

PEC 2: BIOLOGIA MOLECULAR

DIEGO VÁZQUEZ ZAMBRANO

PREGUNTA 2

a.

Original: MARIACENALATAS.LVGS

1. MA.IACENALATAS.LVGS
2. MARIACENALATAS.LV..
3. MARTACENARATAS.LV.K
4. MARIACENALATAS.LD.K
5. MARTACANALATGQAN.LD
6. MARTACENARATAS.LV.K
7. MARTACATALATAS.LV..
8. MARTACENALATAS.LD..

El orden de menor a mayor efecto sería $(2=4) < 8 < (3=6) < 7 < 5 < 1$

b.

Lo primero que hemos hecho es determinar qué secuencia de aminoácidos determina la secuencia Original: MARIACENALATAS.LVGS

Luego vemos las secuencias de nucleótidos (nt):

1. MA.IACENALATAS.LVGS
2. MARIACENALATAS.LV..
3. MARTACENARATAS.LV.K
4. MARIACENALATAS.LD.K
5. MARTACANALATGQAN.LD
6. MARTACENARATAS.LV.K
7. MARTACATALATAS.LV..
8. MARTACENALATAS.LD..

0.	ATG	GCA	AGA	ATA	GCT	TGC	GAA	AAC	GCA	CTC	GCA	ACG	GCA	AGC	TAA	TTA	GTT	GGA	TCA
1.	ATG	GCA	TGA	ATA	GCT	TGC	GAA	AAC	GCA	CTC	GCA	ACG	GCA	AGC	TAA	TTA	GTT	GGA	TCA
2.	ATG	GCC	AGA	ATC	GCT	TGC	GAA	AAC	GCA	CTC	GCA	ACA	GCA	AGC	TAA	TTA	GTC	TGA	TAA
3.	ATG	GCA	AGA	ACA	GCT	TGC	GAA	AAC	GCA	CGC	GCA	ACG	GCA	AGC	TAA	TTA	GTT	TGA	AAA
4.	ATG	GCC	CGA	ATC	GCT	TGC	GAG	AAC	GCA	CTT	GCA	ACA	GCA	AGT	TAA	TTG	GAT	TGA	AAA
5.	ATG	GCC	AGA	ACA	GCT	TGC	GCA	AAC	GCA	CTC	GCC	ACG	GGG	CAA	GCT	AAT	TAG	TTG	GAT
6.	ATG	GCA	AGG	ACA	GCT	TGT	GAA	AAC	GCA	CGC	GCA	ACG	GCG	AGC	TAA	TTA	GTT	TGA	AAA
7.	ATG	GCA	AGA	ACA	GCT	TGC	GCA	ACC	GCA	CTC	GCA	ACA	GCA	AGC	TAA	TTA	GTT	TGA	TAA
8.	ATG	GCA	AGA	ACA	GCT	TGC	GAA	AAC	GCA	CTC	GCA	ACG	GCA	AGC	TAA	TTA	GAT	TGA	TAA

En verde están subrayados los cambios sinónimos y en magenta cuando generan un cambio de aminoácidos. En naranja se han subrayado los cambios que generan un codón STOP, y en gris la secuencia posterior a ese codón que ya no la leemos.

Lo importante son los cambios en los aminoácidos que podemos observar en la parte a y b de la pregunta.

c.

La secuencia de aminoácidos codificada por la secuencia nucleotídica 0 es:

MARIACENALATAS.LVGS

Para MARIACENALATAS:

$$1 \times 4 \times 6 \times 4 \times 4 \times 2 \times 2 \times 2 \times 4 \times 6 \times 4 \times 4 \times 4 \times 6 = 967,680$$

Para MARIACENALATAS.LVGS:

$$1 \times 4 \times 6 \times 4 \times 4 \times 2 \times 2 \times 2 \times 4 \times 6 \times 4 \times 4 \times 4 \times 6 \times 6 \times 4 \times 4 \times 6 = 4,031,232$$

PREGUNTA 3

a.

En la población experimental de 100 individuos, tenemos 70 moscas de la cepa 1 (AA) y 30 moscas de la cepa 2 (aa).

Para calcular las frecuencias génicas (p y q) y las frecuencias genotípicas (P, H y Q) en la siguiente generación, podemos aplicar el equilibrio de Hardy-Weinberg.

Las frecuencias genotípicas iniciales son:

$$P = 70/100 = 0.7$$

$$H = 0 \text{ (no hay individuos heterocigotos en la población inicial)}$$

$$Q = 30/100 = 0.3$$

$$\text{Calculamos las frecuencias génicas iniciales: } p = P + 0.5H = 0.7 + 0.5(0) = 0.7 \quad q = Q + 0.5H = 0.3 + 0.5(0) = 0.3$$

$$\text{Ahora podemos calcular las frecuencias genotípicas en la siguiente generación: } P' = p^2 = (0.7)^2 = 0.49 \quad H' = 2pq = 2(0.7)(0.3) = 0.42 \quad Q' = q^2 = (0.3)^2 = 0.09$$

Por lo tanto, las frecuencias genotípicas en la siguiente generación serán:

$$P' = 0.49$$

$$H' = 0.42$$

$$Q' = 0.09$$

Para las frecuencias génicas:

$$p' = P' + 0.5H' = 0.49 + 0.5(0.42) = 0.49 + 0.21 = 0.7$$

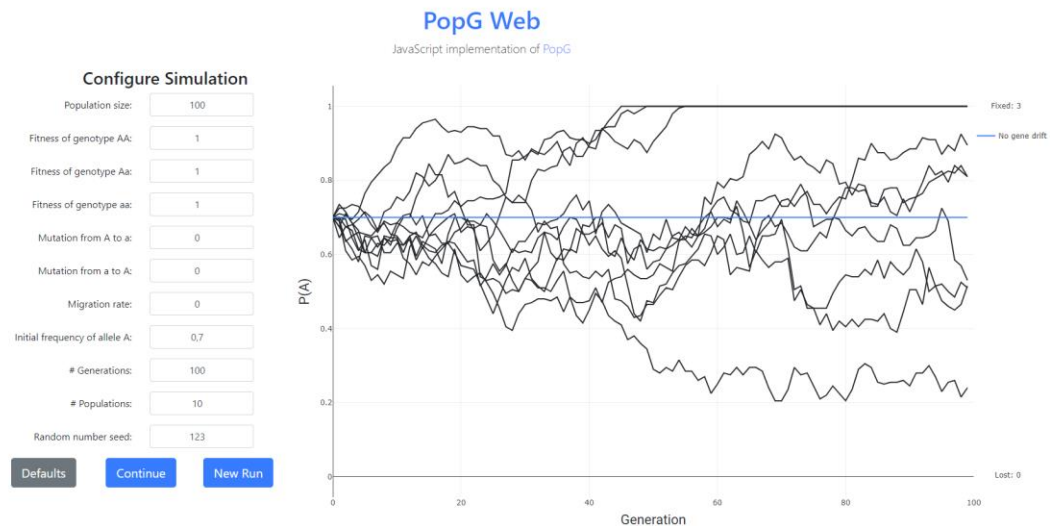
$$q' = Q' + 0.5H' = 0.09 + 0.5(0.42) = 0.09 + 0.21 = 0.3$$

Entonces, las frecuencias génicas en la siguiente generación también serán:

$$p' = 0.7$$

$$q' = 0.3$$

b.



c.

1. Population size (Tamaño de la población):

- Valor: 100

- Explicación: Se elige un tamaño de población de 100 individuos para simular una población de tamaño moderado. Este tamaño permite que haya variabilidad genética y suficientes individuos para observar posibles cambios en las frecuencias génicas a lo largo de las generaciones.

2. Fitness of genotype AA (Aptitud del genotipo AA):

- Valor: 1

- Explicación: Se establece una aptitud de 1 para el genotipo AA, lo que significa que no hay diferencias en la supervivencia o reproducción de los individuos con este genotipo en comparación con los demás genotipos. Esto se basa en la información previa mencionada en el enunciado, donde se indica que el gen en estudio no tiene efectos sobre la eficacia biológica.

3. Fitness of genotype Aa (Aptitud del genotipo Aa):

- Valor: 1

- Explicación: Al igual que en el caso anterior, se establece una aptitud de 1 para el genotipo Aa, lo que implica que no hay diferencias en la supervivencia o reproducción de los individuos heterocigotos en comparación con los demás genotipos.

4. Fitness of genotype aa (Aptitud del genotipo aa):

- Valor: 1

- Explicación: Se asigna una aptitud de 1 para el genotipo aa, indicando nuevamente que no hay diferencias en la supervivencia o reproducción de los individuos con este genotipo en relación con los otros genotipos.

5. Mutation from A to a (Mutación de A a a):

- Valor: 0

- Explicación: Se establece que no hay mutación del alelo A al alelo a en esta simulación. Esto implica que los alelos se mantienen constantes y no se introducen nuevos alelos en la población a través de mutaciones.

6. Mutation from a to A (Mutación de a a A):

- Valor: 0

- Explicación: Al igual que en el caso anterior, se establece que no hay mutación del alelo a al alelo A en esta simulación. Se supone que los alelos se mantienen constantes sin cambios mutacionales.

7. Migration rate (Tasa de migración):

- Valor: 0

- Explicación: Se establece una tasa de migración de 0, lo que significa que no hay migración de individuos entre las poblaciones simuladas. Esta configuración asume que la población experimental está aislada y no hay flujo genético proveniente de otras poblaciones.

8. Initial frequency of allele A (Frecuencia inicial del alelo A):

- Valor: 0.7

- Explicación: Se establece una frecuencia inicial del alelo A de 0.7, lo que implica que inicialmente el alelo A es más común en la población. Esta elección refleja la proporción de 70 moscas de la cepa 1 (genotipo AA) de un total de 100 individuos en la población experimental mencionada en el enunciado.

9. Generations (Número

de generaciones):

- Valor: 100

- Explicación: Se establece un total de 100 generaciones para simular el paso del tiempo y observar la evolución de las frecuencias génicas a lo largo del tiempo.

10. Populations (Número de poblaciones):

- Valor: 10

- Explicación: Se simularán 10 poblaciones para observar posibles variaciones entre ellas. Esto permite obtener una idea de la variabilidad y la estabilidad de los resultados a través de diferentes poblaciones simuladas

11. Random number seed (Semilla aleatoria):

- Valor: 123

- Explicación: Se utiliza una semilla aleatoria de valor 123 para permitir la reproducibilidad de los resultados. Al utilizar la misma semilla aleatoria, se pueden replicar los resultados exactos en ejecuciones posteriores del programa GenPop.

Estos valores específicos se eligen para adaptarse a las condiciones planteadas en el enunciado y pueden variar según los objetivos de la simulación y los datos disponibles.

d.

Si después de 25 generaciones en el laboratorio todos los individuos de la población experimental son aa, esto indica que el alelo a ha alcanzado una frecuencia del 100% en la población. Esto puede ocurrir debido a la deriva genética.

La deriva genética es un fenómeno aleatorio que puede tener un impacto significativo en las poblaciones pequeñas, donde las fluctuaciones aleatorias pueden conducir a cambios en las frecuencias génicas a lo largo del tiempo. En este caso, la población experimental inicial tenía 30 individuos de la cepa 2 (genotipo aa) y, debido a la deriva genética, el alelo a se ha vuelto predominante en la población después de 25 generaciones.

La deriva genética es más pronunciada en poblaciones pequeñas debido a que los eventos aleatorios tienen un mayor impacto en la composición genética de la población. En este escenario, es posible que la población experimental haya experimentado una disminución en su tamaño efectivo debido a factores como la reproducción selectiva, la mortalidad diferencial o la falta de apareamiento al azar.

PREGUNTA 6

a.

Para tener una mejor idea del tiempo transcurrido desde la divergencia entre dos secuencias nucleotídicas alineadas, el estadístico más adecuado es K (número de sustituciones sinónimas por sitio). Este estadístico considera las sustituciones de nucleótidos que no alteran el aminoácido codificado, lo que indica cambios neutrales en la secuencia. Estas sustituciones sinónimas tienden a acumularse a lo largo del tiempo a una tasa relativamente constante, lo que nos permite realizar estimaciones del tiempo transcurrido desde la divergencia.

Por otro lado, el estadístico p (número de sustituciones no sinónimas por sitio) tiene en cuenta las sustituciones de nucleótidos que alteran el aminoácido codificado, lo que puede resultar en cambios funcionales o selectivos. Estas sustituciones no sinónimas pueden estar sujetas a fuerzas evolutivas más complejas y variables, como la selección natural, por lo que no proporcionan una estimación confiable del tiempo transcurrido desde la divergencia.

En resumen, el estadístico K es más adecuado para estimar el tiempo transcurrido desde la divergencia entre dos secuencias nucleotídicas alineadas, ya que considera cambios neutrales que se acumulan de manera constante.

b.

Cuando se comparan secuencias codificadoras de proteínas y se desea reconstruir un árbol filogenético, el estadístico más apropiado es p (número de sustituciones no sinónimas por sitio). Esto se debe a que las sustituciones no sinónimas pueden reflejar cambios funcionales y adaptativos en las proteínas, lo que puede proporcionar información relevante sobre las relaciones evolutivas entre las secuencias.

Las sustituciones no sinónimas pueden estar sujetas a la selección natural y a la presión evolutiva para conservar ciertas funciones proteicas o adquirir nuevas funciones. Al analizar estas sustituciones no sinónimas en secuencias codificadoras de proteínas, podemos inferir la similitud y divergencia evolutiva entre las especies y construir un árbol filogenético que refleje su relación evolutiva.

En resumen, el estadístico p es más adecuado para reconstruir un árbol filogenético cuando se comparan secuencias codificadoras de proteínas, ya que captura cambios funcionales y adaptativos relevantes en la evolución de las proteínas.

c.

Basándome en la información proporcionada, parece que el estudiante A ha obtenido un puntaje de 0 y el estudiante B ha obtenido un puntaje de 10.

La razón detrás de estas calificaciones se basa en la interpretación de los árboles filogenéticos presentados por cada estudiante.

El estudiante A ha creado un árbol en el que todas las ramas son iguales, lo que implica que no ha tenido en cuenta la información de distancia entre las especies procariotas al construir el árbol. Esto sugiere una falta de consideración por las diferencias evolutivas y la divergencia entre las secuencias del gen ribosomal 16S. Por lo tanto, se le ha otorgado un puntaje de 0 porque no ha logrado representar de manera precisa las relaciones filogenéticas basadas en la secuencia del gen.

Por otro lado, el estudiante B ha tenido en cuenta las diferencias de distancia al construir el árbol, lo que ha llevado a la representación de ramas con diferentes longitudes. Al considerar las diferencias de distancia, el estudiante B ha logrado capturar de manera más precisa las relaciones evolutivas entre las especies procariotas basadas en la secuencia del gen ribosomal 16S. Por lo tanto, se le ha otorgado un puntaje de 10 porque ha representado de manera más precisa las relaciones filogenéticas.

Pregunta 4

a.

Esto se debe a que hay posiciones que no son totalmente sinónimas. Véase el ejemplo del primer aminoácido. En la secuencia A tenemos CTC y en la secuencia B tenemos CTG. En este caso tenemos 2/3 posición sinónima, porque, aunque cambia un nucleótido, no supone un cambio de aminoácido. Al no ser estos números enteros, podemos obtener un número total no entero de posiciones sinónimas.

b.

nº codón 12 33 40 41 52 59 64 66 72 100 101 105 112 125 130 186 216 233 245 246
 secuencia A CTC GGT GCG TCT GCG GGT CCA GGA TCA CAG CAC AAT ATG TTA AAA GGT CCG CGA ATA GAA
 secuencia B CTG GGG GCC AGT GCC GGG CCG GGT TCG CAA CTA AAC ATA TTG AAG GGG TCG AGA ATT GAG

 L G A S A G P G S Q H N M L K G P R I E
 L G A S A G P G S Q L N I L K G S R I E

- 12: 1 sinónimo
- 33: 1 sinónimo
- 40: 1 sinónimo
- 41: 1 sinónimo y 1 no sinónimo
- 52: 1 sinónimo
- 59: 1 sinónimo
- 64: 1 sinónimo
- 66: 1 sinónimo
- 72: 1 sinónimo
- 100: 1 sinónimo
- 101: 1,5 no sinónimo y 0.5 sinónimo
- 105: 1 sinónimo
- 112: 1 no sinónimo
- 125: 1 sinónimo
- 130: 1 sinónimo
- 186: 1 sinónimo
- 216: 1 no sinónimo
- 233: 1 sinónimo
- 245: 1 sinónimo
- 246: 1 sinónimo

Sd=17,5 cambios sinónimos
 Nd=4,5 cambios no sinónimos

c.

Para calcular K deberíamos utilizar la fórmula de Jukes y Cantor: $K = -3/4 \ln (1 - 4/3 p)$.

Para ello primero necesitamos calcular Ps y Pa. En este caso ya tenemos Pa porque nos lo proporciona el enunciado:

Según el enunciado tenemos 520,66 posiciones no sinónimas

Por lo tanto, para obtener el número de posiciones sinónimas hacemos lo siguiente:
 $250 \times 3 = 750$ que es el número máximo de posiciones sinónimas que se pueden obtener.

Ahora hacemos la ecuación para obtener el número de posiciones sinónimas:

$$S = 750 - 520,66 = 229,34$$

$$P_s = 17,5 / 229,34 = 0,07630$$

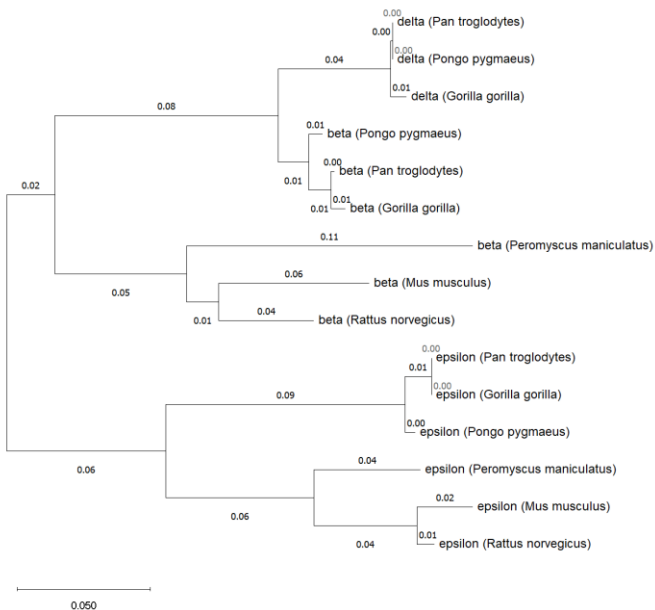
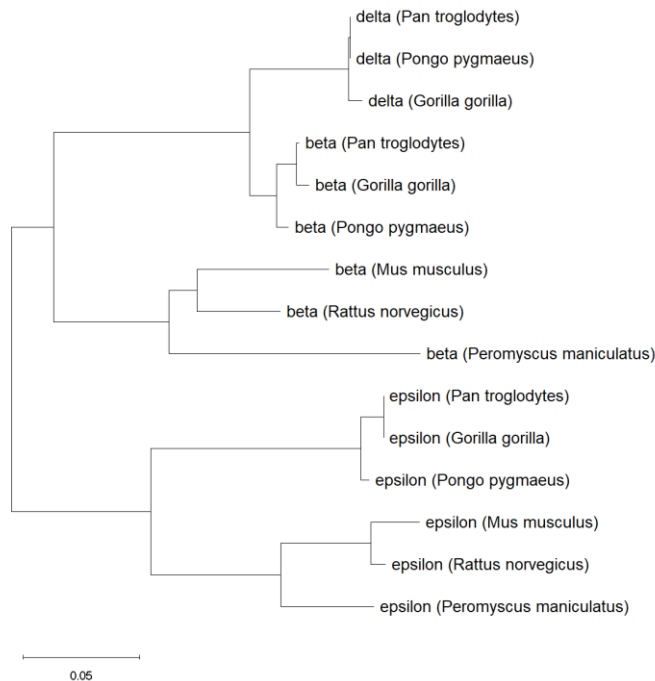
$$P_a = 4,5 / 520,66 = 0,00864$$

$$K_s = -3/4 \ln (1 - 4/3 \times 0,07630) = 0,08046$$

$$K_a = -3/4 \ln (1 - 4/3 \times 0,00864) = 0,008690$$

Pregunta 5

a.



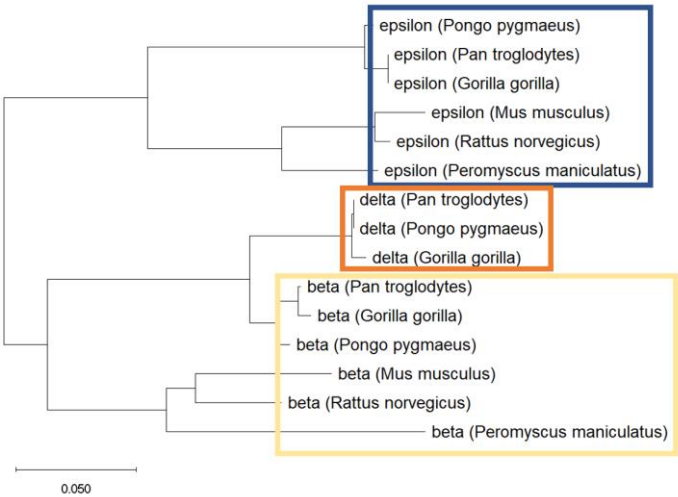
a.

En el árbol filogenético creado utilizando MEGA, el clado que alberga el gen “delta” muestra la menor cantidad de nodos ancestrales compartidos, lo cual indica su origen evolutivo relativamente reciente.

c.

Para esta pregunta he creado una nueva alineación en la cual la secuencia épsilon (*Pongo pygmaeus*) es la primera, para así facilitar la visualización de los ortólogos y parálogos.

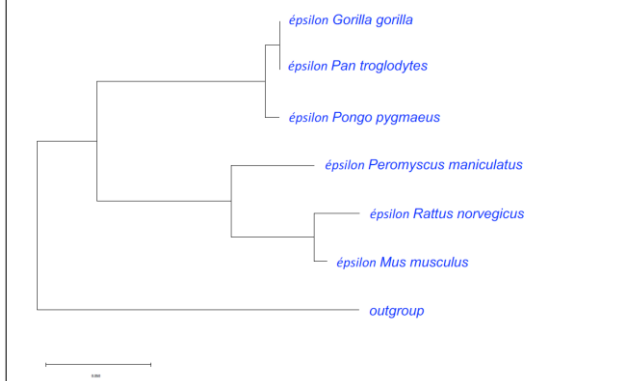
En recuadros de distintos colores se han enmarcado los genes que son ortólogos entre sí. Los ortólogos tienen un ancestro común mientras que los parálogos provienen de un evento de duplicación genética y no pertenecen al mismo clado. Por tanto, todos los que estén en el recuadro amarillo o naranja son parálogos de épsilon (*P. pygmaeus*).



Ortólogos de épsilon (<i>P. pygmaeus</i>)	Parálogos de épsilon (<i>P. pygmaeus</i>)
Épsilon (<i>P. troglodytes</i>)	Delta (<i>P. pygmaeus</i>)
Épsilon (<i>G. gorilla</i>)	Delta (<i>P. troglodytes</i>)
Épsilon (<i>M. musculus</i>)	Delta (<i>G. gorilla</i>)
Épsilon (<i>R. norvegicus</i>)	Beta (<i>P. pygmaeus</i>)
Épsilon (<i>P. maniculatus</i>)	Beta (<i>G. gorilla</i>)
	Beta (<i>P. troglodytes</i>)
	Beta (<i>M. musculus</i>)
	Beta (<i>R. norvegicus</i>)
	Beta (<i>P. maniculatus</i>)

d.

Figura 5.1.



Los nodos representan puntos de divergencia evolutiva y las ramas representan las relaciones de parentesco entre las especies o grupos taxonómicos.

Las ramas principales del árbol representan eventos de especiación, indicando los puntos en los que las especies se dividieron y divergieron en linajes separados. Podemos ver que todos los genes *épsilon* divergieron del outgroup, y a su vez los genes divergieron en los dos órdenes que conocemos como roedores y primates.

El gen *épsilon* divergió primero en los roedores, donde podemos ver dos ramas, una que contiene a *M. musculus* y *R. norvegicus* y otra con *P. maniculatus*.

La diferencia en la longitud de las ramas de los roedores es mayor que aquella de los primates, lo cual nos indica que el porcentaje de homología de la proteína codificada por el gen *épsilon* es mayor entre los primates que entre los roedores. Por ejemplo, es muy significativo como no existe rama para las *épsilon* de *G. gorilla* y *P. troglodytes*, indicando que el evento de divergencia entre los dos es relativamente reciente y que la diferencia genética entre ellos es ínfima.

Como otro dato a notar, ambos grupos contienen dos nodos de divergencia evolutiva. En términos de distancia evolutiva, la *épsilon* de *R. norvegicus* es el que ha evolucionado más recientemente. Es más, todas las *épsilon* del orden primates son anteriores a las *épsilon* de los roedores, en términos de divergencia evolutiva.