

‘‘Wstępow do obliczeniowej biologii molekularnej’’  
(J. Tiuryn, wykład nr.10, 29 grudnia 2005)

## Spis treści

<b>6 Ewolucja sekwencji DNA</b>	<b>71</b>
6.1 Model Jukesa i Cantora (1969) . . . . .	71
6.2 Model Kimury (1980) . . . . .	73

## 6 Ewolucja sekwencji DNA

Przedstawimy dwa teoretyczne modele dynamiki substytucji nukleotydów w sekwencjach DNA.

### 6.1 Model Jukesa i Cantora (1969)

Jest to najprostszy model. Zakładamy, że każdy nukleotyd z równym prawdopodobieństwem  $\alpha$  zmienia się w jeden z trzech pozostałych nukleotydów. Zatem nukleotyd ten pozostaje niezmienny z prawdopodobieństwem  $1 - 3\alpha$ . Niech  $p_A(t)$  oznacza prawdopodobieństwo, że w chwili  $t$  obserwujemy nukleotyd  $A$  ( $t = 0, 1, 2, \dots$ ). Ta definicja wymaga podania czy w chwili początkowej  $t_0 = 0$  nukleotyd  $A$  był obserwowany. Przyjmijmy narazie, że  $p_A(0) = 1$ . Otrzymujemy następujące rekurencyjne równanie:

$$p_A(t + 1) = (1 - 3\alpha)p_A(t) + \alpha(1 - p_A(t)).$$

Zatem zmiana prawdopodobieństwa przy przejściu od  $t$  do  $t + 1$  wynosi

$$p_A(t + 1) - p_A(t) = -4\alpha p_A(t) + \alpha.$$

Przechodząc w powyższym równaniu do czasu ciągłego przyjmujemy, że dla małej chwili czasowej  $\Delta t$  zmiana prawdopodobieństwa przy przejściu od  $t$  do  $t + \Delta t$  zależy proporcjonalnie od długości przyrostu czasu (czyli od  $\Delta t$ ):

$$p_A(t + \Delta t) - p_A(t) = \Delta t \cdot [-4\alpha p_A(t) + \alpha] + o(\Delta t),$$

gdzie  $o(\Delta t)$  jest wielkością, która asymptotycznie jest rzędu mniejszego niż  $\Delta t$ , tzn.  $\lim_{\Delta t \rightarrow 0} o(\Delta t)/\Delta t = 0$ .

Dzieląc obie strony przez  $\Delta t$  i przechodząc z  $\Delta t$  do zera otrzymujemy następujące równanie różniczkowe

$$\frac{dp_A(t)}{dt} = -4\alpha p_A(t) + \alpha$$

z warunkiem początkowym  $p_A(0) = 1$ .

Jest to równanie różniczkowe postaci

$$\frac{dx(t)}{dt} = a \cdot x(t) + b \quad (a \neq 0)$$

z warunkiem początkowym  $x(t_0) = x_0$ . Takie równanie ma dokładnie jedno rozwiązanie wyrażające się wzorem

$$x(t) = -\frac{b}{a} + \left(x_0 + \frac{b}{a}\right)e^{a(t-t_0)}.$$

Podstawiając odpowiednie wartości dostajemy rozwiązanie naszego równania:

$$p_A(t) = \frac{1}{4} + \frac{3}{4} \cdot e^{-4\alpha t}.$$

Zauważmy, że wybierając warunek początkowy  $p_A(0) = 0$  dostajemy rozwiązanie

$$p_A(t) = \frac{1}{4} - \frac{1}{4} \cdot e^{-4\alpha t}.$$

Warunek  $p_A(0) = 0$  oznacza, że w chwili początkowej był nukleotyd  $G$ ,  $C$  lub  $T$ . Tak więc takie samo równanie dostaniemy, gdy początkowo był  $G$ , jak i gdy początkowo był  $C$  lub też  $T$ . Zauważmy również, że takie same równania dostaniemy, gdy w powyższych wzorach zastąpimy  $A$  przez jakąkolwiek literę  $G, C, T$ .

Wprowadzając zapis  $p_{ij}(t)$  na prawdopodobieństwo tego, że w chwili  $t$  obserwujemy nukleotyd  $j$ , przy założeniu że w chwili 0 obserwowany był  $i$ , otrzymujemy następujące równania.

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

Mamy oczywiście, dla dowolnych  $i$  oraz  $t$ ,

$$\sum_{j \in \{A, G, C, T\}} p_{ij}(t) = 1$$

Zauważmy również, że

$$\lim_{t \rightarrow \infty} p_{ij}(t) = \frac{1}{4}.$$

Zatem w stanie równowagi otrzymujemy jednostajny rozkład prawdopodobieństwa na zbiorze nukleotydów.

## 6.2 Model Kimury (1980)

Ze względu na rodzaj zasad nukleotydy dzielą się na dwie grupy: *puryny* ( $A, G$ ; są to związki dwupierścieniowe) oraz *pirymidyny* ( $C, T$ ; związki jednopierścieniowe). Mamy wówczas dwa rodzaje podstawień: *tranzycje* (przeprowadzają puryny na puryny oraz pirymidyny na pirymidyny), oraz *transwersje* (przeprowadzają puryny na pirymidyny oraz pirymidyny na puryny). Swierdzono, że tranzycje występują częściej niż transwersje. Niech  $\alpha$  będzie prawdopodobieństwem zmiany nukleotydu  $X$  na  $Y$ , gdzie  $X \neq Y$  są jednocześnie purynami lub pirymidynami. Podobnie niech  $\beta$  będzie prawdopodobieństwem przeprowadzenia puryny na pirymidynę lub na odwrót. Z powyższej uwagi wynika, że możemy przyjąć iż  $\alpha > \beta$ .

Mamy następujące równania rekurencyjne

$$\begin{aligned} p_A(t+1) &= (1 - \alpha - 2\beta)p_A(t) + \beta p_T(t) + \beta p_C(t) + \alpha p_G(t), \\ p_T(t+1) &= \beta p_A(t) + (1 - \alpha - 2\beta)p_T(t) + \alpha p_C(t) + \beta p_G(t), \\ p_C(t+1) &= \beta p_A(t) + \alpha p_T(t) + (1 - \alpha - 2\beta)p_C(t) + \beta p_G(t), \\ p_G(t+1) &= \alpha p_A(t) + \beta p_T(t) + \beta p_C(t) + (1 - \alpha - 2\beta)p_G(t). \end{aligned}$$

Dodatkowo mamy warunki początkowe  $p_X(0) = 1$ , gdzie  $X$  jest jednym z nukleotydów  $\{A, C, G, T\}$ , oraz  $p_Y(0) = 0$ , dla  $Y \neq X$ .

Równania te prowadzą, podobnie jak wcześniej, do układu równań różniczkowych. Podobnie jak wcześniej przyjmijmy oznaczenie  $p_{ij}(t)$  na prawdopodobieństwo tego, że nukleotyd  $j$  był obserwowany w chwili  $t$  przy założeniu, że  $i$  był obserwowany w chwili 0. Ze względu na symetrię modelu mamy tylko trzy klasy rozwiązań:

$$\begin{aligned} p_{ident}(t) &= p_{ii}(t), & \text{gdzie } i \in \{A, C, G, T\} \\ p_{tranzycja}(t) &= p_{ij}(t), & \text{gdzie } i \rightarrow j \text{ jest tranzycją} \\ p_{transwersja}(t) &= p_{ij}(t), & \text{gdzie } i \rightarrow j \text{ jest transwersją.} \end{aligned}$$

Można pokazać, że rozwiązania te przyjmują następującą postać:

$$\begin{aligned} p_{ident}(t) &= \frac{1}{4} + \frac{1}{4} \cdot e^{-4\beta t} + \frac{1}{2} \cdot e^{-2(\alpha+\beta)t}, \\ p_{tranzycja}(t) &= \frac{1}{4} + \frac{1}{4} \cdot e^{-4\beta t} - \frac{1}{2} \cdot e^{-2(\alpha+\beta)t}, \\ p_{transwersja}(t) &= \frac{1}{4} - \frac{1}{4} \cdot e^{-4\beta t}. \end{aligned}$$

Zauważmy, że  $p_{ident}$  jest funkcją malejącą, a  $p_{transwersja}$  jest rosnącą. Natomiast  $p_{tranzycja}$  nie jest monotoniczna (o ile  $\alpha > \beta$ ). Mamy  $p_{tranzycja}(0) = 0$  oraz  $\lim_{t \rightarrow \infty} p_{tranzycja}(t) = 1/4$ . Ponadto dla  $\alpha > \beta$ ,  $p_{tranzycja}(t) > 1/4$ , o ile  $t > \ln 2 / (\alpha - \beta)$ .

Oczywiście, podobnie jak w modelu Jukes-Cantora, w granicy (w stanie równowagi) otrzymujemy rozkład jednostajny na zbiorze nukleotydów.