

## Práctico 8 - Familias multigénicas - “Globinas”

**Objetivo:** Mediante el estudio de los genes de Globinas en primates, familiarizarse con la identificación de familias multigénicas, su problemática y limitaciones.

Una familia génica o familia multigénica es un grupo de loci cromosómicos cuya secuencia de nucleótidos es similar y derivan de una secuencia común ancestral. Puede incluir copias de genes ligeramente diferentes y/o pseudogenes más variables, en uno o varios cromosomas.

Los genes codificantes para las globinas representan un ejemplo clásico de familia multigénica. En general, estas proteínas portan un grupo “heme” y se caracterizan por unirse y transportar oxígeno. Presentan dominios homólogos en varios taxa: a) hemoglobinas tetraméricas de vertebrados (componente proteico principal de los eritrocitos), b) flavohemoglobinas en microorganismos, c) hemoglobinas homodiméricas bacterianas, d) leghemoglobinas en plantas (asociadas al metabolismo del nitrógeno en plantas con rhizomas), e) hemoglobinas de invertebrados, f) mioglobinas monoméricas usualmente encontradas en el músculo animal entre otras.

En primates existen varias clases de globinas dentro de la familia, generalmente designadas con letras griegas (por ejemplo,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\xi$  y Mioglobina MB). Recientemente se ha identificado una nueva clase, la citoglobina (CYGB). En humanos, el gen que la codifica parece expresarse en todos los tejidos y se localiza en el brazo largo del par cromosómico 17. Se sugiere que su tamaño está conservado en varios mamíferos y es de 190 aminoácidos.

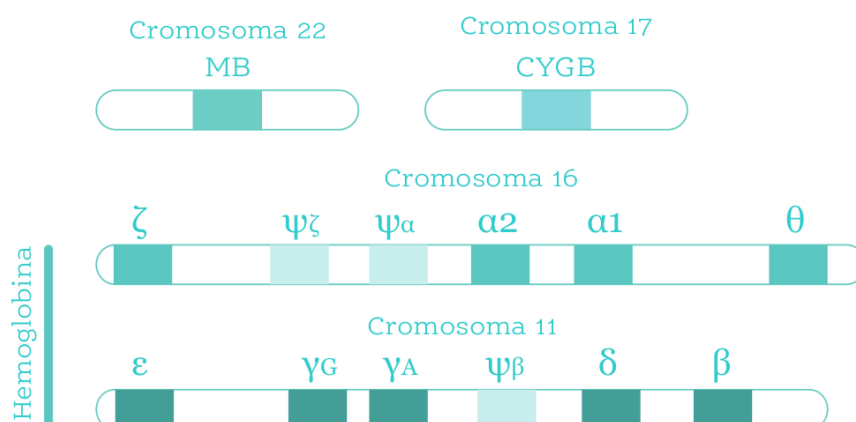


Figura 1 - Localización de los genes codificantes para globinas en humanos.

## Conceptos generales

- ¿Qué procesos podrían estar involucrados en la aparición de las familias multigénicas?
- ¿Qué ventajas conferirían estas familias génicas?

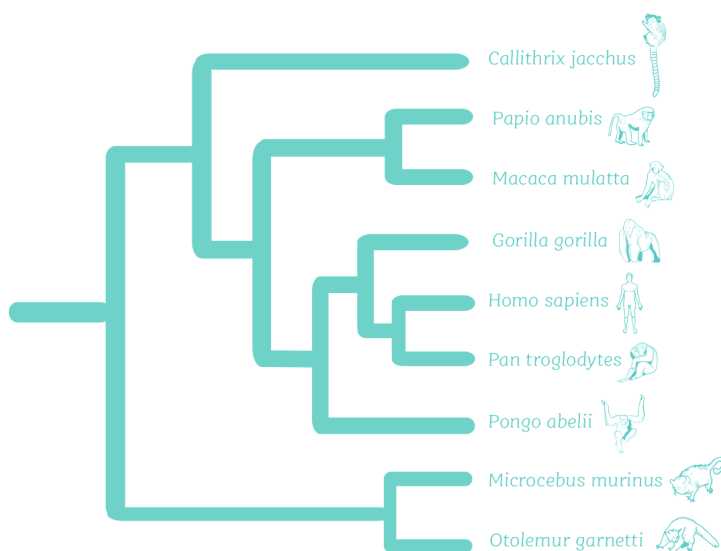
## Análisis del alineamiento de secuencias proteicas de globinas, identificación de regiones conservadas

Cargue el alineamiento en el programa MEGA X: Ir a **Align** → **Edit/Build Alignment** → **Retrieve a sequence from a file** → cargar el archivo “**globinas\_aa.fas**” y alinear: Ir a **Alignment** y elegir cualquiera de las opciones (ClustalW o MUSCLE).

- Identifique los sitios conservados en todas las secuencias y sugiera su posición en la proteína. Justifique.
- ¿Visualiza algún patrón particular en la secuencia de estas proteínas? Interprete.
- Selecciones y copie (ctr + c) la región conservada (al menos 50 aa.) en una de las secuencias y haga un blastp (blast de sec. aminoacídicas) contra la base de datos del genbank (<http://blast.ncbi.nlm.nih.gov/>). ¿A qué corresponde dicha región?
- Opcional : Una forma de identificar a que corresponden dichas regiones es mediante la base del genbank, “Conserved Domain Database (CDD)” (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). La búsqueda de dominios similares en tu secuencia está automáticamente incorporada al blastp.

## Reconstrucción filogenética de las globinas

Cargue los datos en el programa MEGA (**File / Open Data / globinas\_nt.meg**). Esta base de datos tiene las secuencias codificadas (ADNc) reportadas de genes miembros de la familia de las globinas en los siguientes primates: *Homo sapiens*, *Gorilla gorilla*, *Pongo abelii*, *Macaca mulatta*, *Callithrix jacchus*, *Papio anubis*, *Pan troglodytes*, *Microcebus murinus* y *Otolemur garnettii* (ver Figura 2).



## Figura 2 - Cladograma de las especies incluidas los datos

- En base a los datos disponibles (ver Fig 1 y 2), ¿cómo piensa usted que ocurrió la evolución de los diferentes tipos de globinas dentro de los primates? Genere un esquema.

Realice la reconstrucción filogenética en base a los datos que se le brindan (Método: *Neighbor-Joining*, Gaps: *Pairwise deletion*, Bootstrap: *1000 pseudorélicas*).

- ¿Son los resultados coincidentes con el esquema que usted generó?
- Identifique aquellos nodos que se corresponden con eventos de especiación y/o duplicación.
- Analice e interprete algunos casos particulares:
  - ◆ clase delta en humanos
  - ◆ ausencia de  $\gamma$ -globina en *M. murinus* y *O. garnettii*
  - ◆ ausencia de  $\mu$ -globina en *M. murinus*
  - ◆ relación de las citoglobinas con las otras globinas
- Si le interesara, ¿en cuál cromosoma buscaría la  $\mu$ -globina en humanos?
- ¿Qué información adicional sería útil para establecer el origen de un nuevo gen de globinas?

## Estimación del parámetro $\omega$ (dN/dS) entre clases y dentro de las clases

Los genes ya están agrupados en las clases correspondientes en la base de datos. Calcular las distancias media dentro (ortólogos) y entre clases (parálogos).

En MEGA... Ir a *Distances* → *Compute within group mean* y luego *Compute between group mean*, incluyendo sustituciones sinónimas (dS) y no-sinónimas (dN), estimadas con el modelo de Kumar (*LogDet (Tamura-Kumar)*) y Gaps: *Pairwise deletion*.

- ¿Cuál espera sea el resultado dentro y entre clases? ¿Por qué? ¿Qué relación tiene esto con el árbol reconstruido? ¿Cómo piensa que será la estimación de la tasa sinónima y por qué?
- A partir de los valores de distancia media sinónima y no-sinónima calcule la relación entre ellas ( $\approx$  dN/dS) dentro de las clases y entre clases, y tome nota.
- Interprete ambos resultados, relacione los mismos con procesos evolutivos dados en clase.