkan elemen-elemen matriks probabilitas transisi $P(t)$ dari Jukes-Cantor (3.4), maka untuk setiap $i,\ j\in S$, dan $t\geq 0$ diperoleh

$$p_{ij}*(t)=p_{ji}(t)=\begin{cases}\frac{1}{4}+\frac{3}{4}e^{-4\alpha t}, & j=i\\ \frac{1}{4}-\frac{1}{4}e^{-4\alpha t}, & j\neq i\end{cases}=\begin{cases}\frac{1}{4}+\frac{3}{4}e^{-4\alpha t}, & i=j\\ \frac{1}{4}-\frac{1}{4}e^{-4\alpha t}, & i\neq j\end{cases}=p_{ij}(t).$$

Atau dengan kata lain $P*(t)=P(t)$ untuk $t\geq 0$. Sehingga berdasarkan Definisi 3 pada subbab 2.4.1.2, model Jukes-Cantor merupakan model yang bersifat *time-reversible*.

Berdasarkan 1) dan 2) maka kedua kondisi yang harus dipenuhi oleh Jukes-Cantor terpenuhi.

Model Jukes-Cantor diperoleh dengan mengestimasi jarak antara dua barisan berbeda dengan panjang yang sama menggunakan *maximum likelihood*.

Berdasarkan model *likelihood* (2.8) pada subbab 2.4.2,

$$L=\ln\varphi_{x_1}+\ln\varphi_{x_2}+...+\ln\varphi_{x_n}+\ln p_{x_1y_1}(t)+\ln p_{x_2y_2}(t)+\mathrm{K}+\ln p_{x_ny_n}(t).$$

Pandang elemen-elemen matriks probabilitas transisi $P(t)$ (3.4) dan *stationary probability distribution* φ (3.5) dari Jukes-Cantor. Maka model *likelihood* untuk Jukes-Cantor dapat ditulis menjadi

$$\begin{aligned}\ln L &= \ln\frac{1}{4}+\ln\frac{1}{4}+\ldots+\ln\frac{1}{4}+m_1\ln p_{sama}(t)+m_2\ln p_{beda}(t)\\ &= n\ln\frac{1}{4}+m_1\ln p_{sama}(t)+m_2\ln p_{beda}(t)\end{aligned} \qquad (3.6)$$

dimana $m_1+m_2=n$, dengan

m_1 = banyaknya substitusi ke *state* yang sama,

m_2 = banyaknya substitusi ke *state* yang berbeda,

$p_{sama}(t)$ = probabilitas *site* berubah ke *state* yang sama,

$p_{beda}(t)$ = probabilitas *site* berubah ke *state* yang berbeda,

untuk $t \geq 0$.

Dengan memaksimumkan model *likelihood* (3.6) maka akan diperoleh

$$\frac{d(\ln L)}{dt}=0$$

$$\frac{d\left(n\ln\frac{1}{4}+m_1\ln p_{sama}(t)+m_2\ln p_{beda}(t)\right)}{dt}=0$$

$$m_1\frac{p_{sama}'(t)}{p_{sama}(t)}+m_2\frac{p_{beda}'(t)}{p_{beda}(t)}=0 \qquad (3.7)$$

Berdasarkan matriks probabilitas transisi $P(t)$ dari Jukes-Cantor (3.4) maka

$$p_{sama}(t)=\frac{1}{4}+\frac{3}{4}e^{-4\alpha t}\ \Rightarrow\ p_{sama}'(t)=\frac{d\left(p_{ii}(t)\right)}{dt}=-3\alpha e^{-4\alpha t}\text{, dan} \tag{3.8}$$

$$p_{beda}(t)=\frac{1}{4}-\frac{1}{4}e^{-4\alpha t}\ \Rightarrow\ p_{beda}'(t)=\frac{d\left(p_{ij}(t)\right)}{dt}=\alpha e^{-4\alpha t}. \tag{3.9}$$

Substitusikan (3.8) dan (3.9) ke (3.7), maka model *likelihood* (3.7) sekarang menjadi

$$m_1\frac{-3\alpha e^{-4\alpha t}}{\frac{1}{4}+\frac{3}{4}e^{-4\alpha t}}+m_2\frac{\alpha e^{-4\alpha t}}{\frac{1}{4}-\frac{1}{4}e^{-4\alpha t}}=0 \tag{3.10}$$

Karena α merupakan suatu konstanta yang bernilai positif maka

$$m_1\frac{-3e^{-4\alpha t}}{\frac{1}{4}+\frac{3}{4}e^{-4\alpha t}}+m_2\frac{e^{-4\alpha t}}{\frac{1}{4}-\frac{1}{4}e^{-4\alpha t}}=0.$$

Misal $z=e^{-4\alpha t}$, maka

$$m_1\frac{-3z}{\frac{1}{4}+\frac{3}{4}z}+m_2\frac{z}{\frac{1}{4}-\frac{1}{4}z}=0$$

$$-3m_1\frac{z}{1+3z}+m_2\frac{z}{1-z}=0$$

$$-3m_1(1-z)z+m_2(1+3z)z=0$$

$$-3m_1z+3m_1z^2+m_2z+3m_2z^2=0$$

$$(3m_1+3m_2)z^2-(3m_1-m_2)z=0$$

diperoleh

$$z=\frac{3m_1-m_2}{3m_1+3m_2}.$$

Dengan mensubstitusikan kembali nilai $z = e^{-4\alpha t}$ maka

$$e^{-4\alpha t} = \frac{3m_1 - m_2}{3m_1 + 3m_2}$$

$$-4\alpha t = \ln\frac{3m_1 - m_2}{3m_1 + 3m_2}$$

$$t = -\frac{1}{4\alpha}\ln\frac{3m_1 - m_2}{3m_1 + 3m_2}$$

$$= -\frac{1}{4\alpha}\ln\frac{3m_1 - m_2 + 3m_2 - 3m_2}{3m_1 + 3m_2}$$

$$= -\frac{1}{4\alpha}\ln\frac{(3m_1 + 3m_2) - (-m_2 - 3m_2)}{3m_1 + 3m_2}$$

$$= -\frac{1}{4\alpha}\ln\left(1 - \frac{4m_2}{3m_1 + 3m_2}\right)$$

$$= -\frac{1}{4\alpha}\ln\left(1 - \frac{4}{3}\frac{m_2}{m_1 + m_2}\right)$$

$$= -\frac{1}{4\alpha}\ln\left(1 - \frac{4}{3}r\right),$$

dimana $r = \dfrac{m_2}{m_1 + m_2} = \dfrac{m_2}{n} = \dfrac{\text{banyaknya substitusi ke state yang berbeda}}{\text{panjang barisan}}$.

Karena pada subbab 2.4.2 t didefinisikan sebagai jarak antara dua barisan (d) maka diperoleh

$$d = -\frac{1}{4\alpha}\ln\left(1 - \frac{4}{3}r\right).$$

Dengan pemilihan nilai $\alpha = \frac{1}{3}$ maka diperoleh model Jukes-Cantor yang umum digunakan untuk menghitung jarak antara dua barisan yang berbeda sebagai berikut

$$d = -\frac{3}{4}\ln\left(1 - \frac{4}{3}r\right) \tag{3.11}$$

dimana $r = \dfrac{\text{banyaknya substitusi ke state yang berbeda}}{\text{panjang barisan}}$.

3.2 METODE KIMURA

Metode Kimura menghasilkan suatu model yang biasa disebut sebagai *Kimura 2-Parameter model* atau model dua parameter. Model ini merupakan generalisasi dari model Jukes-Cantor yang telah dibahas pada subbab sebelumnya. Model ini mengasumsikan tingkat substitusi yang berbeda antara transisi dengan transversi sehingga probabilitas transisi antara substitusi transisi dengan substitusi transversi tersebut juga berbeda, misal masing-masing sama dengan $f(t)$ dan $g(t)$. Sehingga dengan diasumsikan berjalan sesuai dengan suatu *regular continuous-time Markov chain* maka pada Penn, Osnat [8] ditentukan matriks probabilitas transisi dari Kimura sebagai berikut

$$P(t)=\begin{array}{cccccc} & A & C & G & T & \\ & \left[\begin{array}{cccc} 1-f(t)-2g(t) & g(t) & f(t) & g(t) \\ g(t) & 1-f(t)-2g(t) & g(t) & f(t) \\ f(t) & g(t) & 1-f(t)-2g(t) & g(t) \\ g(t) & f(t) & g(t) & 1-f(t)-2g(t) \end{array}\right] & & & & \begin{array}{c} A \\ C \\ G \\ T \end{array} \end{array}$$

(3.12)

dengan $f(t)$ dan $g(t)$ merupakan fungsi-fungsi bernilai non negatif yang diferensiabel untuk setiap $t \geq 0$.

Sama seperti pada pembahasan model Jukes-Cantor sebelumnya, akan dicari bentuk elemen-elemen dari matriks $P(t)$ (3.12).

Karena Kimura diasumsikan berjalan sesuai dengan *regular Markov chain* maka seluruh elemen $P(t)$ dari Kimura diferensiabel untuk setiap $t \geq 0$ sehingga elemen-elemen matriks $P(t)$ dari Kimura tersebut dapat diturunkan terhadap t sebagai berikut

$$P'(t)=\begin{bmatrix} -f'(t)-2g'(t) & g'(t) & f'(t) & g'(t) \\ g'(t) & -f'(t)-2g'(t) & g'(t) & f'(t) \\ f'(t) & g'(t) & -f'(t)-2g'(t) & g'(t) \\ g'(t) & f'(t) & g'(t) & -f'(t)-2g'(t) \end{bmatrix}.$$

Saat $t=0$,

$$P'(0)=\begin{bmatrix} -f'(0)-2g'(0) & g'(0) & f'(0) & g'(0) \\ g'(0) & -f'(0)-2g'(0) & g'(0) & f'(0) \\ f'(0) & g'(0) & -f'(0)-2g'(0) & g'(0) \\ g'(0) & f'(0) & g'(0) & -f'(0)-2g'(0) \end{bmatrix}.$$

Misal didefinisikan suatu *matrix of instantaneous change* sebagai berikut

$$Q = P'(0) = \begin{bmatrix} -\alpha-2\beta & \beta & \alpha & \beta \\ \beta & -\alpha-2\beta & \beta & \alpha \\ \alpha & \beta & -\alpha-2\beta & \beta \\ \beta & \alpha & \beta & -\alpha-2\beta \end{bmatrix},$$

dimana $\alpha = f'(0)$ dan $\beta = g'(0)$ merupakan konstanta-konstanta tingkat substitusi yang bernilai positif, dengan α merupakan tingkat substitusi dari transisi dan β merupakan tingkat substitusi dari transversi.

Berdasarkan lemma pada subbab 2.4.1.2, untuk $t \geq 0$ berlaku

$$P'(t) = P(t)P'(0).$$

Maka

$$\begin{bmatrix} -f'(t)-2g'(t) & g'(t) & f'(t) & g'(t) \\ g'(t) & -f'(t)-2g'(t) & g'(t) & f'(t) \\ f'(t) & g'(t) & -f'(t)-2g'(t) & g'(t) \\ g'(t) & f'(t) & g'(t) & -f'(t)-2g'(t) \end{bmatrix}$$

$$= \begin{bmatrix} 1-f(t)-2g(t) & g(t) & f(t) & g(t) \\ g(t) & 1-f(t)-2g(t) & g(t) & f(t) \\ f(t) & g(t) & 1-f(t)-2g(t) & g(t) \\ g(t) & f(t) & g(t) & 1-f(t)-2g(t) \end{bmatrix}.$$

$$\begin{bmatrix} -\alpha-2\beta & \beta & \alpha & \beta \\ \beta & -\alpha-2\beta & \beta & \alpha \\ \alpha & \beta & -\alpha-2\beta & \beta \\ \beta & \alpha & \beta & -\alpha-2\beta \end{bmatrix}.$$

Pandang elemen pada baris ke satu kolom ke tiga dari perkalian matriks tersebut,

$$f'(t) = \left(1-f(t)-2g(t)\right)\alpha + g(t)\beta + f(t)(-\alpha-2\beta) + g(t)\beta$$

$$f'(t) = \alpha - \alpha f(t) - 2\alpha g(t) + \beta g(t) - \alpha f(t) - 2\beta f(t) + \beta g(t)$$

$$\frac{d(f(t))}{dt} = \alpha - 2(\alpha + \beta) f(t) - 2(\alpha - \beta) g(t). \qquad (3.13)$$

Pandang kembali elemen pada baris ke satu kolom ke dua dari perkalian matriks tersebut,

$$g'(t) = (1 - f(t) - 2g(t))\beta + g(t)(-\alpha - 2\beta) + f(t)\beta + g(t)\alpha$$

$$= \beta - \beta f(t) - 2\beta g(t) - \alpha g(t) - 2\beta g(t) + \beta f(t) + \alpha g(t)$$

$$\frac{d(g(t))}{dt} = \beta - 4\beta g(t)$$

$$\frac{d(g(t))}{\beta - 4\beta g(t)} = dt.$$

Integralkan kedua ruas maka

$$\int \frac{d(g(t))}{\beta - 4\beta g(t)} = \int dt$$

$$\frac{\ln(\beta - 4\beta g(t))}{-4\beta} = t + c_1$$

$$\ln(\beta - 4\beta g(t)) = -4\beta t - 4\beta c_1$$

$$\beta - 4\beta g(t) = e^{-4\beta t - 4\beta c_1}$$

$$= e^{-4\beta t} e^{-4\beta c_1}$$

$$= ce^{-4\beta t}$$

$$-4\beta g(t) = ce^{-4\beta t} + \beta$$

$$g(t)=\frac{1}{4}-\frac{c}{4\beta}e^{-4\beta t}. \tag{3.14}$$

Anggap nilai awal $g(0)=0$. Sehingga untuk $t=0$,

$$g(0)=\frac{1}{4}-\frac{c}{4\beta}e^{-4\beta\cdot 0}=0,$$

diperoleh nilai $c=\beta$. Substitusikan nilai c ke (3.14) diperoleh

$$g(t)=\frac{1}{4}-\frac{1}{4}e^{-4\beta t}. \tag{3.15}$$

Substitusikan (3.15) ke (3.13) diperoleh

$$\begin{aligned}\frac{d(f(t))}{dt}&=\alpha-2(\alpha+\beta)f(t)-2(\alpha-\beta)\left(\frac{1}{4}-\frac{1}{4}e^{-4\beta t}\right)\\&=\alpha-2(\alpha+\beta)f(t)-\frac{\alpha-\beta}{2}+\frac{\alpha-\beta}{2}e^{-4\beta t}\\&=-2(\alpha+\beta)f(t)+\frac{\alpha+\beta}{2}+\frac{\alpha-\beta}{2}e^{-4\beta t}\end{aligned}$$

$$\frac{d(f(t))}{dt}+f(t)\big(2(\alpha+\beta)\big)=\frac{\alpha+\beta}{2}+\frac{\alpha-\beta}{2}e^{-4\beta t}. \tag{3.16}$$

Persamaan diferensial (3.16) ini merupakan suatu persamaan diferensial linier. Untuk menyelesaikannya, kalikan kedua ruas dengan faktor integrasi $e^{\int 2(\alpha+\beta)dt}$.

$$\left(\frac{d(f(t))}{dt}+f(t)\big(2(\alpha+\beta)\big)\right)e^{\int 2(\alpha+\beta)dt}=\left(\frac{\alpha+\beta}{2}+\frac{\alpha-\beta}{2}e^{-4\beta t}\right)e^{\int 2(\alpha+\beta)dt}$$

$$\frac{d(f(t))}{dt}e^{\int 2(\alpha+\beta)dt}+f(t)\big(2(\alpha+\beta)\big)e^{\int 2(\alpha+\beta)dt}=\frac{\alpha+\beta}{2}e^{\int 2(\alpha+\beta)dt}+\frac{\alpha-\beta}{2}e^{-4\beta t}e^{\int 2(\alpha+\beta)dt}$$

$$\frac{d\left(f(t)\right)}{dt}e^{\int 2(\alpha+\beta)dt}+f(t)\frac{d\left(e^{\int 2(\alpha+\beta)dt}\right)}{dt}=\frac{\alpha+\beta}{2}e^{\int 2(\alpha+\beta)dt}+\frac{\alpha-\beta}{2}e^{-4\beta t}e^{\int 2(\alpha+\beta)dt}$$

$$\frac{d\left(f(t)e^{\int p(t)dt}\right)}{dt}=\frac{\alpha+\beta}{2}e^{\int 2(\alpha+\beta)dt}+\frac{\alpha-\beta}{2}e^{-4\beta t}e^{\int 2(\alpha+\beta)dt}$$

$$d\left(f(t)e^{\int p(t)dt}\right)=\frac{\alpha+\beta}{2}e^{\int 2(\alpha+\beta)dt}+\frac{\alpha-\beta}{2}e^{-4\beta t}e^{\int 2(\alpha+\beta)dt}dt$$

Integralkan kedua ruas maka

$$f(t)e^{\int 2(\alpha+\beta)dt}=\int\left(\frac{\alpha+\beta}{2}+\frac{\alpha-\beta}{2}e^{-4\beta t}\right)e^{\int 2(\alpha+\beta)dt}dt$$

$$f(t)e^{2(\alpha+\beta)t}=\int\left(\frac{\alpha+\beta}{2}+\frac{\alpha-\beta}{2}e^{-4\beta t}\right)e^{2(\alpha+\beta)t}dt$$

$$=\int\left(\frac{\alpha+\beta}{2}e^{2(\alpha+\beta)t}+\frac{\alpha-\beta}{2}e^{2(\alpha-\beta)t}\right)dt$$

$$=\frac{\alpha+\beta}{2}\cdot\frac{e^{2(\alpha+\beta)t}}{2(\alpha+\beta)}+\frac{\alpha-\beta}{2}\frac{e^{2(\alpha-\beta)t}}{2(\alpha-\beta)}+c$$

$$=\frac{1}{4}e^{2(\alpha+\beta)t}+\frac{1}{4}e^{2(\alpha-\beta)t}+c$$

$$f(t)=\frac{1}{4}+\frac{1}{4}e^{-4\beta t}+ce^{-2(\alpha+\beta)t} \qquad (3.17)$$

Anggap nilai awal $f(0)=0$. Sehingga untuk $t=0$,

$$f(0)=\frac{1}{4}+\frac{1}{4}e^{-4\beta\cdot 0}+ce^{-2(\alpha+\beta)\cdot 0}=0,$$

diperoleh nilai $c=-\frac{1}{2}$. Substitusikan nilai c ke (3.17) diperoleh

$$f(t)=\frac{1}{4}+\frac{1}{4}e^{-4\beta t}-\frac{1}{2}e^{-2(\alpha+\beta)t} \tag{3.18}$$

Dengan mensubstitusikan (3.15) dan (3.18) ke (3.12) maka diperoleh elemen-elemen matriks probabilitas transisi $P(t)$ dari Kimura sebagai berikut

(1) $p_{AA}(t)=p_{CC}(t)=p_{GG}(t)=p_{TT}(t)=\frac{1}{4}+\frac{1}{4}e^{-4\beta t}+\frac{1}{2}e^{-2(\alpha+\beta)t}$,

(2) $p_{AG}(t)=p_{GA}(t)=p_{CT}(t)=p_{TC}(t)=\frac{1}{4}+\frac{1}{4}e^{-4\beta t}-\frac{1}{2}e^{-2(\alpha+\beta)t}$, dan

(3) $p_{AC}(t)=p_{CA}(t)=p_{AT}(t)=p_{TT}(t)=p_{CG}(t)=p_{GC}(t)=p_{GT}(t)=p_{TG}(t)$

$$=\frac{1}{4}-\frac{1}{4}e^{-4\beta t}. \tag{3.19}$$

untuk $t\geq 0$.

Berikutnya akan ditunjukkan dua kondisi yang harus dipenuhi oleh Kimura dalam menggambarkan substitusi nukleotida.

1) *Stationary probability distribution* φ yang bersifat unik

Untuk mencari *stationary probability distribution* $\varphi=(\varphi_A,\varphi_C,\varphi_G,\varphi_T)$ dari Kimura, perhatikan sistem persamaan linier yang terbentuk dari

$$\varphi Q=[\varphi_A\ \varphi_C\ \varphi_G\ \varphi_T]\begin{bmatrix}-\alpha-2\beta & \beta & \alpha & \beta\\ \beta & -\alpha-2\beta & \beta & \alpha\\ \alpha & \beta & -\alpha-2\beta & \beta\\ \beta & \alpha & \beta & -\alpha-2\beta\end{bmatrix}=[0\quad 0\quad 0\quad 0]$$
$$=\mathbf{0}$$

berikut:

(1) $(-\alpha-2\beta)\varphi_A+\beta\varphi_C+\alpha\varphi_G+\beta\varphi_T=0$

(2) $\beta\varphi_A+(-\alpha-2\beta)\varphi_C+\beta\varphi_G+\alpha\varphi_T=0$

(3) $\alpha\varphi_A+\beta\varphi_C+(-\alpha-2\beta)\varphi_G+\beta\varphi_T=0$

(4) $\beta\varphi_A+\alpha\varphi_C+\beta\varphi_G+(-\alpha-2\beta)\varphi_T=0$

Solusi untuk sistem persamaan linier tersebut adalah

$$\varphi_A=\varphi_C=\varphi_G=\varphi_T\,.$$

Akan dicari $\varphi_A=\varphi_C=\varphi_G=\varphi_T$ yang memenuhi kondisi

$$\sum_{i\in S}\varphi_i=\varphi_A+\varphi_C+\varphi_G+\varphi_T=1 \text{ dengan } \varphi_i\geq 0 \text{ untuk setiap } i\in S\,.$$

Substitusikan $\varphi_A=\varphi_C=\varphi_G=\varphi_T$ ke $\sum_{i\in S}\varphi_i=\varphi_A+\varphi_C+\varphi_G+\varphi_T=1$ diperoleh

$$\varphi_A=\varphi_C=\varphi_G=\varphi_T=\frac{1}{4}.$$

Berdasarkan Definisi 1 pada subbab 2.4.1.2 maka s*tationary probability distribution* dari Kimura adalah

$$\varphi=\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right). \tag{3.20}$$

$\varphi=\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)$ ini merupakan satu-satunya s*tationary probability distribution* dari Kimura sehingga Kimura memiliki *stationary probability distribution* φ yang bersifat unik.

2) Model yang bersifat *time-reversible*

Akan dibuktikan bahwa Kimura merupakan model yang bersifat *time-reversible*.

Bukti.

Misalkan $\pi = (\pi_A, \pi_C, \pi_G, \pi_T)$ merupakan suatu vektor probabilitas awal dari Kimura dengan $\pi = \varphi = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$. Berdasarkan Definisi 2 pada subbab 2.4.1.2, maka reversed *Markov chain* dari Kimura didefinisikan oleh

$$p_{ij}{}^*(t) = \frac{\varphi_j p_{ji}(t)}{\varphi_i},$$

untuk setiap $i, j \in S$, dan $t \geq 0$.

Karena $\varphi_i = \varphi_j = \frac{1}{4}$ untuk setiap $i, j \in S$ maka

$$p_{ij}{}^*(t) = p_{ji}(t),$$

untuk setiap $i, j \in S$, dan $t \geq 0$.

Berdasarkan elemen-elemen matriks probabilitas transisi $P(t)$ dari Kimura (3.22), maka untuk $t \geq 0$ diperoleh

(1) $p_{AA}{}^*(t) = p_{CC}{}^*(t) = p_{GG}{}^*(t) = p_{TT}{}^*(t) = \frac{1}{4} + \frac{1}{4}e^{-4\beta t} + \frac{1}{2}e^{-2(\alpha+\beta)t}$,

(2) $p_{AG}{}^*(t) = p_{GA}{}^*(t) = p_{CT}{}^*(t) = p_{TC}{}^*(t) = \frac{1}{4} + \frac{1}{4}e^{-4\beta t} - \frac{1}{2}e^{-2(\alpha+\beta)t}$, dan

(3) $p_{AC}*(t) = p_{CA}*(t) = p_{AT}*(t) = p_{TT}*(t) = p_{CG}*(t) = p_{GC}*(t) = p_{GT}*(t) = p_{TG}*(t)$

$$= \frac{1}{4} - \frac{1}{4}e^{-4\beta t}.$$

Atau dengan kata lain $P*(t) = P(t)$ untuk $t \geq 0$. Sehingga berdasarkan Definisi 3 pada subbab 2.4.1.2, model Kimura merupakan model yang bersifat *time-reversible*.

Berdasarkan 1) dan 2) maka kedua kondisi yang harus dipenuhi oleh Kimura terpenuhi.

Seperti yang telah disebutkan sebelumnya, model Kimura merupakan generalisasi dari model Jukes-Cantor. Hal ini dikarenakan pada tingkat substitusi transisi yang sama dengan tingkat substitusi transversi, yaitu pada pemilihan nilai $\alpha = \beta$, maka model Kimura akan memiliki bentuk yang sama dengan Jukes-Cantor. Berikut diberikan penjelasannya.

Berdasarkan model *likelihood* (2.8) pada subbab 2.4.2,

$$L = \ln\varphi_{x_1} + \ln\varphi_{x_2} + ... + \ln\varphi_{x_n} + \ln p_{x_1y_1}(t) + \ln p_{x_2y_2}(t) + \mathrm{K} + \ln p_{x_ny_n}(t).$$

Pandang elemen-elemen matriks probabilitas transisi $P(t)$ (3.19) dan *stationary probability distribution* φ (3.20) dari Kimura. Maka model *likelihood* untuk Kimura dapat ditulis menjadi

$$\ln L = \ln\frac{1}{4} + \ln\frac{1}{4} + ... + \ln\frac{1}{4} + m_1 \ln p_{sama}(t) + m_2 \ln p_{transisi}(t) + m_3 \ln p_{transversi}(t)$$

$$= n\ln\frac{1}{4} + m_1 \ln p_{sama}(t) + m_2 \ln p_{transisi}(t) + m_3 \ln p_{transversi}(t), \qquad (3.21)$$

dimana $m_1 + m_2 + m_3 = n$, dengan

m_1 = banyaknya substitusi ke *state* yang sama,

m_2 = banyaknya substitusi transisi,

m_3 = banyaknya substitusi transversi,

$p_{sama}(t)$ = probabilitas *site* berubah ke *state* yang sama,

$p_{transisi}(t)$ = probabilitas substitusi transisi,

$p_{transversi}(t)$ = probabilitas substitusi transversi,

untuk $t \geq 0$.

Dengan memaksimumkan model *likelihood* (3.21) maka akan diperoleh

$$\frac{d(\ln L)}{dt} = 0$$

$$\frac{d\left(n\ln\frac{1}{4} + m_1 \ln p_{sama}(t) + m_2 \ln p_{transisi}(t) + m_3 \ln p_{transversi}(t)\right)}{dt} = 0$$

$$m_1 \frac{p_{sama}'(t)}{p_{sama}(t)} + m_2 \frac{p_{transisi}'(t)}{p_{transisi}(t)} + m_3 \frac{p_{transversi}'(t)}{p_{transversi}(t)} = 0. \quad (3.22)$$

Berdasarkan matriks probabilitas transisi $P(t)$ dari Kimura (3.19) maka

$$p_{sama}(t) = \frac{1}{4} + \frac{1}{4}e^{-4\beta t} + \frac{1}{2}e^{-2(\alpha+\beta)t} \Rightarrow p_{sama}'(t) = -\beta e^{-4\beta t} - (\alpha+\beta)e^{-2(\alpha+\beta)t} \quad (3.23)$$

$$p_{transisi}(t) = \frac{1}{4} + \frac{1}{4}e^{-4\beta t} - \frac{1}{2}e^{-2(\alpha+\beta)t} \Rightarrow p_{transisi}'(t) = -\beta e^{-4\beta t} + (\alpha+\beta)e^{-2(\alpha+\beta)t} \quad (3.24)$$

$$p_{transversi}(t) = \frac{1}{4} - \frac{1}{4}e^{-4\beta t} \Rightarrow p_{transversi}'(t) = \beta e^{-4\beta t}. \quad (3.25)$$

Substitusikan (3.23), (3.24), dan (3.25) ke (3.22), maka model *likelihood* (3.22) sekarang menjadi

$$m_1 \frac{-\beta e^{-4\beta t} - (\alpha+\beta) e^{-2(\alpha+\beta)t}}{\frac{1}{4}+\frac{1}{4}e^{-4\beta t}+\frac{1}{2}e^{-2(\alpha+\beta)t}} + m_2 \frac{-\beta e^{-4\beta t} + (\alpha+\beta) e^{-2(\alpha+\beta)t}}{\frac{1}{4}+\frac{1}{4}e^{-4\beta t}-\frac{1}{2}e^{-2(\alpha+\beta)t}} + m_3 \frac{\beta e^{-4\beta t}}{\frac{1}{4}-\frac{1}{4}e^{-4\beta t}} = 0 \qquad (3.26)$$

dimana α dan β merupakan konstanta-konstanta tingkat substitusi yang bernilai positif dengan α merupakan tingkat substitusi dari transisi dan β merupakan tingkat substitusi dari transversi.

Pilih $\alpha = \beta$, maka model (3.26) menjadi

$$m_1 \frac{-\alpha e^{-4\alpha t} - 2\alpha e^{-4\alpha t}}{\frac{1}{4}+\frac{1}{4}e^{-4\alpha t}+\frac{1}{2}e^{-4\alpha t}} + m_2 \frac{-\alpha e^{-4\alpha t} + 2\alpha e^{-4\alpha t}}{\frac{1}{4}+\frac{1}{4}e^{-4\alpha t}-\frac{1}{2}e^{-4\alpha t}} + m_3 \frac{\alpha e^{-4\alpha t}}{\frac{1}{4}-\frac{1}{4}e^{-4\alpha t}} = 0$$

$$m_1 \frac{-3\alpha e^{-4\alpha t}}{\frac{1}{4}+\frac{3}{4}e^{-4\alpha t}} + m_2 \frac{\alpha e^{-4\alpha t}}{\frac{1}{4}-\frac{1}{4}e^{-4\alpha t}} + m_3 \frac{\alpha e^{-4\alpha t}}{\frac{1}{4}-\frac{1}{4}e^{-4\alpha t}} = 0$$

$$m_1 \frac{-3\alpha e^{-4\alpha t}}{\frac{1}{4}+\frac{3}{4}e^{-4\alpha t}} + (m_2+m_3) \frac{\alpha e^{-4\alpha t}}{\frac{1}{4}-\frac{1}{4}e^{-4\alpha t}} = 0,$$

atau

$$m_1 \frac{-3\alpha e^{-4\alpha t}}{\frac{1}{4}+\frac{3}{4}e^{-4\alpha t}} + m_4 \frac{\alpha e^{-4\alpha t}}{\frac{1}{4}-\frac{1}{4}e^{-4\alpha t}} = 0, \qquad (3.27)$$

dimana m_1 adalah banyaknya substitusi ke *state* yang sama dan $m_4 = m_2 + m_3$ adalah banyaknya substitusi ke *state* yang berbeda. Model (3.27) ini merupakan model yang sama seperti yang diberikan oleh model *maximum likelihood* Jukes-Cantor (3.10). Sehingga dengan kata lain, untuk $\alpha = \beta$, Kimura akan menghasilkan model yang sama dengan model Jukes-Cantor.

Berikut diberikan model Kimura untuk menghitung jarak antara dua barisan.

$$d = -\frac{1}{2}\ln\left(\left(1-2r_1-r_2\right)\sqrt{1-2r_2}\right), \tag{3.28}$$

dimana

$$r_1 = \frac{\text{banyaknya substitusi transisi}}{\text{panjang barisan}}, \text{ dan } r_2 = \frac{\text{banyaknya substitusi transversi}}{\text{panjang barisan}}.$$

Perolehan model Kimura (3.28) tersebut tidak akan dijelaskan lebih lanjut pada skripsi ini, namun dapat dilihat perolehan selengkapnya pada Kimura, Motoo [6].

Model Jukes-Cantor (3.11) dan model Kimura (3.28) ini selanjutnya digunakan untuk menghitung jarak tiap pasang barisan yang akan disimulasikan dalam bentuk program pada Bab IV terhadap suatu data barisan DNA.