

## **La Historia Evolutiva Escrita en los Genes de las Globinas. Genes que se duplican, saltan y divergen: éstos son los elementos claves en la evolución de los agrupamientos de genes de las globinas.**

Lewin, R., 1981. Evolutionary history written in globin genes. *Science*, 214, 426-429 (traducido)

Los genes que codifican para las globinas, los componentes proteicos de la hemoglobina, han sido más profundamente estudiados que ninguna otra familia de genes en organismos superiores por dos razones. Primero, porque ofrecen un modelo muy bueno en el cual investigar la operación de los genes que se encienden y se apagan en diferentes estadios del desarrollo. Y segundo, porque el conocimiento detallado acerca de la familia de las globinas prometía algunas visiones útiles dentro de la naturaleza de una serie de enfermedades sanguíneas clínicamente importantes, las talasemias.

Mientras que los resultados de este considerable esfuerzo a escala mundial han colmado muchas de las metas principales del trabajo, aportaron también un resultado inesperado. Ahora es posible bosquejar en detalles reveladores la historia evolutiva de la familia de las globinas en los últimos 500 millones de años. La historia de la familia génica de las globinas contiene muchas lecciones acerca de los mecanismos moleculares de la evolución de los genes, lecciones que deberían ser aplicables a todos los genes de los organismos superiores.

En estos días se ha vuelto una verdad manifiesta decir que el material genético de organismos superiores se encuentra en un estado de cambio dinámico mucho mayor de lo que alguna vez se podría haber predicho, y un descubrimiento muy reciente en los genes de las globinas subraya esto. 'El locus de las globinas está constantemente emitiendo copias de sus genes y bombardeando el resto del genoma con ellas', dice Philip Leder, anteriormente en el Instituto Nacional de la Salud (NHI), Bethesda, y ahora en la Escuela Médica de Harvard. 'Y lo que se aplica a las globinas debe aplicarse a otros genes también', sugiere.

Paradójicamente, la segunda principal característica de la historia de la familia de las globinas es la estabilidad, tanto en la estructura de los genes individuales como en la de la familia como unidad. 'El agrupamiento de genes se mantiene estable sobre largos períodos de tiempo,' dice Alec Jeffreys, de la Universidad de Leicester, Inglaterra, 'pero cuando ocurre el cambio, éste parece ir a saltos.'

La evolución de los genes de las globinas, por lo tanto, aparentemente está gobernada por una combinación de un claro potencial de cambio y la propiedad demostrada de una casi constancia a largo plazo. Tal vez el estilo desigual del modo global de cambio evolutivo en la familia de las globinas es consecuencia de la resolución de estas dos características interactuantes.

La hemoglobina es un tetrámero con dos moléculas tipo  $\alpha$ -globina y dos tipo  $\beta$ -globina. (Cada subunidad de cinco proteínas contiene un grupo hemo que sirve para combinarse con el oxígeno mientras la hemoglobina atraviesa los tejidos.) Los humanos poseemos cinco globinas diferentes del tipo  $\beta$ ,  $\epsilon$ ,  $G\gamma$ ,  $A\gamma$ ,  $\delta$ , y  $\beta$ , las cuales son utilizadas durante la vida embrionaria temprana, la vida fetal, la infancia muy temprana, y la vida adulta. Sólo existen tres genes del tipo  $\alpha$ , el  $\zeta$  embrionario y dos  $\alpha$  de los adultos.

Los agrupamientos (o 'clusters') humanos de las globinas  $\alpha$  y  $\beta$  están localizados en cromosomas diferentes (16 y 11, respectivamente), y el agrupamiento  $\alpha$  es apenas de la mitad del tamaño del de su compañera. En ambos agrupamientos los genes se encuentran dispuestos a lo largo del cromosoma en el orden en el cual se expresan durante el desarrollo.

Las preguntas acerca de la evolución de los genes de las globinas caen en dos áreas principales. La primera se relaciona con la estructura de los propios genes. La segunda está enfocada en la estructura de los agrupamientos de genes.

Las similitudes entre los genes de la  $\alpha$ - y  $\beta$ -globinas indican que surgieron de un único gen ancestral, probablemente por una simple duplicación. Las diferencias detalladas entre los genes ubicaron la fecha de su divergencia en alrededor de 500 millones de años atrás -es decir, en el alba de la historia de los vertebrados.

Como muchos genes eucariotas estudiados hasta el momento, los genes de las globinas están interrumpidos por regiones no codificantes, o intrones. Los genes de la  $\alpha$ - y de la  $\beta$ -globinas poseen dos de tales secuencias intercaladas, en posiciones homólogas. (Esto es consistente con el origen común de los genes.) A diferencia de las regiones codificantes de los genes, las secuencias intercaladas son conocidas por su rápida acumulación de mutaciones. Los intrones de la  $\alpha$ - y de la  $\beta$ -globinas entran dentro de este cuadro, ya que sus secuencias han divergido considerablemente. Lo que es inusual, sin embargo, es que los tamaños de las regiones no codificantes han cambiado relativamente poco.

‘La primera secuencia intercalada es de un largo de unos 116 a 130 pares de bases en todos los genes de  $\alpha$ - y  $\beta$ -globinas de mamíferos estudiados hasta el momento, a pesar de la edad extrema de la duplicación del gen de la  $\beta$ -globina’, dice Jeffreys. ‘El segundo intrón también varía poco dentro de las familias de los genes de la  $\alpha$ - y  $\beta$ -globinas, aunque la secuencia  $\alpha$  es consistentemente más corta que la  $\beta$ .’ Una inferencia obvia es que el tamaño de los intrones es importante para alguna función aún no identificada. Tal restricción parece no aplicarse a la mayoría de los intrones hasta ahora examinados de otros genes.

Una intrigante idea, sólo recientemente desarrollada, conlleva la clara implicación de que durante su larga historia los genes de las globinas de vertebrados perdieron un intrón: en un momento existieron tres, y no dos como ahora observamos.

Mitiko Go, de la Universidad de Kyushu, Japón, se interesó en la noción de que las regiones codificantes de los genes, los exones, se corresponden con los dominios estructurales en la proteína resultante. Esto se relaciona con las especulaciones de Walter Gilbert de la Universidad de Harvard y Colin Blake de la Universidad de Oxford, Inglaterra, de que los genes se encuentran ordenados en módulos en la evolución, a partir de minigenes que corresponden a dominios proteicos estructuralmente estables.

Aunque la globina no posee dominios proteicos distintivos, sus aminoácidos están ordenados en regiones compactas discretas, cuatro en todas ellas, de acuerdo con Go. Mostró que la primera región se corresponde con la primera secuencia codificante, así como la última con el tercer exón. Pero las dos regiones del medio están codificadas por un único exón, el del medio de los tres. Tal vez ‘la presión selectiva eliminó la secuencia intercalada que estaba presente en un gen ancestral’, especuló Go en su artículo de comienzos de este año [*Nature (London)* 291, 90 (1981)].

Al mes siguiente un segundo artículo de *Nature* ofrecía algunos datos que corroboraban esto determinados a partir de una fuente inesperada: los nódulos radicales de las plantas leguminosas. La fijación del nitrógeno es la principal función de los nódulos radicales, y la presencia de una proteína que elimina eficazmente al oxígeno permite que el proceso continúe sin obstáculos. Esa proteína es la leghemoglobina, una molécula que por algún tiempo ha sido reconocida como muy similar en estructura a la hemoglobina de los vertebrados. Cuando K. A. Marcker y sus colegas de la Universidad de Aarhus, Dinamarca, publicaron en junio su análisis del gen de la leghemoglobina [*Nature (London)* 291, 677 (1981)], estuvo claro que la similitud no era producto de la evolución convergente, como usualmente se suponía. Los genes de vertebrados y plantas eran estrechamente homólogos, excepto que el gen de la leghemoglobina posee tres intrones (y por lo tanto cuatro exones), y no dos.

Marcker concluyó que sus resultados no apoyaban la idea de que los exones se corresponden con los dominios funcionales de las proteínas. Lo que ellos no sabían cuando escribieron esto es que el intrón ‘extra’ de la leghemoglobina cae precisamente en la división en dos regiones compactas del exón central de la globina de los vertebrados, como fue revelado por el análisis de Go. Una simple pérdida del intrón extra de la leghemoglobina haría que el gen fuera equivalente a la globina de los vertebrados.

La sección de la molécula de globina que está codificada por el exón central es responsable de la unión con el componente hemo de la hemoglobina, con los contactos principales distribuidos de manera equitativa entre lo que Go ha identificado como la segunda y la tercera región compacta de la estructura de la proteína. Tal vez bien atrás en la historia evolutiva temprana existía una pequeña proteína de unión al grupo hemo que era codificada por un gen con dos exones. Más tarde, otros exones fueron añadidos para dar el gen primordial de la globina, el cual perdió entonces el intrón que divide a la región de unión al grupo hemo. La similitud de parte de la estructura de ciertas proteínas de citocromos (que también se unen a grupos hemo) con el exón central de la globina apoya la ida de un antiguo gen de unión al grupo hemo, el cual viajó entonces a lo largo de varios caminos evolutivos para dar una familia de diferentes proteínas de unión al grupo hemo.

La pregunta exasperante es, ¿cómo las leguminosas adquirieron su gen tipo globina? ‘Una ancestralidad común que se remonte a la época de la división entre plantas y animales parece estar descartada’, dice Jeffreys. ‘Quizás haya habido una transferencia horizontal en tiempos recientes’; sugiere, ‘tal vez el gen ha sido transportado por un insecto que portaba un virus patógeno de las plantas.’ Esta intrigante posibilidad implicaría que la hemoglobina de los insectos posee también tres intrones y que la línea evolutiva que llevó a los vertebrados perdió el intrón ‘extra’ antes de que el gen primordial se duplicara para establecer los grupos  $\alpha$  y  $\beta$ . El análisis de la globina de los insectos es aguardado con considerable interés.

La heroica devoción al análisis físico de los loci de las globinas en años recientes, en particular en los laboratorios de Leder, Tom Maniatis, en el Instituto de Tecnología de California (ahora en Harvard), Oliver Smithies, Universidad de Wisconsin, Jeffreys, y otros, ha abierto el camino para una arqueología molecular considerable. ‘Ahora es posible trazar la evolución de los agrupamientos de las globinas con cierto detalle, en particular del agrupamiento de la  $\beta$ -globina, desde el comienzo de la era de los mamíferos’, dice Jeffreys.

El agrupamiento de la  $\beta$ -globina humana es el más complejo de los analizados hasta ahora. Además de los cinco genes funcionales de globinas, el locus de 60 kilobases también contiene dos de los llamados pseudogenes. Los pseudogenes son relativamente recientes para la biología molecular y aún son de alguna manera enigmáticos. La mayoría poseen un parecido cercano con un gen conocido, aunque están claramente discapacitados en su estructura por

adiciones y deleciones que impedirían la transcripción y la traducción normales. Algunos investigadores han sugerido que los pseudogenes podrían ser importantes en la regulación de la actividad de genes vecinos. Otros los consideran como productos que han divergido de la duplicación de genes, no necesariamente relictos del cambio evolutivo, sino, en cambio, genes nuevos potenciales.

Leder señala que los pseudogenes serían unidades muy fértiles de cambio evolutivo. ‘Una vez que un gen duplicado se ve liberado de la presión selectiva es libre de divergir en su secuencia de cualquier forma’, dice. ‘Los pseudogenes poseen la gran ventaja de llevar con ellos todas las señales de procesamiento necesarias para la normal generación, tales como los sitios de inicio de la transcripción y los sitios de corte de los intrones.’

Los pseudogenes poseen incluso un mayor potencial para el cambio evolutivo que lo que implica esta descripción, como será aparente más tarde.

A pesar de que el complejo de la  $\beta$ -globina humana contiene un número relativamente grande de genes activos, el 95 por ciento del locus está compuesto por ADN que no codifica para proteínas. ¿Cuál es el rol de este ADN extra, si es que lo tiene? Los pseudogenes constituyen sólo una pequeña proporción de la región, aunque podrían existir más pseudogenes. Algo de ese ADN está compuesto por representantes de familias muy conocidas de secuencias repetitivas. Y el restante es ADN sin función conocida ni secuencias comparables.

‘Queríamos poner a prueba la hipótesis de que este ADN extra es ‘ADN chatarra’, dice Jeffreys, ‘así que comparamos el loci  $\beta$  en humanos, gorilas y babuinos.’ Jeffreys y sus colegas razonaron que si se trataba de ADN chatarra, en los más de 20 a 40 millones de años de evolución representada por los humanos, los grandes simios y los monos del Viejo Mundo, era esperable que variara tanto la secuencia como la cantidad global de ADN intergénico. ‘Descubrimos que el agrupamiento de genes es remarcablemente estable’, reporta Jeffreys. ‘El patrón y el tamaño global del agrupamiento es el mismo, y la tasa de sustitución de nucleótidos es de un cuarto a un quinto de la que se esperaría para ADN inservible.’ El ADN no codificante parece por tanto no ser chatarra, pero qué función podría cumplir es aún un misterio.

Un paso aún más atrás en el tiempo evolutivo, a hace 40 millones de años, cuando los monos del Nuevo y del Viejo Mundo se separaron, presenta de pronto un cuadro diferente. Los genes  $\epsilon$  y  $\gamma$  se encuentran mucho más cerca el uno del otro; el gen  $\gamma$  aún no se ha duplicado; y el complejo entero es sustancialmente más pequeño.

Aún otro paso atrás en el pasado, a hace 70 millones de años, a los comienzos del orden de los primates y cerca de la diversificación explosiva de los mamíferos, lleva a un agrupamiento  $\beta$  muy simple. El lemur marrón (un prosimio) posee sólo uno de cada uno de los genes  $\epsilon$ ,  $\gamma$  y  $\beta$  contenidos dentro de un agrupamiento muy pequeño. Posee también un pseudogen. La similitud entre el agrupamiento de genes del lemur y el del conejo indica que este ordenamiento posiblemente refleja muy estrechamente el agrupamiento  $\beta$  básico de los mamíferos, establecido tal vez hace 80 a 90 millones de años atrás.

El incremento en la complejidad del agrupamiento de la  $\beta$ -globina sigue el patrón conceptualmente simple de duplicación de genes seguida por la diversificación de la secuencia del nuevo gen, para producir eventualmente un pariente que actúa en un punto diferente del programa de desarrollo. De este modo, en humanos el gen  $\beta$  original se encuentra ahora acompañado por cuatro vecinos estrechamente relacionados (además de los pseudogenes).

La historia no es tan simple y directa, sin embargo. Por ejemplo, aunque el gen  $\gamma$  se duplicó hace unos 20 a 40 millones de años, las secuencias de ambos genes en humanos son tan similares que parecen sugerir una separación muy reciente. La causa de esta anomalía, aparentemente, es el fenómeno de evolución concertada. Los genes repetidos en tandem, como lo son los dos  $\gamma$ , intercambian secuencias por la acción de varios mecanismos, con el resultado de que sus estructuras permanecen muy similares.

Otra rareza es la ausencia de la proteína  $\delta$ -globina en los monos del Viejo Mundo, aún cuando existe un gen en el lugar correcto del agrupamiento. Al parecer el gen ha sido apagado recientemente. En otras palabras, se trata ahora de un pseudogen.

Una de las mayores curiosidades es el pseudogen del lemur. La sección de adelante, 5', del gen se parece al gen  $\epsilon$ , mientras que la sección de atrás, 3', es del tipo  $\beta$ . Jeffreys y sus colegas especulan que el mecanismo de corrección entró dos veces en acción, una vez con el gen  $\beta$  como ‘modelo’ y otra con el gen  $\epsilon$  como modelo, y el resultado es un pseudogen en mosaico.

Los eventos más tempranos en la evolución de los genes de las globinas son los más difíciles de discernir. La suposición de que el gen primordial se duplicó para dar un par en tandem está apoyada por la disposición simple de un único gen  $\alpha$  y un único gen  $\beta$  como vecinos en uno de los cromosomas del anfibio *Xenopus tropicalis*.

La pregunta siguiente es, ¿cómo se separaron los agrupamientos  $\alpha$  y  $\beta$  en diferentes cromosomas? ¿Fue una copia de uno de los genes transportada hacia otro cromosoma, donde fue insertada, con la copia original cayendo gradualmente en desuso? ¿O fue quitado uno de los genes de su posición original y reubicado en cualquier parte? ¿O quizás se duplicó todo el conjunto del cromosoma (tetraploidización) en un ancestro antiguo, seguido por el apagado del gen  $\alpha$  en un lugar y del gen  $\beta$  en el otro? El apoyo a este último mecanismo proviene de *X. laevis*, un pariente evolutivo de *X. tropicalis*. El evento evolutivo involucró aquí a la tetraploidización de forma que *X. laevis* posee dos

agrupamientos de globinas, cada uno de los cuales contiene un único gen  $\alpha$  y un único gen  $\beta$ . La divergencia de secuencia en esta especie parece estar acallando a los genes  $\alpha$  y  $\beta$  de un mismo agrupamiento. No obstante, se sabe que la tetraploidización ha sido importante en la evolución de los vertebrados, y el silenciamiento diferencial anteriormente especulado es una firme posibilidad.

La evolución de los genes, ejemplificada por la familia de las globinas, es vista por lo tanto como un asunto dinámico, con los cambios más importantes manifestándose en forma escalonada. La fuente de los cambios es principalmente la duplicación de genes seguida por la divergencia de secuencias que crean así a los miembros relacionados, aunque distintivamente diferentes, de la familia de las globinas. Además de la evolución dentro del locus, el agrupamiento de las globinas puede ser la fuente de nuevos genes en cualquier lugar.

El trabajo en los laboratorios de Leder y Smithies ha revelado en años recientes la existencia de dos pseudogenes  $\alpha$  en el ratón. Uno de ellos era un pseudogen convencional, ya que poseía cambios de bases que impedirían su expresión. El segundo, sin embargo, es el pseudogen más intrigante descubierto hasta el momento: sus intrones han sido perfectamente cortados.

La eliminación de los intrones en uno de los pseudogenes fue la primer sorpresa. La segunda surgió cuando Aya Leder intentó descubrir exactamente dónde se ubicaban los pseudogenes en el agrupamiento de las globinas. No lo estaban. Se encontraban en cromosomas enteramente diferentes, no en el agrupamiento  $\alpha$  como se esperaba.

¿Cómo se movieron los pseudogenes desde su ubicación original hacia cromosomas diferentes? Por varias razones, incluyendo la precisa escisión de los intrones, Philip Leder fue atraído por la sugerencia de Stephen Goff y David Baltimore de que el gen sin secuencias intercaladas fue transportado por la acción de un retrovirus. Existen secuencias típicas de retrovirus a ambos lados del gen, pero su disposición está de alguna forma dividida. Debe ser posible que la inserción del pseudogen y el pasaje del retrovirus por esta región fueran eventos desconectados, en cuyo caso deberá buscarse otra explicación para explicar el movimiento del gen con la pérdida de sus intrones.

La transposición del segundo pseudogen pudo haber ocurrido por un mecanismo mucho más convencional, posiblemente involucrando el entrecruzamiento desigual, un proceso que por razones mecánicas no es fuera de lo común en genes repetidos en tandem.

En su búsqueda de los pseudogenes  $\alpha$  del ratón los trabajadores del NIH aumentaron la sensibilidad de su sistema de detección; en vez de identificar sólo dos, ahora poseen evidencia de unos diez, aunque la ubicación de éstos no ha sido aún determinada. Leder asemeja la emisión constante de pseudogenes por parte del locus de la globina a un volcán: 'Es un modelo tipo Vesubio', bromea. Esto implica que todas las familias de genes son fuentes potenciales de nuevos genes, a través de la transposición, la subsiguiente divergencia de la estructura de los genes, y su eventual reclutamiento para cumplir una nueva función.

Se ha demostrado recientemente que la familia de genes de las histonas emiten copias de genes que se integran fuera del agrupamiento principal (*Science*, Mayo 1, p. 530). este locus parece ceñirse por lo tanto al modelo tipo Vesubio. Lo que ahora se requiere es un análisis general de la estructura de los genes que revele los ecos de las relaciones evolutivas que deben existir entre muchos genes, si el modelo es correcto en sus implicaciones evolutivas.

Los genes de las globinas han probado ser de hecho una fértil fuente de datos y una agudo estímulo para las ideas en el campo en rápido crecimiento de los aspectos moleculares del cambio evolutivo.

#### **Filogenia de los agrupamientos de la $\beta$ -globina.**

Los mapas de los agrupamientos de la  $\beta$ -globina revelan la secuencia en la que ciertos genes cambiaron a través del tiempo evolutivo. La sorprendente estabilidad del agrupamiento entre los primates del Viejo Mundo, grandes simios y humanos está indicada por el tamaño y la organización constantes del agrupamiento. Las ubicaciones de los pseudogenes se indican con cruces.