ÁRBOLES FILOGENÉTICOS

SEMANA 04

27 de Abril del 2020

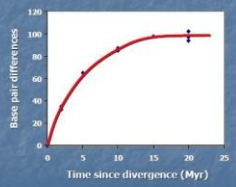
Índice

- Modelos de Arboles Filogenéticos
- Comparación de Modelos

1 ÁRBOLES FILOGENÉTICOS

Midiendo el cambio evolutivo

- Medida simple: Contar el número de sitios diferentes .
- Estimador muy inexacto:
  - Sitios pueden tener sustituciones repetidas.
  - Divergencia de secuencias llega a ser menos exacta en su estimación



Modelo Jukes-Cantor (JC)

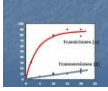
- Asume que las cuatro bases tienen igual frecuencia y que las sustituciones son igualmente probables.

$$P_J = \begin{bmatrix} - & \alpha & \alpha & \alpha \\ \alpha & - & \alpha & \alpha \\ \alpha & \alpha & - & \alpha \\ \alpha & \alpha & \alpha & - \end{bmatrix} \qquad f = [¼\ ¼\ ¼\ ¼]$$

Modelo Kimura de 2 parámetros (K2P)

- Toma en cuenta diferencias entre transiciones vs. transversiones

$$P_J = \begin{bmatrix} - & \beta & \alpha & \beta \\ \beta & - & \beta & \alpha \\ \alpha & \beta & - & \beta \\ \beta & \alpha & \beta & - \end{bmatrix} \qquad f = [¼\ ¼\ ¼\ ¼]$$



Existen varios modelos que estiman la tasa de cambio entre las secuencias de nucleótidos

Felsenstein (1981) (F81)

- Toma en cuenta diferencias en la composicion de las bases.
- Porcentaje (G + C) puede variar entre 25% - 75%.
- F81 permite que la frecuencia de cada nucleotido sea diferente. No permite variación en las frecuencias entre genes y especies.

$$P_J = \begin{bmatrix} - & \pi_C(1-e^{-\beta t}) & \pi_G(1-e^{-\beta t}) & \pi_T(1-e^{-\beta t}) \\ \pi_A(1-e^{-\beta t}) & - & \pi_G(1-e^{-\beta t}) & \pi_T(1-e^{-\beta t}) \\ \pi_A(1-e^{-\beta t}) & \pi_C(1-e^{-\beta t}) & - & \pi_T(1-e^{-\beta t}) \\ \pi_A(1-e^{-\beta t}) & \pi_C(1-e^{-\beta t}) & \pi_G(1-e^{-\beta t}) & - \end{bmatrix}$$

$$f = [\pi_A\ \pi_C\ \pi_G\ \pi_T]$$

Hasegawa, Kishino y Yano (1985) (HKY85)

- Esencialmente mezcla modelos K2P and F81 , permitiendo la ocurrencia de transiciones y transversiones a distintas tasas y a su vez permitiendo que la frecuencia de bases varie.

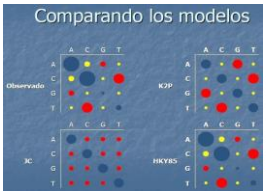
$$P_J = \begin{bmatrix} - & \pi_C\beta & \pi_G\alpha & \pi_T\beta \\ \pi_A\beta & - & \pi_G\beta & \pi_T\alpha \\ \pi_A\alpha & \pi_C\beta & - & \pi_T\beta \\ \pi_A\beta & \pi_C\alpha & \pi_G\beta & - \end{bmatrix} \qquad f = [\pi_A\ \pi_C\ \pi_G\ \pi_T]$$

Los modelos también pueden estimar la frecuencia de los nucleótidos

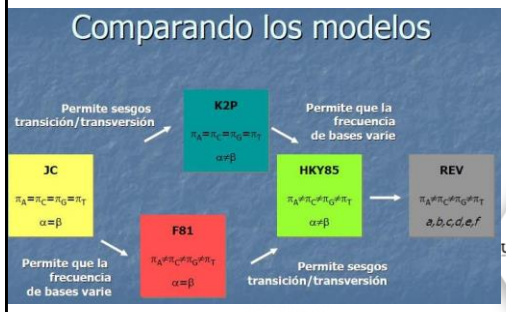
modelo General reversible (REV)

- Modelo más general – cada sustitución tiene su propia probabilidad.

$$P_t = \begin{bmatrix} - & \pi_C a & \pi_G b & \pi_T c \\ \pi_A a & - & \pi_G d & \pi_T e \\ \pi_A b & \pi_C d & - & \pi_T f \\ \pi_A c & \pi_C e & \pi_G f & - \end{bmatrix} \qquad f = [\pi_A \; \pi_C \; \pi_G \; \pi_T]$$

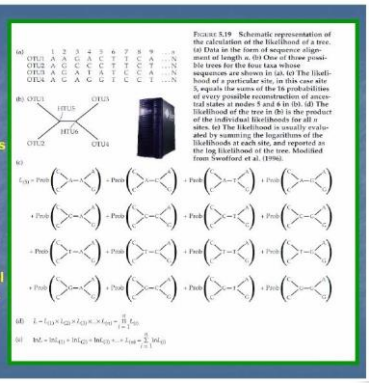


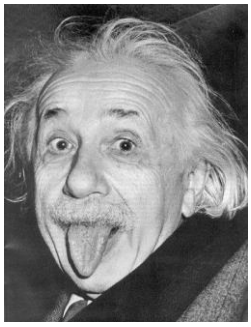
Se debe escoger el modelo que mejor se ajuste a los datos



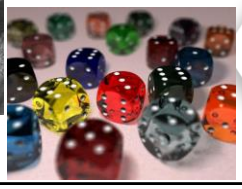
Existen en la actualidad modelos de sustitución nucleotídica

1. Calcule la probabilidad para cada sitio.
2. Sume los valores de L para todos los sitios en el árbol.
3. Compare los valores de L para todos los arboles posibles.
4. Elija el arbol con el valor mas alto de L.



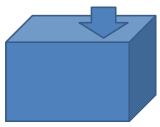


The theory says a lot, but does not really bring us any closer to the secret of the '***old one***'. I, at any rate, am convinced that ***He*** **does not throw dice.**



Jugando a los dados

- 90 dados normales + 10 dados trucados



Observation	Fair	Biased
⚀	1/6	1/21
⚁	1/6	2/21
⚂	1/6	3/21
⚃	1/6	4/21
⚄	1/6	5/21
⚅	1/6	6/21

- Probabilidad de sacar un dado trucado? ❼ hipótesis=trucado

$$\Pr(trucado) = \frac{10}{90+10} = 0.1$$

- Sacamos y lanzamos 2 dados: 4 y 6

- Probabilidad de que no estén trucados:

$$\Pr(4y6|buenos) = \frac{1}{6}x\frac{1}{6} = \frac{1}{36}$$

0,0278

- Probabilidad de que estén trucados:

$$\Pr(4y6|trucados) = \frac{4}{21}x\frac{6}{21} = \frac{24}{441}$$

0,0544

.: Dados trucados

Observation	Fair	Biased
⚀	1/6	1/21
⚁	1/6	2/21
⚂	1/6	3/21
⚃	1/6	4/21
⚄	1/6	5/21
⚅	1/6	6/21

Probabilidad de los datos dado el modelo, i.e. el LIKELIHOOD

Hallando la probabilidad posterior: Probabilidad del modelo dados los datos

$$\Pr(trucado) = \frac{10}{90+10} = 0.1 \qquad \Pr(4y6|trucados) = \frac{4}{21} x \frac{6}{21} = \frac{24}{441}$$

$$\Pr(trucados|4y6) = \frac{\Pr(trucados)\ \Pr(4y6|trucados)}{\Pr\ (4y6)}$$

$$\Pr(4y6|trucados)\ x0.1 + \Pr(4y6|buenos)\ x0.9$$

$$\Pr(4y6|trucados) = \frac{4}{21} x \frac{6}{21} = \frac{24}{441} \qquad \Pr(4y6|buenos) = \frac{1}{6} x \frac{1}{6} = \frac{1}{36}$$

$$\Pr(trucados|4y6) = 0.17877 > \text{P (trucados)= 0.10}$$

Conclusiones

- ✓ Modelos que estiman la tasa de cambio entre las secuencias de nucleótidos (JP y K2P).
- ✓ Los modelos también pueden estimar la frecuencia de los nucleótidos (F81 y HKY85).
- ✓ Probabilidad de modelos de sustitución nucleotídica.

Gracias

ERES MÁS QUE CAPAZ DE ENCONTRAR EL ÉXITO, PERO SOLO OCURRIRÁ SI TE PONES A ELLO !!