MO640 - Biologia Computacional
Modelos de Evolução

André Atanasio Maranhão Almeida

Ata de aula: 13/09/2004

Resumo

Nesta aula foram abordados dois modelos de evolução, Jukes-Cantor e Kimura. Roberto Hiroshi Higa, José Augusto Amgarten Quitzau, George Barreto Pereira Bezerra, Leandro Rodrigues Magalhães Marco, Pedro de Vasconcellos Castro e Cleber Valgas Gomes Mira foram convidados a expor o assunto.

1 Introdução

Com o entendimento da forma como funciona a variação de moléculas como o DNA e os peptídeos, surgiu o interesse e a necessidade de modelar matematicamente o processo de evolução de moléculas, no qual um molécula ancestral sofre mutações até transformar-se na molécula observada. O principal objetivos quando se estuda modelos de evolução é quantificar alterações. Aqui dois modelos foram abordados, o modelo de evolução de Jukes-Cantor e o modelo de 2 parâmetros de Kimura.

2 O Modelo de Evolução de Jukes-Cantor

Modelo proposto por T. H. Jukes e C. Cantor em 1969. Este é o modelo mais simples de evolução, onde trata-se apenas um parâmetro, que se refere a probabilidade de um nucleotídeo sofrer mutação em um intervalo de tempo Δt. Tal parâmetro é denotado por α e é proporcional ao tempo transcorrido entre uma molécula ancestral e a obtenção da molécula descendente. Neste modelo, α é diretamente proporcional ao tamanho de Δt, sendo sua relação com o intervalo de tempo dada por $\alpha = u\Delta t$, com u constante.

A partir de α deseja-se obter a probabilidade de uma posição em uma seqüência conter um nucleotídeo b num instante t qualquer, dado que continha um nucleotídeo a no instante t_0. Considera-se t_0 o começo do processo evolutivo e a função $Y(t)$ estima a probabilidade no caso $b \neq a$. É definida também uma função $X(t)$ que estima a probabilidade no caso $b = a$.

Note que não haver mutação e não haver mudança de base em um determinado intervalo de tempo são situações diferentes. Na primeira situação tem-se que em nenhum momento do intervalo de tempo houve alteração na base de determinada posição da seqüência. Já na segunda, durante todo o intervalo de tempo podem ter ocorrido diversas alterações na base de determinado posição da seqüência, porém no fim do intervalo há uma alteração que faz com que a posição volte a possuir a mesma base que no início.

Para se chegar as equações que definem $X(t)$ e $Y(t)$ são consideradas duas situações. Na primeira o nucleotídeo de determinada posição da seqüência no início é igual ao nucleotídeo no fim do intervalo de tempo. Já na segunda, o nucleotídeo de determinada posição da seqüência no início é diferente do nucleotídeo no fim do intervalo de tempo.

Ao se analisar os possíveis cenários para a primeira situação ($A \rightarrow A \rightarrow A$, $A \rightarrow C \rightarrow A$, $A \rightarrow G \rightarrow A$ e $A \rightarrow T \rightarrow A$) chega-se a Equação 1. Ressalta-se que todas estas são situações tendo como origem A. Há situações análogas tendo-se como origem C, G e T.

$$\frac{X(t+\Delta t) - X(t)}{\Delta t} = -3uX(t) + 3uY(t) \tag{1}$$

Ao se analisar os possíveis cenários para a segunda situação ($A \rightarrow A \rightarrow T$, $A \rightarrow C \rightarrow T$, $A \rightarrow G \rightarrow T$ e $A \rightarrow T \rightarrow T$) chega-se a Equação 2.

$$\frac{Y(t+\Delta t) - Y(t)}{\Delta t} = uX(t) - uY(t) \tag{2}$$

Quando Δt tende a zero, obtemos o sistema de equações diferenciais de primeira ordem apresentado na Equação 3.

$$\begin{cases} \frac{dX(t)}{dt} = -3uX(t) + 3uY(t) \\ \frac{dY(t)}{dt} = uX(t) - uY(t) \end{cases} \tag{3}$$

A solução do sistema de equações obedecendo às condições iniciais $X(t_0) = 1$ e $Y(t_0) = 0$, para $t_0 = 0$, é apresentado a seguir.

$$\begin{array}{l} X(t) = \frac{1}{4} + \frac{3}{4}e^{-4ut} \\ Y(t) = \frac{1}{4} - \frac{1}{4}e^{-4ut} \end{array}$$

Note que temos $X(t) + 3Y(t) = 1$ no modelo Jukes-Cantor.

3 O Modelo de 2 Parâmetros de Kimura

O modelo de Jukes-Cantor é simples e fácil de usar, porém pelo que se observa não reflete fielmente a realidade. Foi observado que certos tipos de mutações ocorrem com maior freqüência devido a semelhanças na estrutura química entre os nucleotídeos. De acordo com estas semelhanças os quatro tipos de nucleotídeos que formam as moléculas de DNA podem ser agrupados como *purinas* ou *pirimidinas*. No primeiro grupo encontra-se a *Adenina* (A) e a *Guanina* (G). Já no segundo grupo encontra-se a *Timina* (T) e a *Citosina* (C). Constatou-se que há uma facilidade maior de troca de nucleotídeo por um do mesmo grupo, *transição*, que uma troca de nucleotídeo por um de outro grupo, *transversão*.

Em 1980, Kimura propôs um modelo probabilístico que leva em consideração o fato de mutações poderem ser subdivididas em transições e transversões e que as transições ocorrem com maior freqüência. Kimura estabelece que a probabilidade de ocorrer uma transição num intervalo de tempo fixo Δt, em uma determinada posição de uma seqüência de nucleotídeos é α, o qual é proporcional ao tamanho de Δt e possui a seguinte relação $\alpha = u\Delta t$. É estabelecido também um segundo parâmetro, que determina a probabilidade de ocorrer uma transversão num intervalo de tempo Δt. Este parâmetro tem valor β e possui a seguinte relação com o tamanho do intervalo de tempo $\beta = v\Delta t$. Tanto u quanto v são constantes.

Neste modelo são definidas três funções como segue:

$X(t)$ determina a probabilidade de ao fim de um intervalo de tempo t a posição de uma seqüência possuir o mesmo nucleotídeo que apresentava no instante t_0.

$Y(t)$ determina a probabilidade de ao fim de um intervalo de tempo t a posição de uma seqüência resultar num nucleotídeo diferente do que era apresentado no instante t_0. Além disso, os nucleotídeos diferentes pertencem a um mesmo grupo, caracterizando uma *transição*.

$Z(t)$ determina a probabilidade de ao fim de um intervalo de tempo t a posição de uma seqüência possuir um nucleotídeo diferente do que era apresentado no instante t_0. Além disso, os nucleotídeos diferentes pertencem a grupos diferentes, caracterizando uma *transversão*.

Assim como no modelo anterior, analisando-se cada uma das situações, no caso três, e seus respectivos cenários, chega-se a três equações diferenciais que são dispostas no sistema apresentado a seguir.

$$\begin{cases} \frac{dX(t)}{dt} = -(u+2v)X(t) + uY(t) + 2vZ(t) \\ \frac{dY(t)}{dt} = uX(t) - (u+2v)Y(t) + 2vZ(t) \\ \frac{dZ(t)}{dt} = vX(t) + vY(t) - 2vZ(t) \end{cases}$$

Resolvendo-se o sistema com as condições iniciais $X(t_0) = 1$, $Y(t_0) = Z(t_0) = 0$, $t_0 = 0$, obtemos o seguinte.

$$\begin{aligned} X(t) &= \tfrac{1}{4} + \tfrac{1}{4}e^{-4vt} + \tfrac{1}{2}e^{-2(u+v)t} \\ Y(t) &= \tfrac{1}{4} + \tfrac{1}{4}e^{-4vt} - \tfrac{1}{2}e^{-2(u+v)t} \\ Z(t) &= \tfrac{1}{4} - \tfrac{1}{4}e^{-4vt} \end{aligned}$$

O modelo implica na seguinte equação.

$$X(t) + Y(t) + 2Z(t) = 1$$