

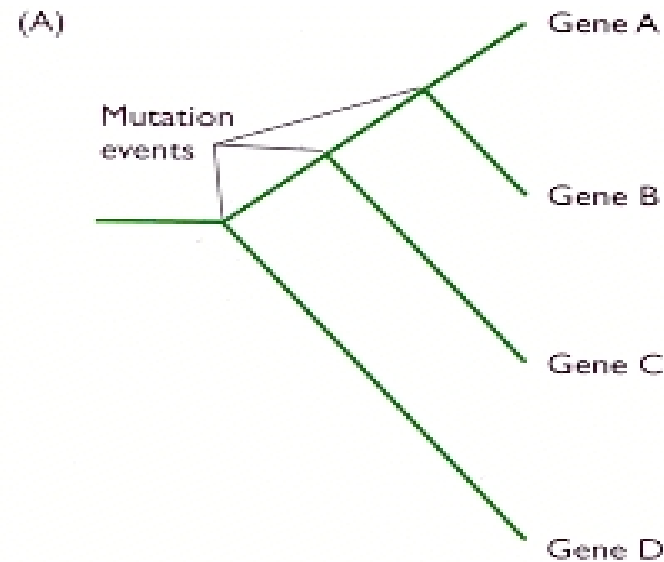
Evolucion molecular y Adaptacion

Curso CEFIRE
Setiembre 2011

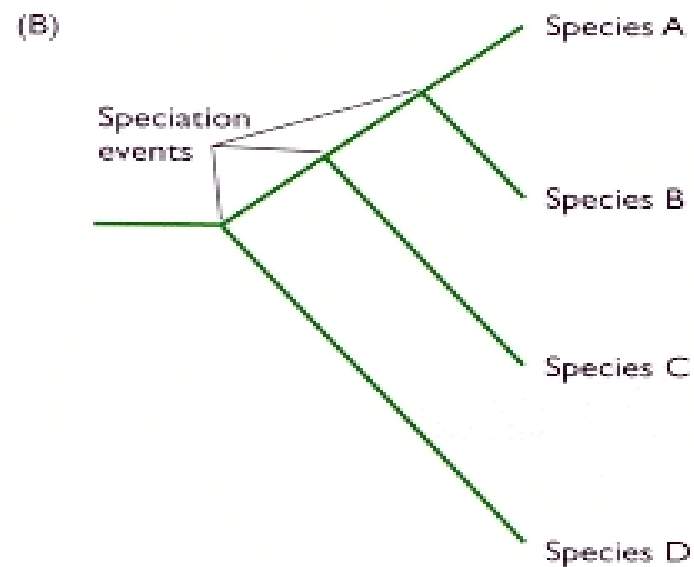
Evolucion molecular

Árbol de genes/especies:

- El resultado de la reconstrucción de una filogenia a partir de secuencias es un árbol de gen
- Existen muchas razones por las cuales árbol de especie y árbol de gen no tienen que coincidir (duplicaciones, alelos polimorficos...)



A gene tree

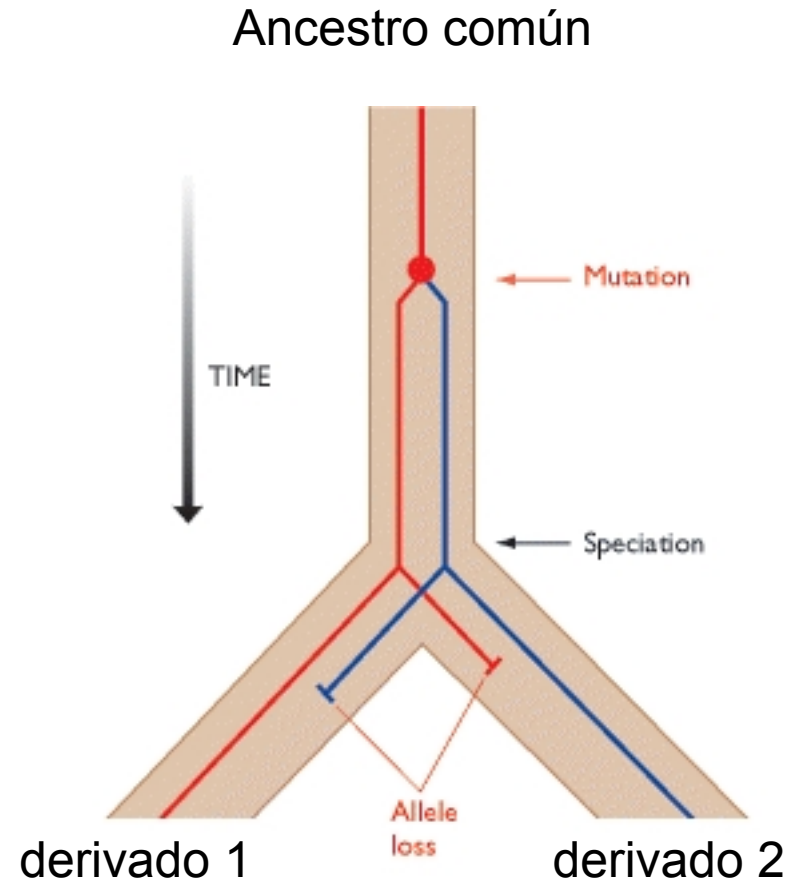


A species tree

Evolucion molecular

Árbol de genes / Árbol de especies

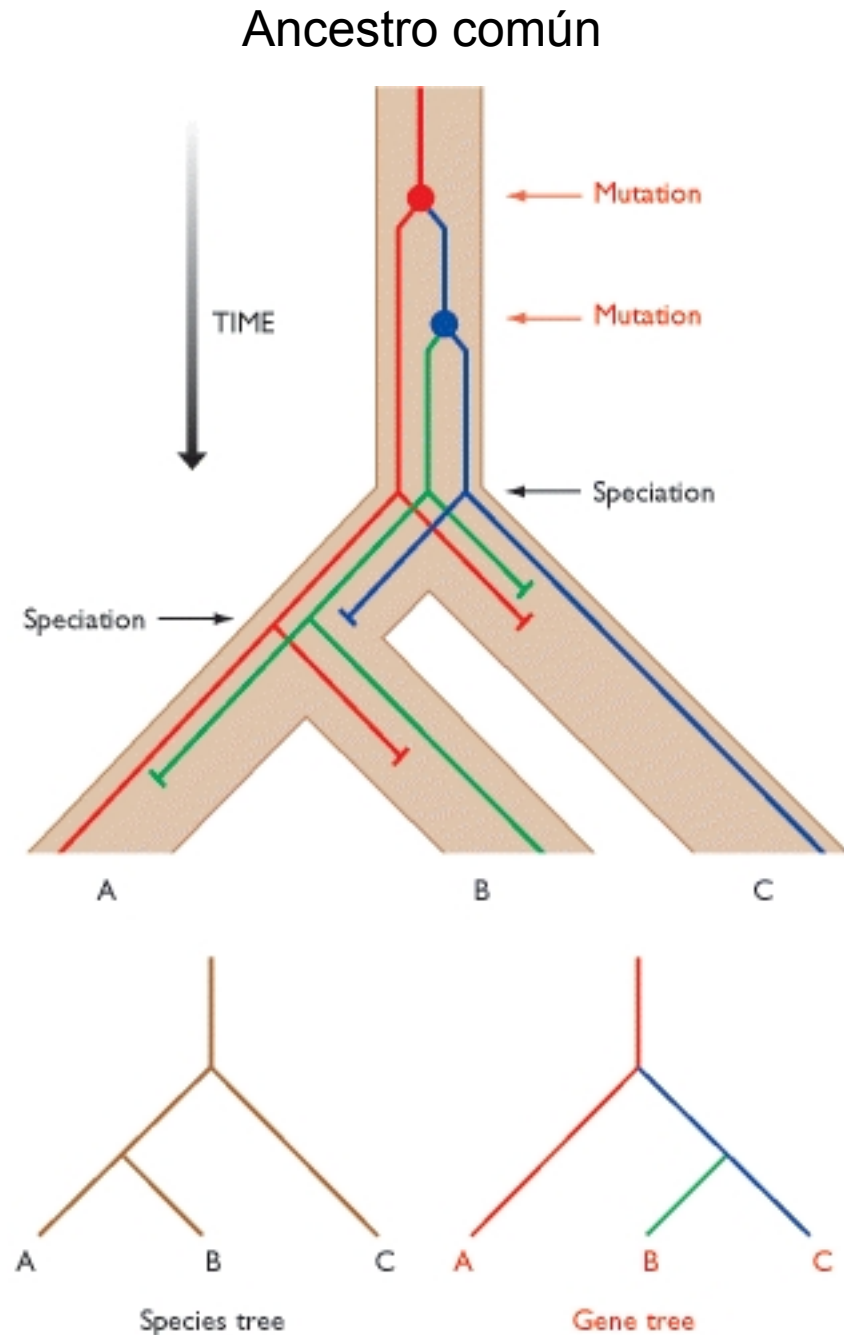
- Misma topología
- Diferentes tiempos de divergencia



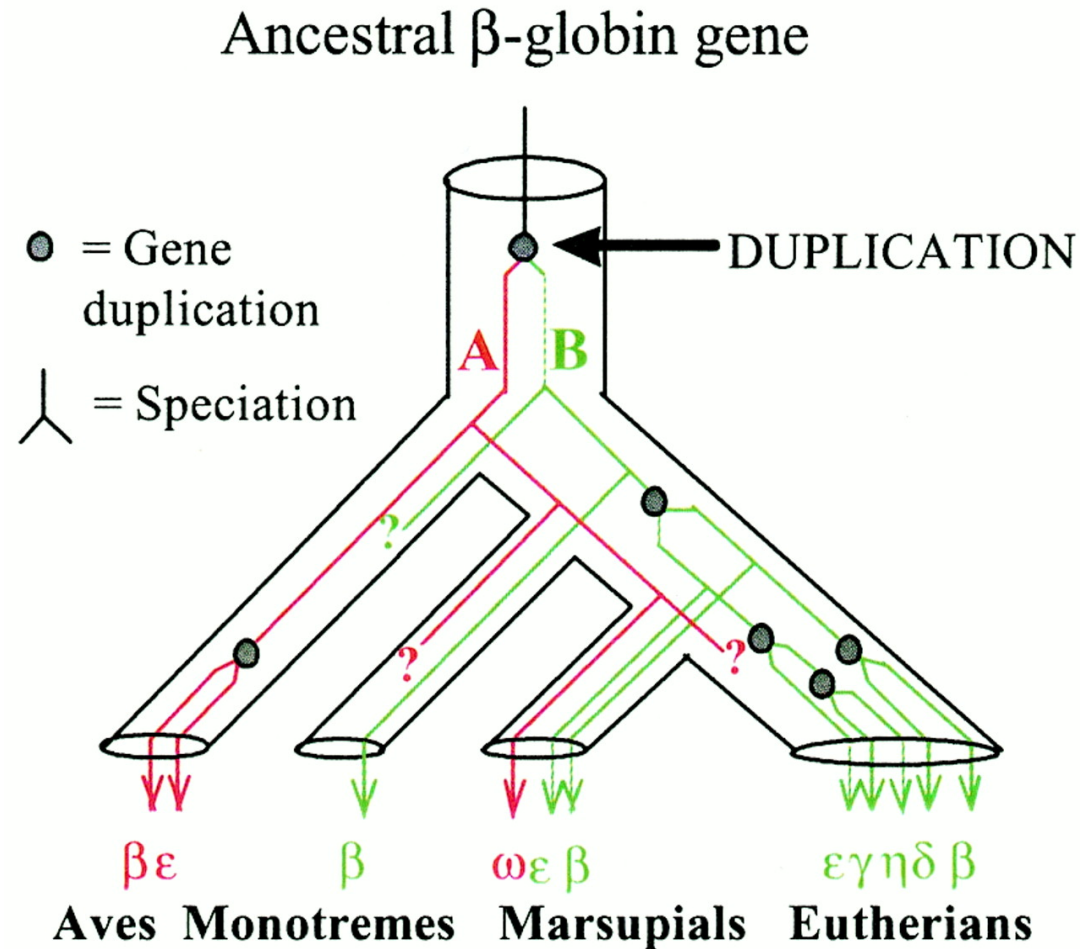
Evolucion molecular

Árbol de genes / Árbol de especies

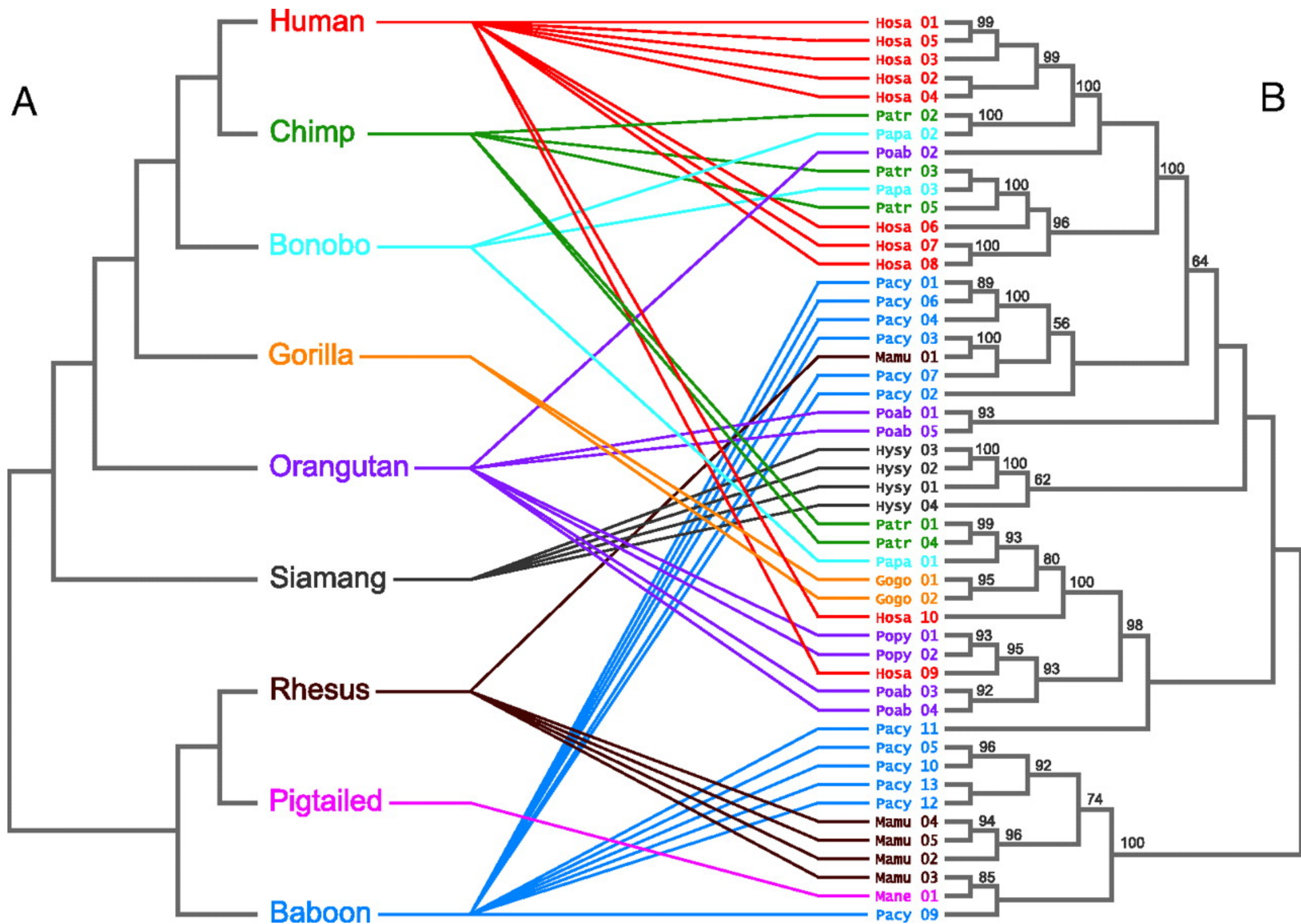
- Diferente topología
- Diferentes tiempos de divergencia



Evolution molecular



□ D. Wheeler et al., "An orphaned mammalian beta-globin gene of ancient evolutionary origin.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 3, pp. 1101-6, Jan. 2001.



D. A. Loisel, M. V. Rockman, G. A. Wray, J. Altmann, and S. C. Alberts, "Ancient polymorphism and functional variation in the primate MHC-DQA1 5' cis-regulatory region.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 44, pp. 16331-6, Oct. 2006.

Evolucion molecular

Reloj molecular

Hipótesis: para una proteína la **tasa de evolución** es aproximadamente constante a lo largo del tiempo y de los linajes.

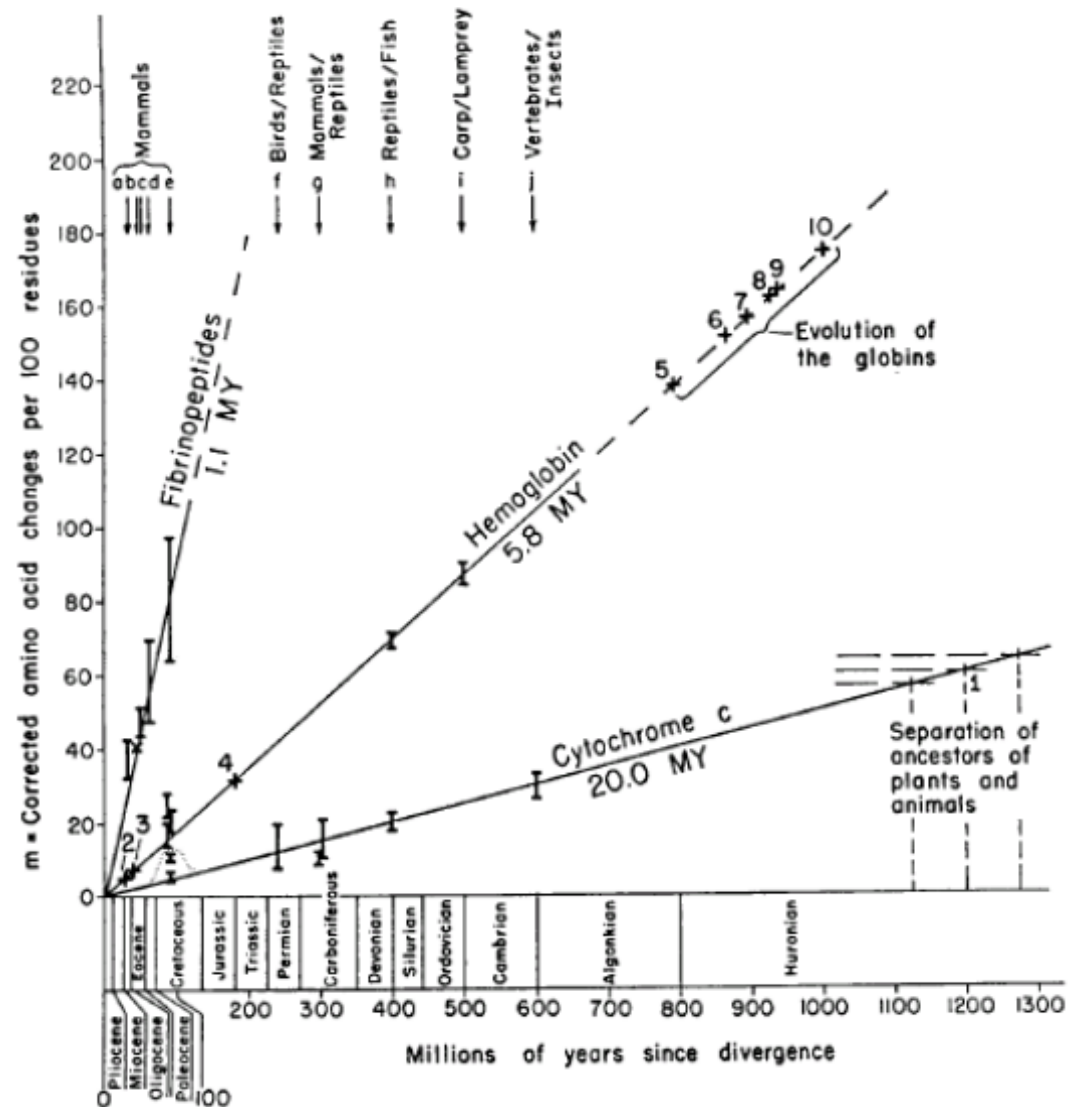
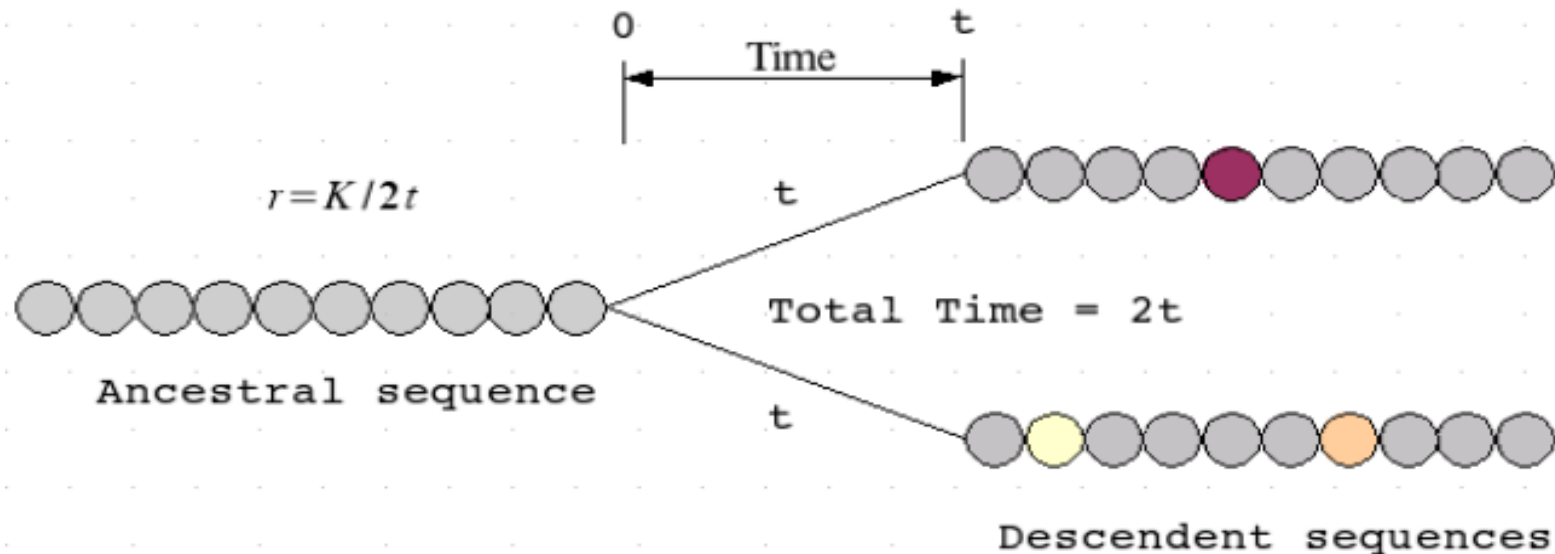


Fig. 8.3. Rates of amino acid substitution in the fibrinopeptides, hemoglobin, and cytochrome c. Comparisons for which no adequate time coordinate is available are indicated by numbered crosses. Point 1 represents a date of 1200 ± 75 MY (million years) for the separation of plants and animals, based on a linear extrapolation of the cytochrome c curve. Points 2–10 refer to events in the evolution of the globin family. The δ/β separation is at point 3, γ/β is at 4, and α/β is at 500 MY (carp/lamprey). From Dickerson (1971).

Evolucion molecular

Reloj molecular

- Las secuencias pueden ser usadas para inferir la edad de los eventos evolutivos.
- La tasa a la que aparecen los cambios entre las secuencias nos permite hacer nos una idea de los mecanismos de evolución molecular.



Evolución molecular

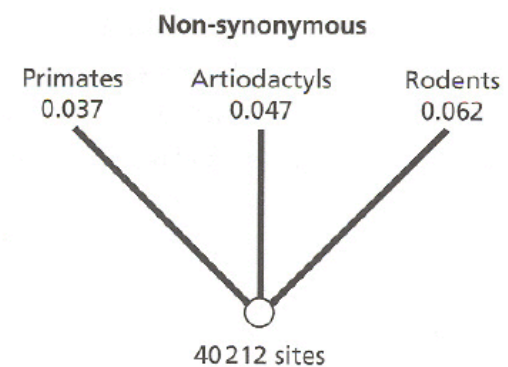
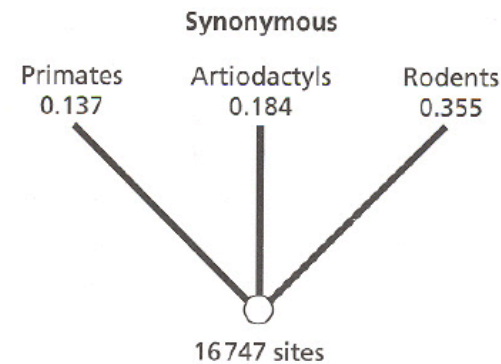
No existe un reloj universal,
variaciones existen para:

- diferentes moléculas,
dependiendo de restricciones
funcionales (expresión)
- diferentes regiones en una
misma molécula (superficie de
la hemoglobina tiene tasa 10
veces mas alta que su interior)

Evolución molecular

No existe un reloj universal,
variaciones existen para:

- diferentes moléculas,
dependiendo de restricciones
funcionales (expresión)
- diferentes regiones en una
misma molécula (superficie de
la hemoglobina tiene tasa 10
veces mas alta que su interior)
- diferentes grupos taxonómicos



Evolución molecular

No existe un reloj universal,
variaciones existen para:

- diferentes moléculas, dependiendo de restricciones funcionales (expresión)
- diferentes regiones en una misma molécula (superficie de la hemoglobina tiene tasa 10 veces mas alta que su interior)
- diferentes grupos taxonómicos
- posición de la base (sinónima, no-sinónima)

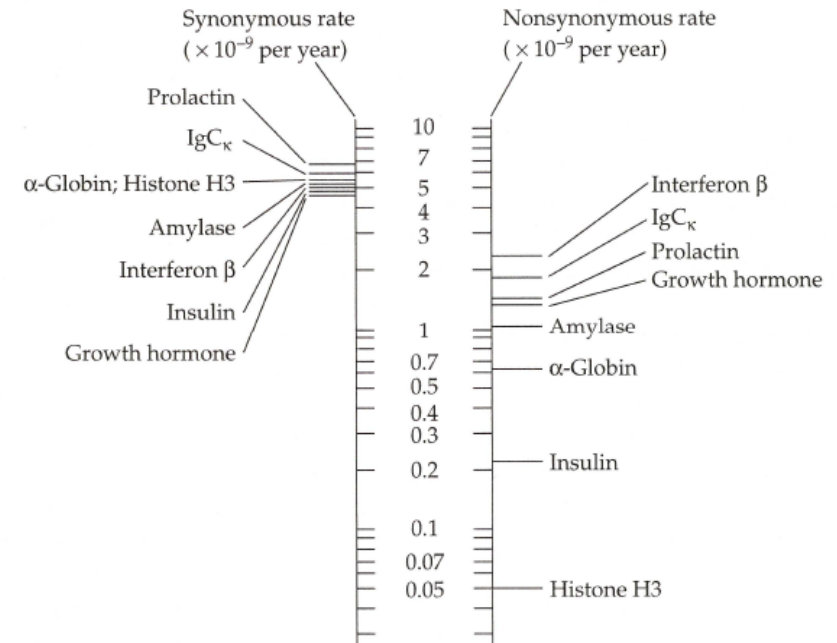


Figure 8.14 Comparison of rates of synonymous and nonsynonymous nucleotide substitutions. Synonymous rates are generally much faster and much more uniform than nonsynonymous rates. (From Kimura 1986.)

Evolucion molecular

Mutaciones sinónimas y no-sinónimas

		2nd base							
		U		C		A		G	
1st base	U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine
		UUC	(Phe/F) Phenylalanine	UCC	(Ser/S) Serine	UAC	(Tyr/Y) Tyrosine	UGC	(Cys/C) Cysteine
		UUA	(Leu/L) Leucine	UCA	(Ser/S) Serine	UAA	Stop (<i>Ochre</i>)	UGA	Stop (<i>Opal</i>)
		UUG	(Leu/L) Leucine	UCG	(Ser/S) Serine	UAG	Stop (<i>Amber</i>)	UGG	(Trp/W) Tryptophan
	C	CUU	(Leu/L) Leucine	CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine
		CUC	(Leu/L) Leucine	CCC	(Pro/P) Proline	CAC	(His/H) Histidine	CGC	(Arg/R) Arginine
		CUA	(Leu/L) Leucine	CCA	(Pro/P) Proline	CAA	(Gln/Q) Glutamine	CGA	(Arg/R) Arginine
		CUG	(Leu/L) Leucine	CCG	(Pro/P) Proline	CAG	(Gln/Q) Glutamine	CGG	(Arg/R) Arginine
	A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine
		AUC	(Ile/I) Isoleucine	ACC	(Thr/T) Threonine	AAC	(Asn/N) Asparagine	AGC	(Ser/S) Serine
		AUA	(Ile/I) Isoleucine	ACA	(Thr/T) Threonine	AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine
		AUG ^[A]	(Met/M) Methionine	ACG	(Thr/T) Threonine	AAG	(Lys/K) Lysine	AGG	(Arg/R) Arginine
	G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine
		GUC	(Val/V) Valine	GCC	(Ala/A) Alanine	GAC	(Asp/D) Aspartic acid	GGC	(Gly/G) Glycine
		GUA	(Val/V) Valine	GCA	(Ala/A) Alanine	GAA	(Glu/E) Glutamic acid	GGA	(Gly/G) Glycine
		GUG	(Val/V) Valine	GCG	(Ala/A) Alanine	GAG	(Glu/E) Glutamic acid	GGG	(Gly/G) Glycine

Evolucion molecular

Mutaciones sinónimas y no-sinónimas

		2nd base							
		U		C		A		G	
1st base	U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine
		UUC	(Phe/F) Phenylalanine	UCC	(Ser/S) Serine	UAC	(Tyr/Y) Tyrosine	UGC	(Cys/C) Cysteine
		UUA	(Leu/L) Leucine	UCA	(Ser/S) Serine	UAA	Stop (Ochre)	UGA	Stop (Opal)
		UUG	(Leu/L) Leucine	UCG	(Ser/S) Serine	UAG	Stop (Amber)	UGG	(Trp/W) Tryptophan
	C	CUU	(Leu/L) Leucine	CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine
		CUC	(Leu/L) Leucine	CCC	(Pro/P) Proline	CAC	(His/H) Histidine	CGC	(Arg/R) Arginine
		CUA	(Leu/L) Leucine	CCA	(Pro/P) Proline	CAA	(Gln/Q) Glutamine	CGA	(Arg/R) Arginine
		CUG	(Leu/L) Leucine	CCG	(Pro/P) Proline	CAG	(Gln/Q) Glutamine	CGG	(Arg/R) Arginine
	A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine
		AUC	(Ile/I) Isoleucine	ACC	(Thr/T) Threonine	AAC	(Asn/N) Asparagine	AGC	(Ser/S) Serine
		AUA	(Ile/I) Isoleucine	ACA	(Thr/T) Threonine	AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine
		AUG ^[A]	(Met/M) Methionine	ACG	(Thr/T) Threonine	AAG	(Lys/K) Lysine	AGG	(Arg/R) Arginine
	G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine
		GUC	(Val/V) Valine	GCC	(Ala/A) Alanine	GAC	(Asp/D) Aspartic acid	GGC	(Gly/G) Glycine
		GUA	(Val/V) Valine	GCA	(Ala/A) Alanine	GAA	(Glu/E) Glutamic acid	GGA	(Gly/G) Glycine
		GUG	(Val/V) Valine	GCG	(Ala/A) Alanine	GAG	(Glu/E) Glutamic acid	GGG	(Gly/G) Glycine

Evolucion molecular

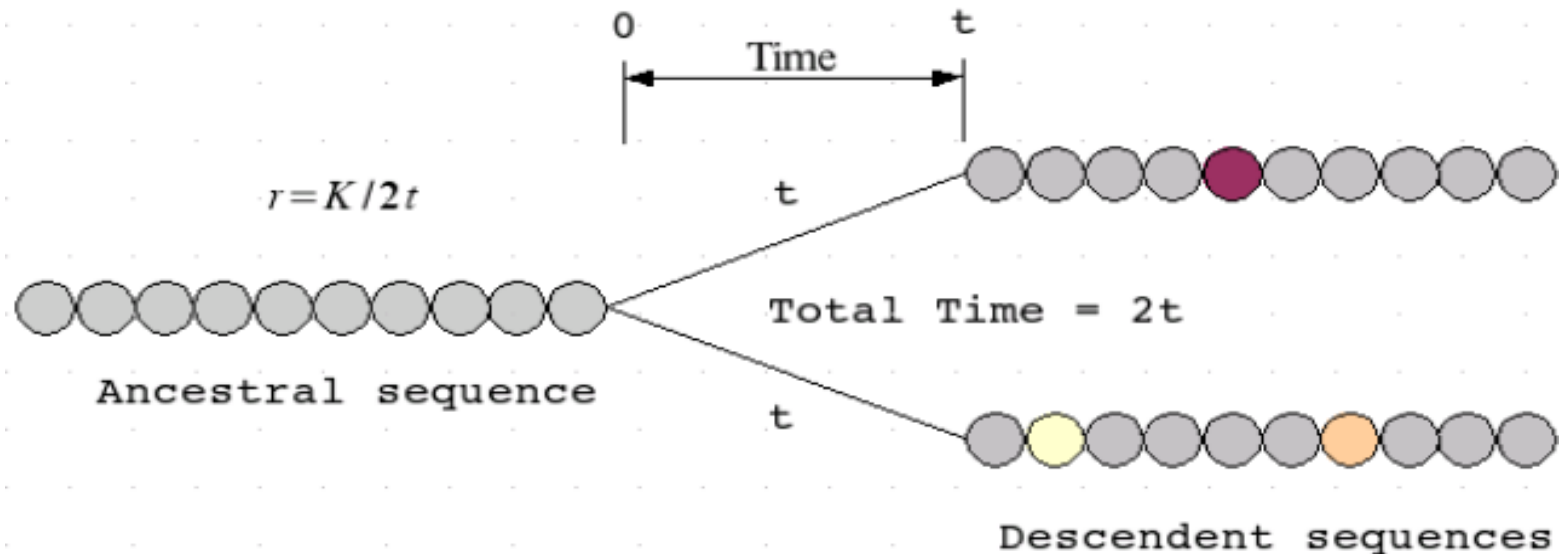
Mutaciones sinónimas y no-sinónimas

		2nd base							
		U		C		A		G	
1st base	U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine
		UUC	(Phe/F) Phenylalanine	UCC	(Ser/S) Serine	UAC	(Tyr/Y) Tyrosine	UGC	(Cys/C) Cysteine
		UUA	(Leu/L) Leucine	UCA	(Ser/S) Serine	UAA	Stop (Ochre)	UGA	Stop (Opal)
		UUG	(Leu/L) Leucine	UCG	(Ser/S) Serine	UAG	Stop (Amber)	UGG	(Trp/W) Tryptophan
	C	CUU	(Leu/L) Leucine	CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine
		CUC	(Leu/L) Leucine	CCC	(Pro/P) Proline	CAC	(His/H) Histidine	CGC	(Arg/R) Arginine
		CUA	(Leu/L) Leucine	CCA	(Pro/P) Proline	CAA	(Gln/Q) Glutamine	CGA	(Arg/R) Arginine
		CUG	(Leu/L) Leucine	CCG	(Pro/P) Proline	CAG	(Gln/Q) Glutamine	CGG	(Arg/R) Arginine
	A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine
		AUC	(Ile/I) Isoleucine	ACC	(Thr/T) Threonine	AAC	(Asn/N) Asparagine	AGC	(Ser/S) Serine
		AUA	(Ile/I) Isoleucine	ACA	(Thr/T) Threonine	AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine
		AUG ^[A]	(Met/M) Methionine	ACG	(Thr/T) Threonine	AAG	(Lys/K) Lysine	AGG	(Arg/R) Arginine
	G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine
		GUC	(Val/V) Valine	GCC	(Ala/A) Alanine	GAC	(Asp/D) Aspartic acid	GGC	(Gly/G) Glycine
		GUA	(Val/V) Valine	GCA	(Ala/A) Alanine	GAA	(Glu/E) Glutamic acid	GGA	(Gly/G) Glycine
		GUG	(Val/V) Valine	GCG	(Ala/A) Alanine	GAG	(Glu/E) Glutamic acid	GGG	(Gly/G) Glycine

Evolucion molecular

Modelos evolutivos

- Cambios en tasas de mutación varían dependiendo de su posición en el ADN. Al nivel de proteínas los patrones de substitución varían dependiendo de la posición en el codon (1era 2nda 3era posición)
- Una forma sencilla de medir la divergencia entre dos secuencias es contar el numero de sitios que difieren:

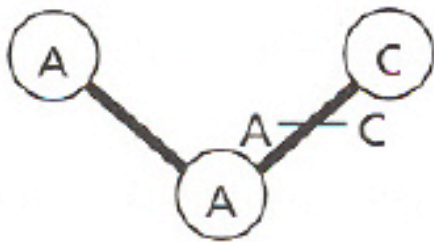


Evolución molecular

Numero de mutaciones observadas no es necesariamente el numero de mutaciones real

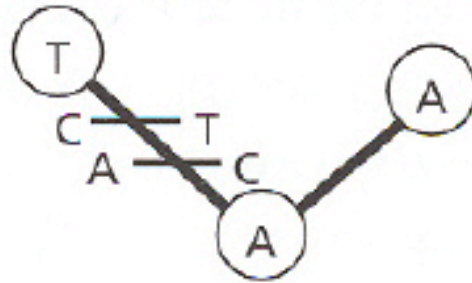
(a) Single substitution

1 change, 1 difference



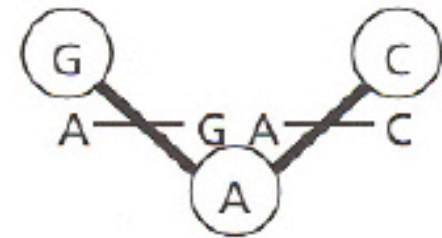
(b) Multiple substitution

2 changes, 1 difference



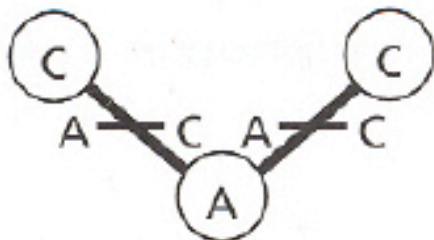
(c) Coincidental substitution

2 changes, 1 difference



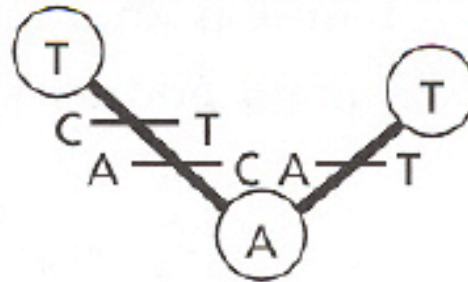
(d) Parallel substitution

2 changes, no difference



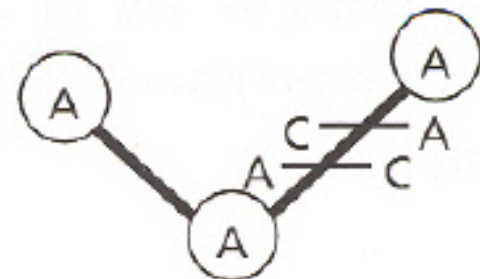
(e) Convergent substitution

3 changes, no difference



(f) Back substitution

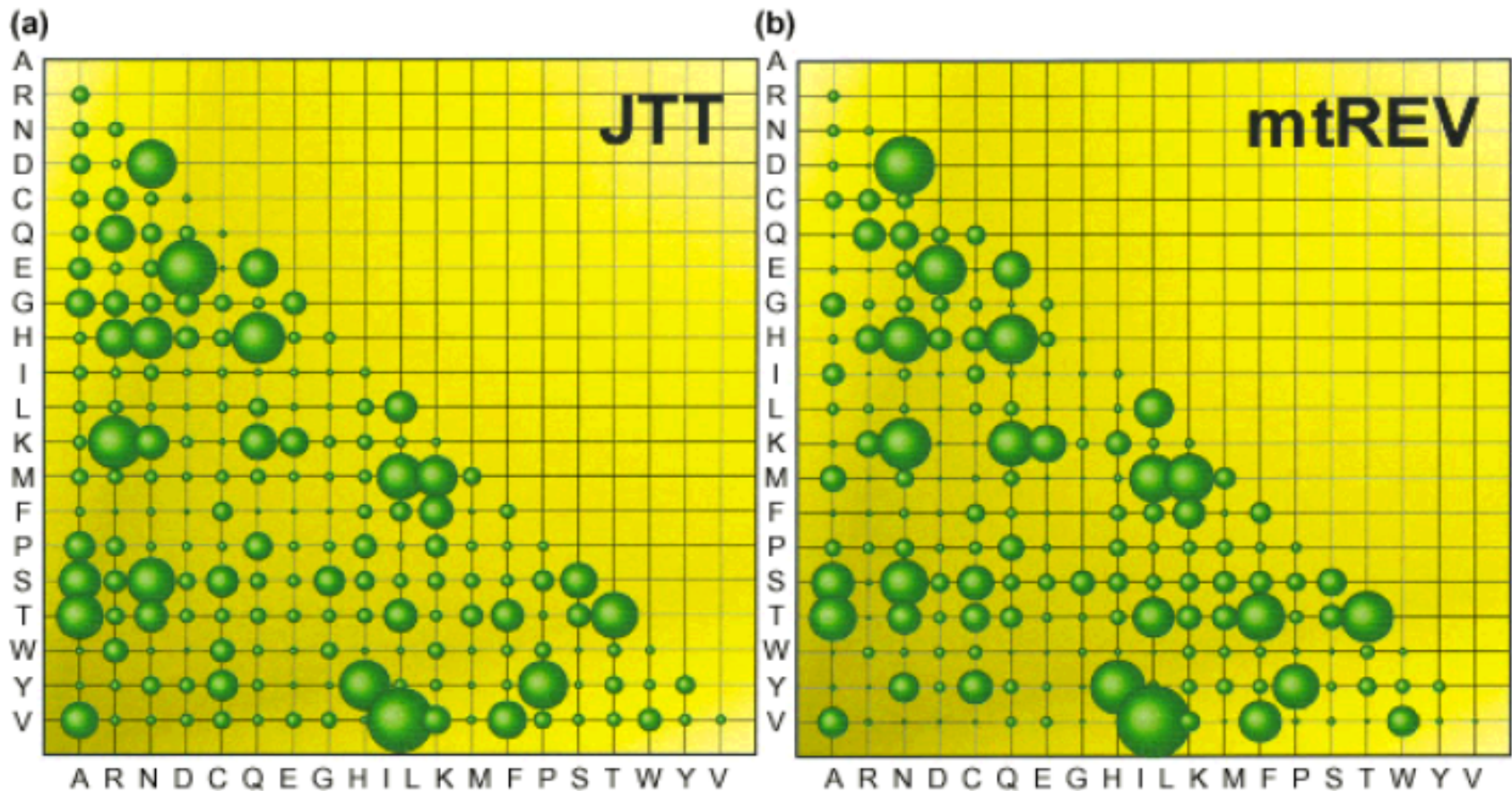
2 changes, no difference



Evolucion molecular

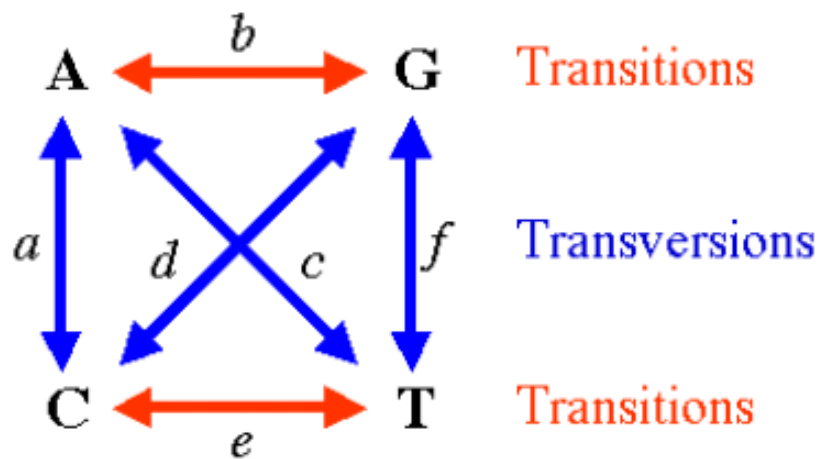
Modelos de evolución de amino ácidos

- Modelos empíricos, basados en observaciones
- La estima del mejor modelo se hace por máxima verosimilitud y depende de las secuencias alineadas.



Evolucion molecular

Modelos ADN



Designation	Rate params	Base frequencies	Number of free params
JC	$a=b=c=d=e=f$	$\pi_A = \pi_C = \pi_G = \pi_T$	1
K80, K2P	$a=c=d=f, b=e$	$\pi_A = \pi_C = \pi_G = \pi_T$	2
TrNef	$a=c=d=f, b, e$	$\pi_A = \pi_C = \pi_G = \pi_T$	3
K81, K3ST	$a=f, b=e, c=d$	$\pi_A = \pi_C = \pi_G = \pi_T$	3
TVMef	$a, c, d, f, b=e$	$\pi_A = \pi_C = \pi_G = \pi_T$	5
TiMef	$a=f, c=d, b, e$	$\pi_A = \pi_C = \pi_G = \pi_T$	4
SYM	a, b, c, d, e, f	$\pi_A = \pi_C = \pi_G = \pi_T$	6
F81	$a=b=c=d=e$	$\pi_A, \pi_C, \pi_G, \pi_T$	4
HKY	$a=c=d=f, b=e$	$\pi_A, \pi_C, \pi_G, \pi_T$	5
TrN	$a=c=d=f, b, e$	$\pi_A, \pi_C, \pi_G, \pi_T$	6
K81uf	$a=f, b=e, c=d$	$\pi_A, \pi_C, \pi_G, \pi_T$	6
TVM	$a, c, d, f, b=e$	$\pi_A, \pi_C, \pi_G, \pi_T$	8
TiM	$a=f, c=d, b, e$	$\pi_A, \pi_C, \pi_G, \pi_T$	7
GTR, REV	a, b, c, d, e, f	$\pi_A, \pi_C, \pi_G, \pi_T$	9

Evolucion molecular

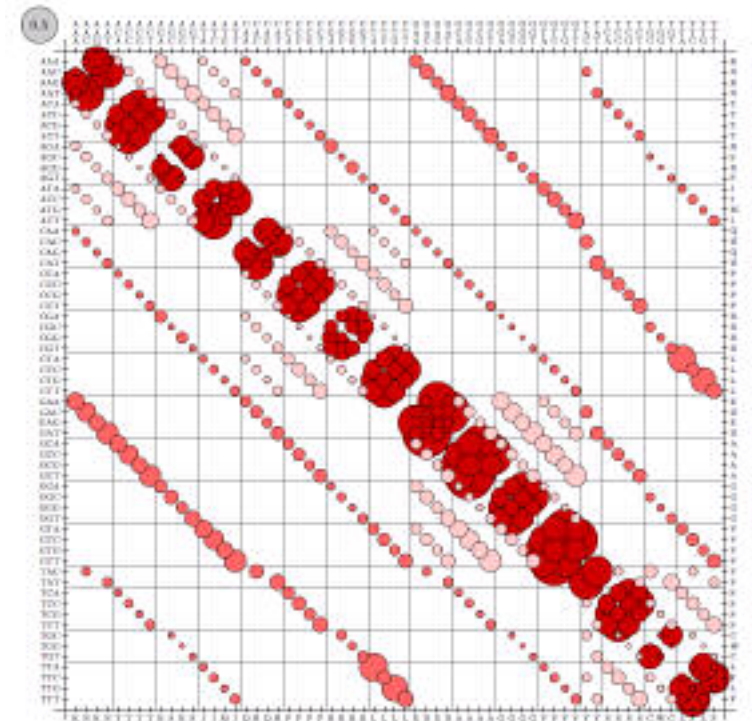
Modelos de codones (ejemplo con modelo $M0$)

$$q_{ij} = \begin{cases} 0, & \text{si } i \text{ y } j \text{ difieren en 2 o 3 posiciones de codon} \\ \mu\pi_j, & \text{si } i \text{ y } j \text{ difieren por una transversión sinónima} \\ \mu\kappa\pi_j, & \text{si } i \text{ y } j \text{ difieren por una transición sinónima} \\ \mu\omega\pi_j, & \text{si } i \text{ y } j \text{ difieren por una transversión no-sinónima} \\ \mu\kappa\omega\pi_j, & \text{si } i \text{ y } j \text{ difieren por una transición no-sinónima} \end{cases}$$

κ : tasa de transición/transversión

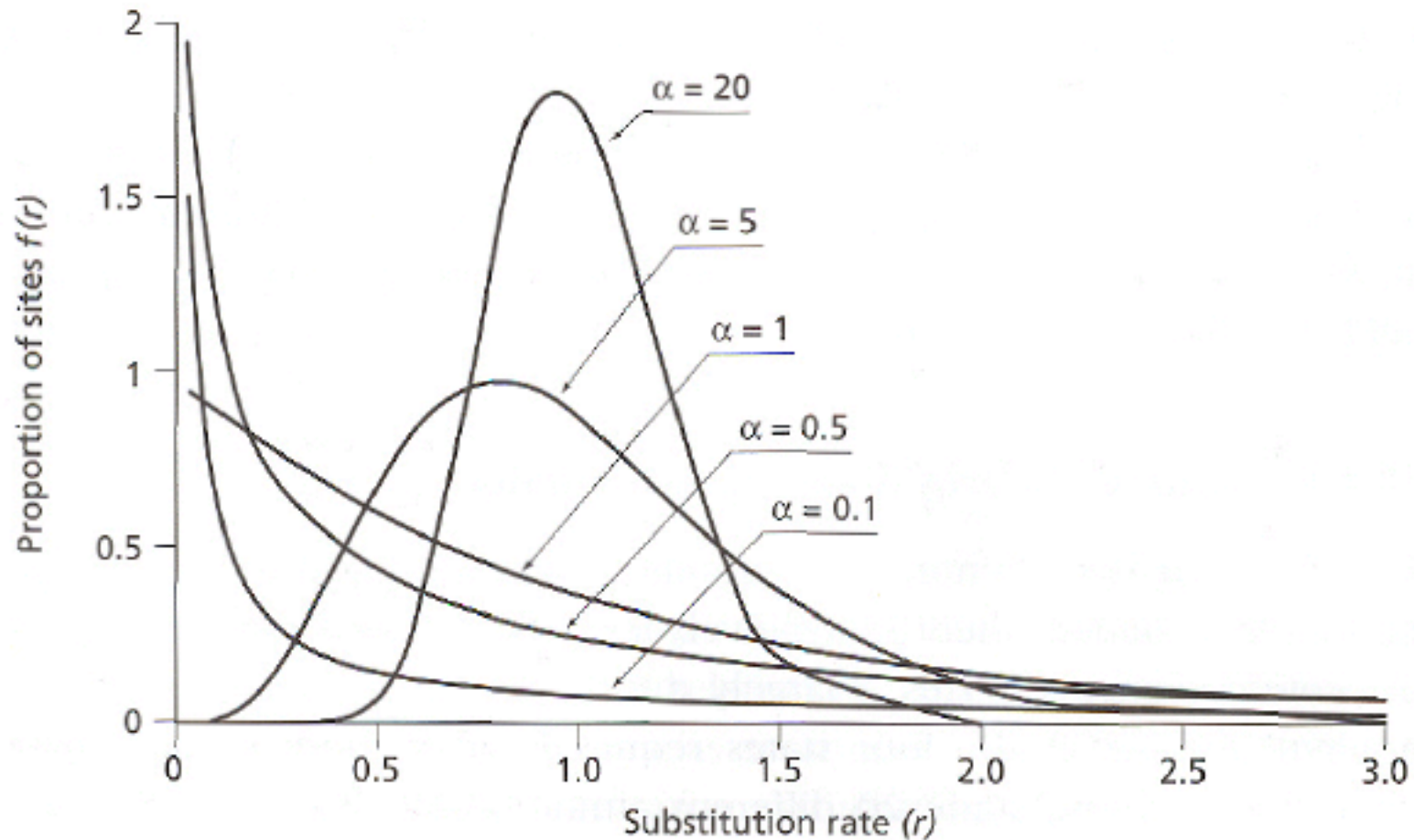
ω : tasa de mutación sinónima/no-sinónima

π_j : frecuencia del codon j



Evolución molecular

Correcciones de heterogeneidad de tasa de mutación a lo largo de una secuencia



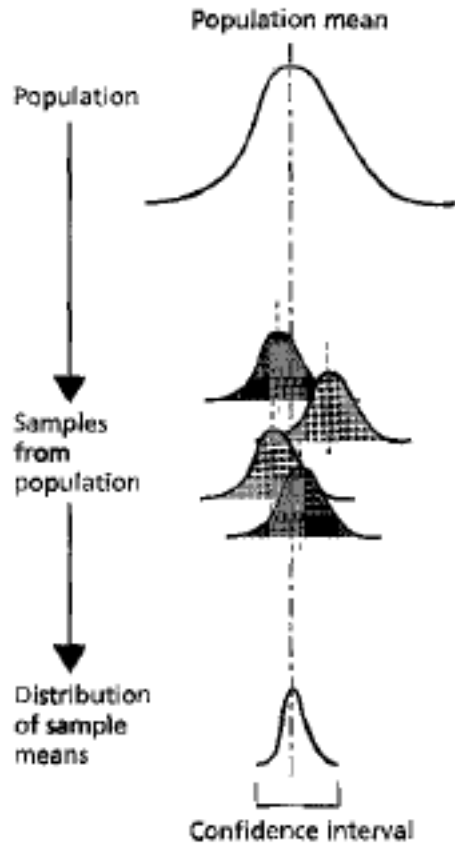
Construcción de árboles

- **Métodos de distancia** (Neighbour-joining, UPGMA)
 - calcular las distancia para cada par de secuencias
 - hacerlo encajar en un árbol
- **Métodos de parsimonia**
 - reconstruir el árbol con el mínimo de cambios posibles
- **Métodos estadísticos**
 - Máxima verosimilitud: dadas las secuencias encontrar el árbol mas probable, $P(D|H)$
 - *Bayesianos*: encontrar el árbol que explique mejor las secuencias, $P(H|D)$

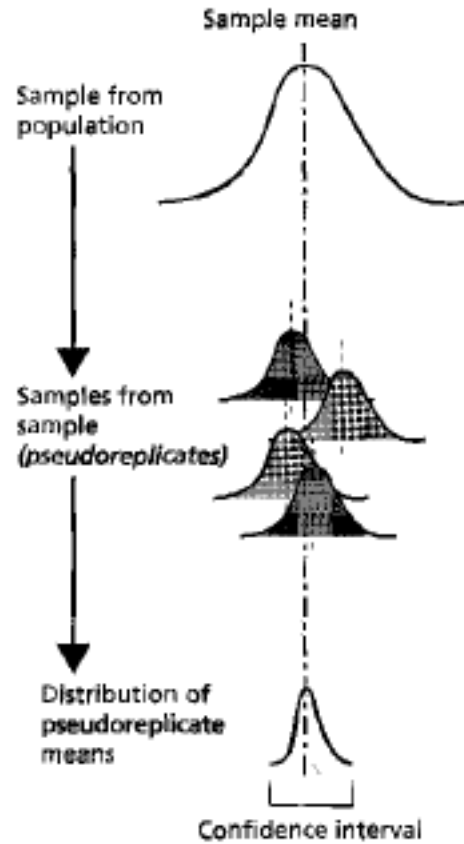
Arbol y confianza

Bootstrapping

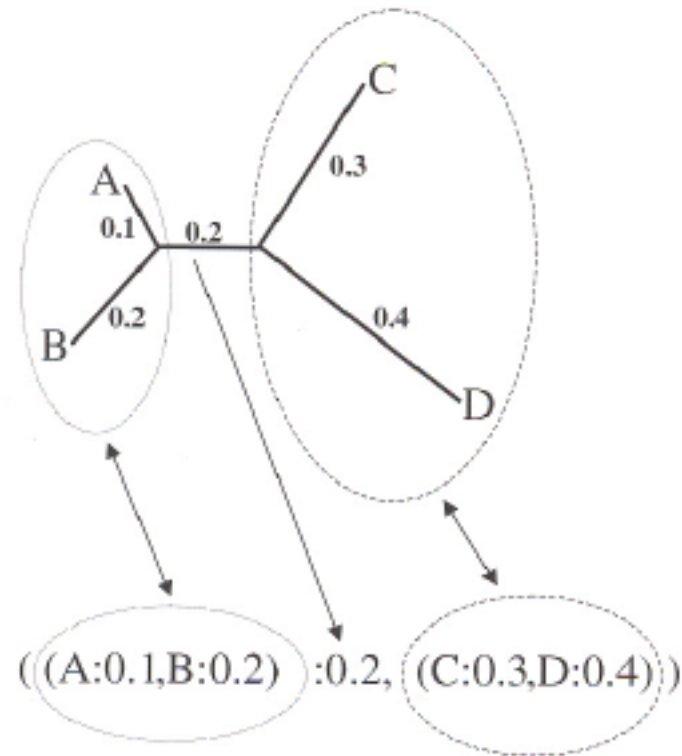
(a) Resampling from population



(b) Resampling from a sample (bootstrap)



$$((A,B),(C,D)) \equiv \begin{array}{c} A & & C \\ & \diagdown & / \\ & & \\ & / & \diagdown \\ B & & D \end{array}$$



Adaptación

Selección positiva, Selección negativa, Relajación

↓
**Relaxation / weak
selection**



Adaptación

Selección positiva, Selección negativa, Relajación



Positive Selection

**Relaxation / weak
selection**



Adaptación

Selección positiva, Selección negativa, Relajación



Positive Selection

**Relaxation / weak
selection**



Adaptación

Selección positiva, Selección negativa, Relajación



Positive Selection

**Relaxation / weak
selection**



Adaptación

Selección positiva, Selección negativa, Relajación



Positive Selection

**Relaxation / weak
selection**



Adaptación

Selección positiva, Selección negativa, Relajación



Positive Selection

**Relaxation / weak
selection**



Adaptación

Selección positiva, Selección negativa, Relajación



Positive Selection



Relaxation / weak selection

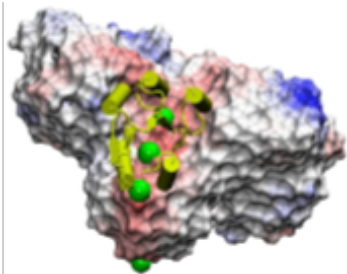


Adaptación

Selección positiva, Selección negativa, Relajación



Positive Selection



Relaxation / weak selection

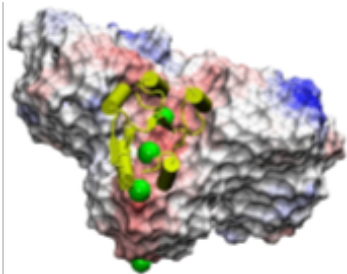


Adaptación

Selección positiva, Selección negativa, Relajación



Positive Selection



Relaxation / weak selection



Purifying selection

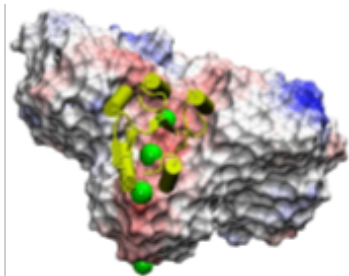


Adaptación

Selección positiva, Selección negativa, Relajación



Positive Selection



Relaxation / weak selection



Purifying selection



Adaptacion

Cómo se mide?

Concepto : Comparar tasa de cambio de algo **funcionalmente relevante (tasa1)** con la tasa de algo neutro, **que escape a selección natural (tasa2)**.

$$\text{presion selectiva} = \frac{\text{tasa1}}{\text{tasa2}}$$

Adaptación

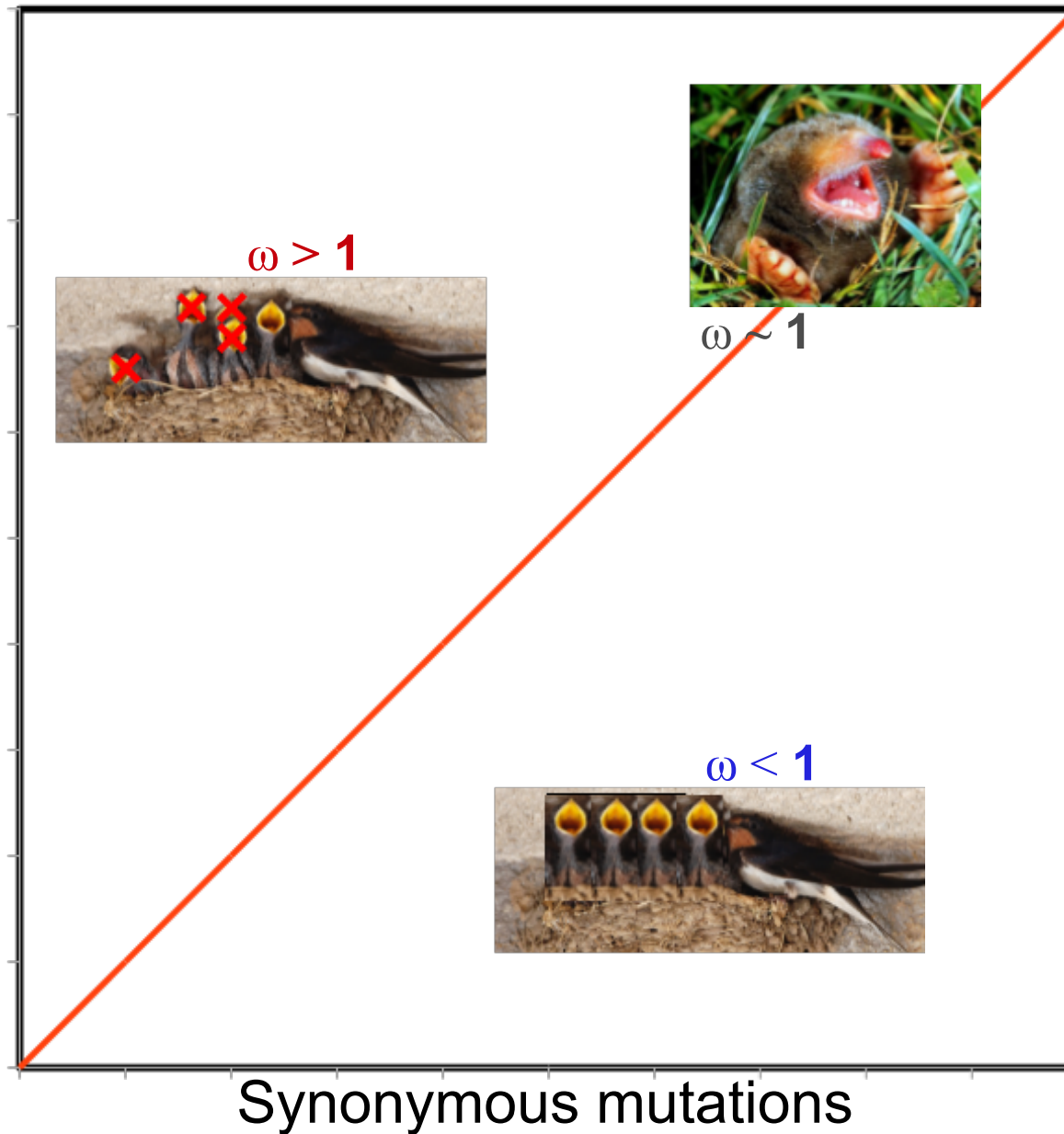
Como se mide?

- en **codones** :
 - estima filogenetica de $\omega = dN/dS$
- en **proteínas** :
 - tasa de evolución relativa (u), no se distingue entre selección positiva y relajación.
- en **promotores** :
 - comparación de la tasa de intrones o partes cercanas de ADN no codificante evolucionando de forma neutral.

Adaptación

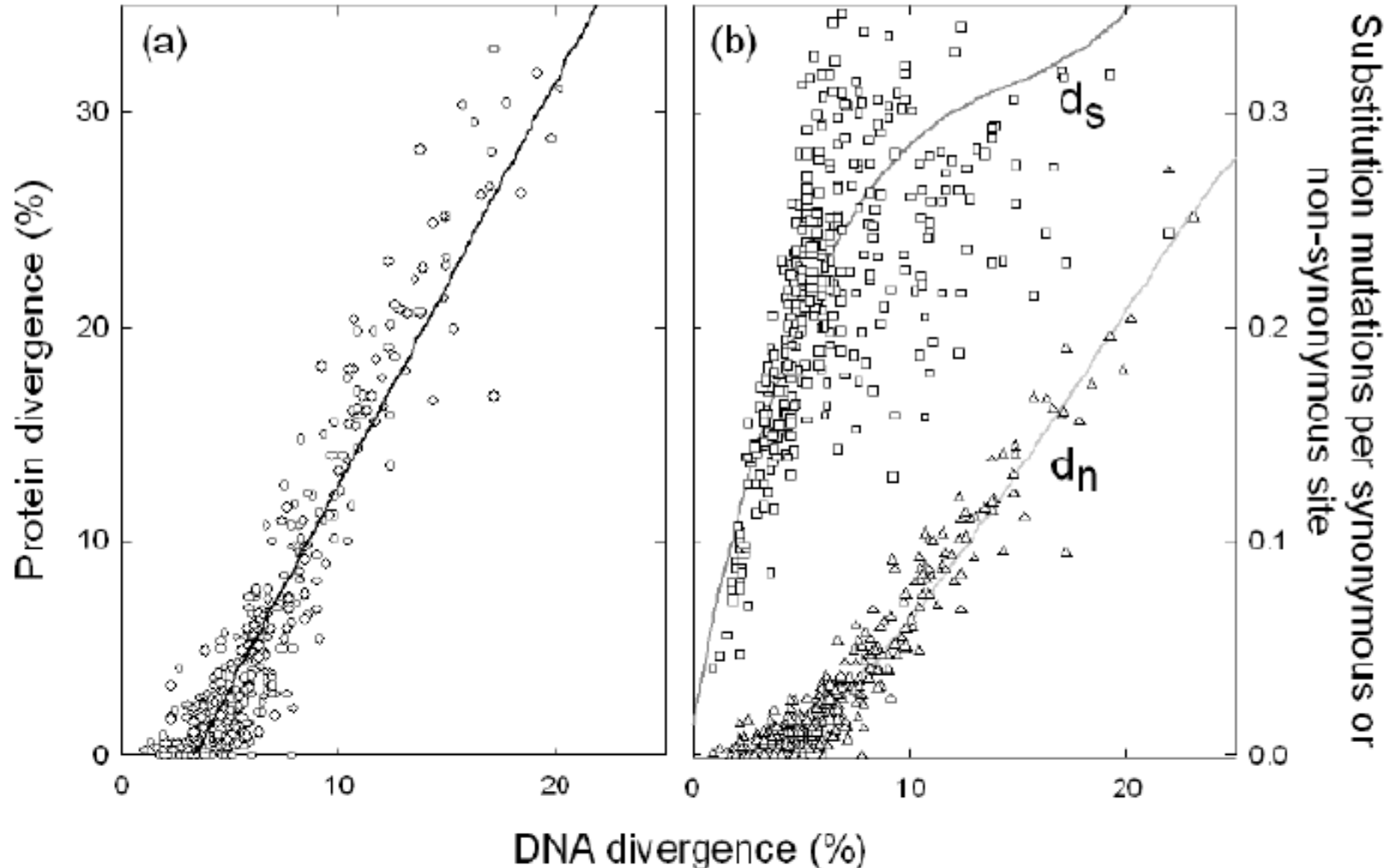
$$\omega = dN/dS$$

Non-synonymous mutations



Adaptacion

Saturación de la tasa de mutación



Adaptación

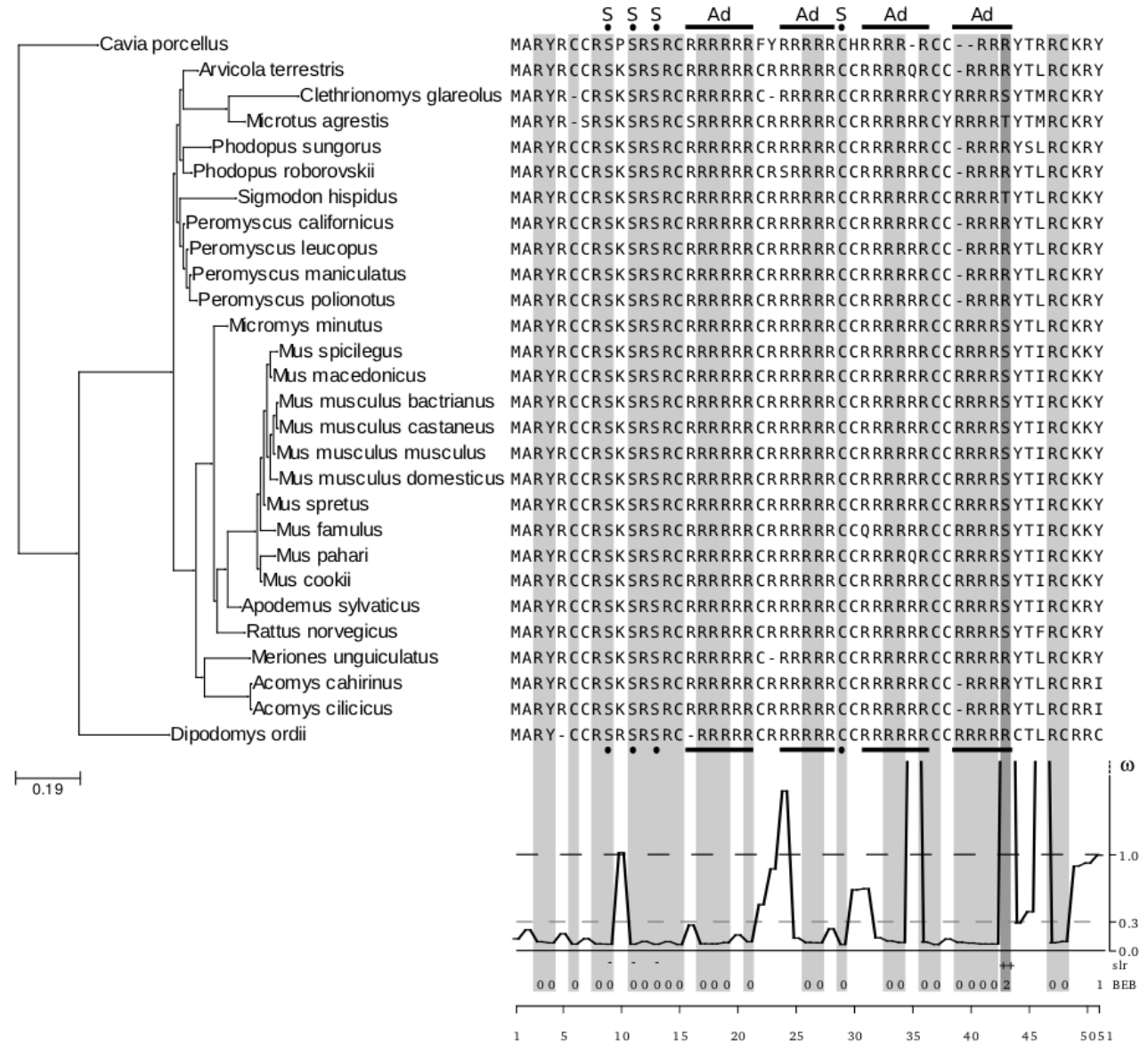
Modelos usados para medir adaptación en codones:

- Sitios
- ramas
- ramas y sitios

Adaptación

Modelos usados para medir adaptación en codones:

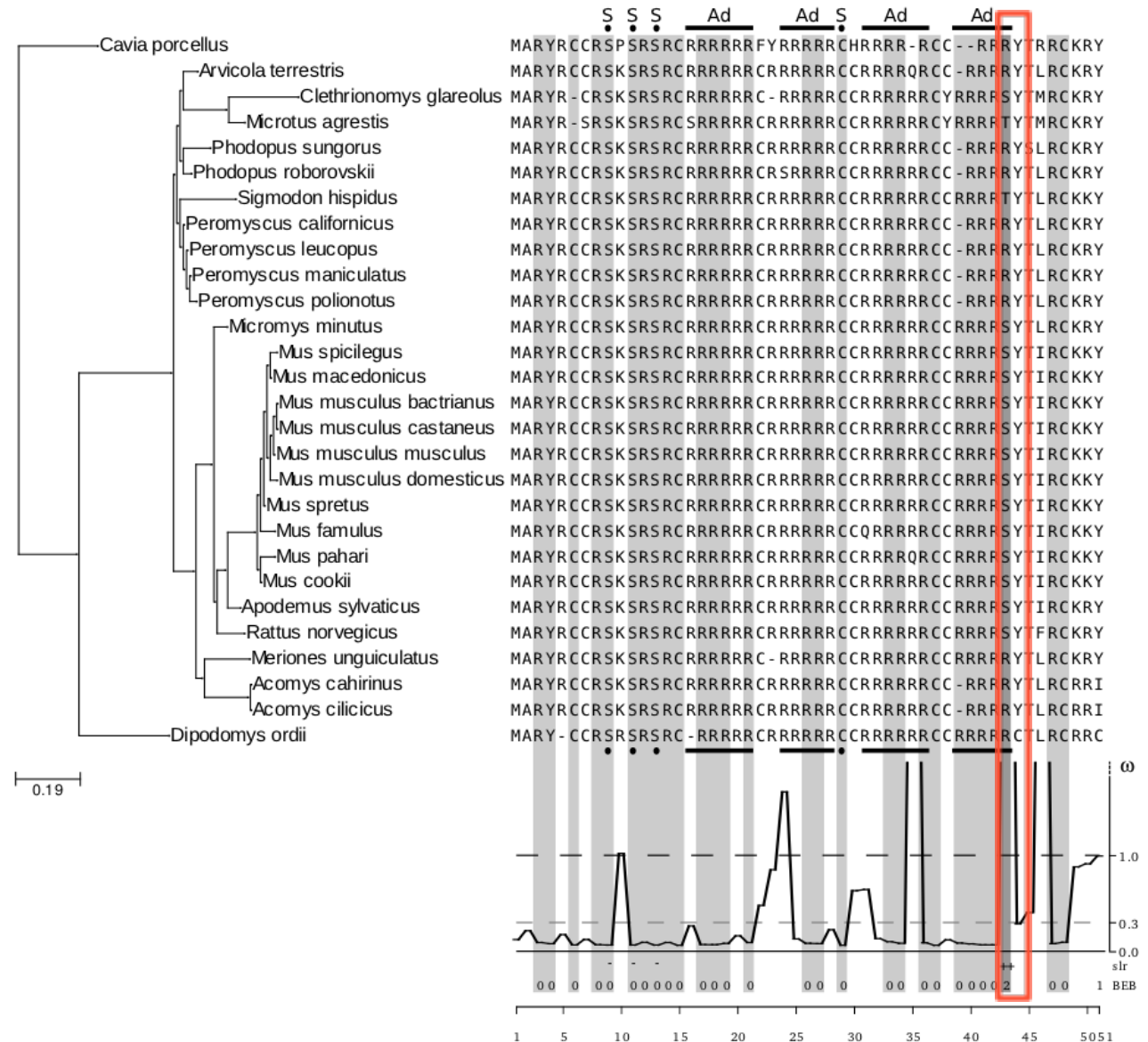
- Sitios



Adaptación

Modelos usados para medir adaptación en codones:

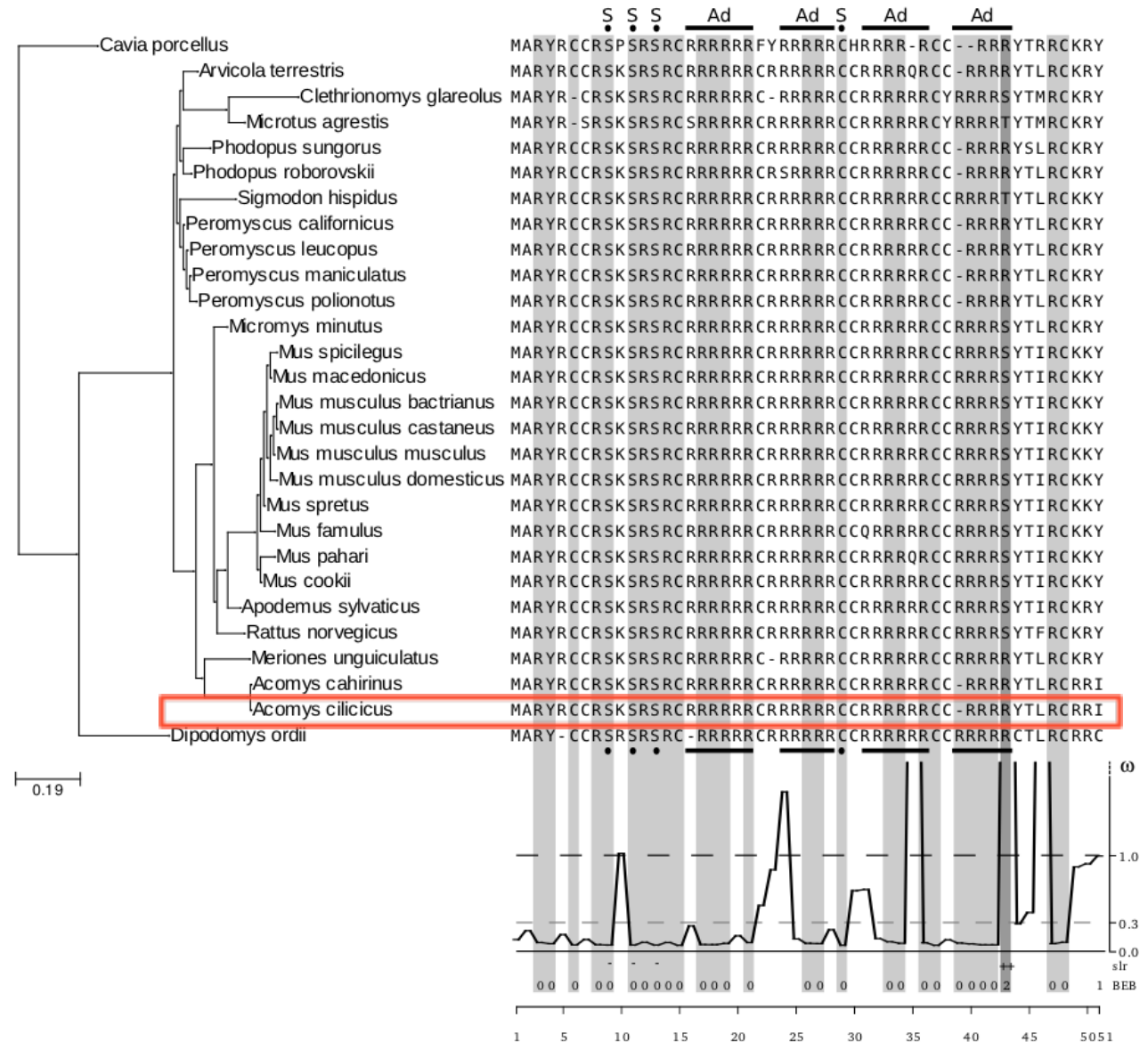
- Sitios



Adaptación

Modelos usados para medir adaptación en codones:

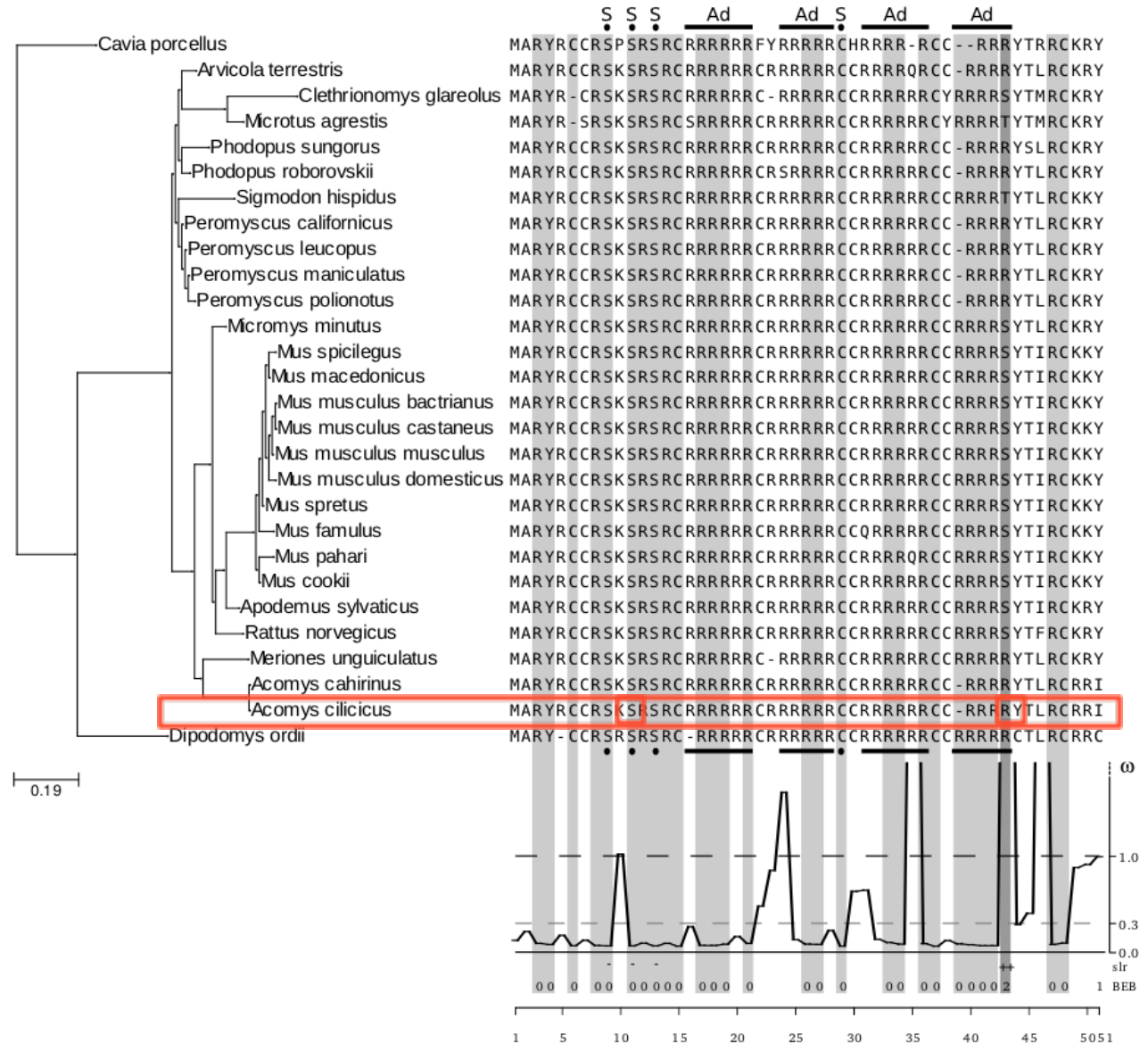
- Sitios
- ramas



Adaptación

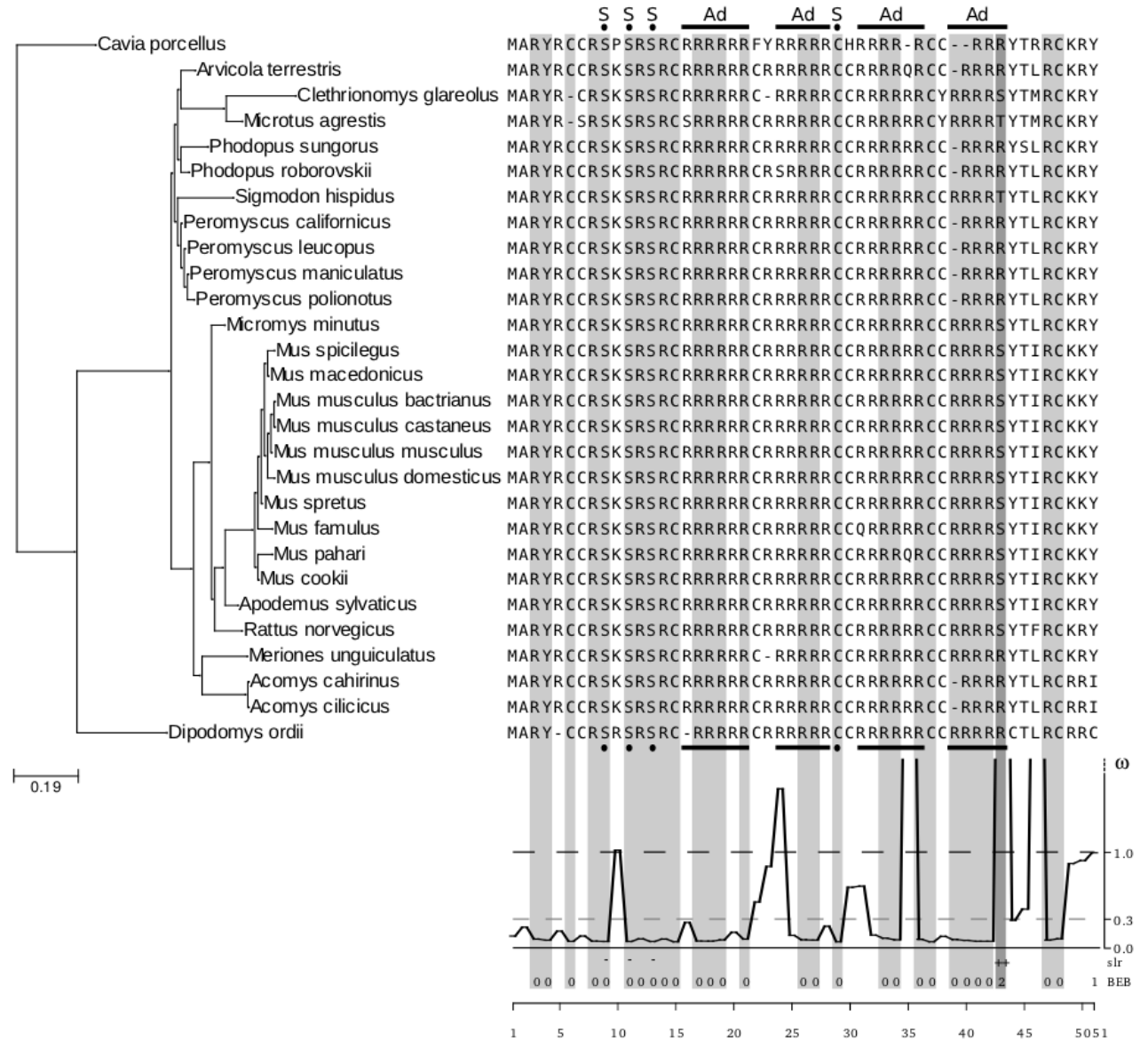
Modelos usados para medir adaptación en codones:

- Sitios
- ramas
- ramas y sitios



Adaptación

a que sirve?



Enlaces/ayuda

- La mayor parte de esta introducción fue sacada del curso de ["Molecular Evolucion Phylogenetics Phylogenomics and Adaptation"](#) por **Hernan Dopazo**. Desde esta misma web se pueden encontrar referencias a libros que cubren la mayor parte de los temas presentados.
- Existen muchos cursos/seminarios en la web:
 - Web de seminarios sobre filogenia: <http://phyloseminar.org/>
 - Nature Scitable: <http://www.nature.com/scitable/topic/evolutionary-genetics-13>
 - NCBI books: <http://www.ncbi.nlm.nih.gov/books/NBK21122/>
 - etc...

Preguntas?

