

La evolución de MOLÉCULAS y de sistemas BIOLÓGICOS

Amanda Castillo Cobián, Agustino Martínez-Antonio y Alexander de Luna Fors

En última instancia la evolución supone cambios en la estructura genética de las poblaciones. Estos cambios, incluso si son seleccionados por sus efectos en el organismo completo, son en principio arbitrarios y dependen de la naturaleza química del material genético.

Introducción

La idea central derivada de la teoría evolutiva propuesta por Darwin, acerca de la evolución de las especies, señala que todos los organismos poseen un ancestro en común y que han evolucionado de éste a partir de modificaciones subsecuentes. Además, el concepto implica que el mecanismo principal por el que la evolución actúa es la selección natural de las variaciones hereditarias. Es decir, Darwin plantea que la selección natural es la responsable directa de la evolución y la adaptación de los organismos a su medio ambiente. A partir de estas ideas se generaron dos campos de estudio dentro de la biología: el estudio de la historia evolutiva de los organismos (filogenia) y la elucidación de las fuerzas evolutivas que moldean la biodiversidad, mediante las adaptaciones que producen.

La selección natural en un sentido amplio se refiere a que algunos individuos en una población presentan una mayor adaptación a un ambiente dado y, por tanto, dejan un mayor número de descendientes que otros individuos de la misma población en las mismas condiciones. Es decir, esta ventaja adaptativa constituye un diferencial entre supervivencia y reproducción de algunos genotipos sobre otros (genotipo: se refiere a un grupo de organismos que comparten la misma información genética).

Los biólogos evolutivos distinguen en la actualidad dos tipos de selección natural: 1) la selección purificadora o negativa, que actúa eliminando cambios o mutan-

Darwin plantea que la selección natural es la responsable directa de la evolución y la adaptación de los organismos a su medio ambiente



tes, y 2) la selección positiva, del tipo darwiniano, que favorece la fijación de cambios en los genomas. A su vez, la selección positiva puede considerarse de dos tipos: la direccional, que promueve la fijación de un alelo ventajoso (un gen en una población); y la balanceadora, que tiende a incrementar el polimorfismo (variación dentro de un mismo gen); es esta última la que más se acerca a lo que Darwin se refirió como la selección natural.

Nacimiento del estudio de la evolución molecular

Se considera que la evolución molecular nace a partir del trabajo de Zuckerkandl y Pauling en 1965, donde subrayan la importancia de la secuencia de aminoácidos de las proteínas como documentos históricos. Más aún, estos autores proponen que el cambio en la secuencia es proporcional al tiempo transcurrido entre especies, lo que derivó en la propuesta del reloj molecular. Sin embargo, la idea de la evolución molecular se puede rastrear aún más atrás en el tiempo, a partir de los estudios de G. H. F. Nuttall en 1904, quien trabajó con proteínas de la sangre y sus reacciones

inmunológicas, proponiendo por primera vez que las comparaciones de estas reacciones nos pueden indicar las relaciones filogenéticas entre los primates y otros grupos de animales. Las implicaciones de estas ideas fueron el eje central que rigió el estudio de la evolución molecular hasta nuestros días. La primera premisa es que podemos estudiar las relaciones filogenéticas entre biomoléculas de la misma forma que estudiamos relaciones entre especies; es decir, que la historia evolutiva de las proteínas o el ADN nos puede reflejar la historia de las especies. La segunda premisa es la existencia de un reloj molecular; es decir, que los cambios fijados en las proteínas y en el ADN pulsan de manera constante y nos dan un registro directo del tiempo transcurrido. Estas ideas sentaron las bases teóricas para el estudio de la evolución a nivel molecular y han tenido impacto en todas sus ramas, desde la propuesta de los modelos de sustitución en proteínas y ADN, hasta los estudios de fecha- ción molecular.

Sin embargo, para lograr comprender el papel de la evolución molecular dentro del desarrollo de la biología evolutiva moderna es necesario estudiarla en el contexto de la denominada “gran síntesis evolutiva” ocurrida entre los años 30s y 50s, y que involucró el trabajo de las mentes más brillantes de su tiempo. La síntesis evolutiva intenta hacer una incorporación de las ideas darwinistas y la genética mendeliana que resultó en el surgimiento de la genética de poblaciones. Los tres principales arquitectos de esta síntesis fueron Sewall G. Wright, Ronald A. Fisher y J. B. S. Haldane, quienes modelaron la distribución de las frecuencias alélicas en las poblaciones como resultado de la interacción de las fuerzas evolutivas, tales como la selección, la mutación, la migración y la deriva génica. La síntesis evolutiva es central en biología ya que nos da un marco teórico de referencia sobre el cual analizar los procesos evolutivos de una manera formal en la biología evolutiva moderna.

A partir de esta época comenzó una acumulación de trabajos en genética de poblaciones que mostraron que existía una gran variación genética en los genes y las proteínas, misma que no podía ser justificada del todo por efectos de la selección natural. Se inició así un debate acerca de la naturaleza del polimorfismo molecular y sobre si la evolución actuaba de la misma forma a nivel organismo y molécula. Los neo-darwinistas consideraban que el tamaño de las poblaciones era grande (infinito para sus modelos estadísticos) y tomaban como modelo de selección, por ejemplo, caracteres con importancia agronómica, sin parecerles importante la mutación como generadora de adaptaciones. Este tipo de discusiones hicieron que los debates



giraran en cómo buscar evidencia de la selección tipo darwiniana en las poblaciones biológicas.

Es durante el período de la síntesis evolutiva que Haldane calcula lo que denomina el costo de la selección natural, que se refiere al límite de substituciones que un mamífero puede poseer en un gen como resultado único de la selección natural. El cálculo resultó en aproximadamente una substitución génica cada 300 generaciones o cada 1 200 años, si el tiempo generacional promedio de un mamífero es de 4 años. Haldane desarrolla, a su vez, el concepto de carga genética, que hace posible el cálculo del número de variantes alélicas que pueden existir en una población natural, y cómo el mantenimiento de esta variación por selección haría que la carga de las mutaciones deletéreas fuese muy grande al pasar el tiempo, tanto que no podría sostener a la población y terminaría por extinguirse. Estas ideas, en conjunto con la acumulación de evidencia empírica del polimorfismo en proteínas y ADN, comenzaron a generar polémica sobre la participación de la selección en el mantenimiento del polimorfismo molecular, lo que culminó con la propuesta alternativa de la teoría neutra.

Teoría neutra

La propuesta de la teoría neutra fue desarrollada por Kimura (1968) en conjunto con King y Jukes (1969). Usando las hemoglobinas de los mamíferos como punto de partida, donde el número promedio de substituciones nucleotídicas, al año por genoma, es de 4×10^9 , calcularon que existe, en promedio, una substitución cada dos años. Si esto fuese cierto sería una tasa de cambio altísima, muy superior al límite propuesto por Haldane, para ser justificado únicamente por la selección natural darwiniana. Por esta razón Kimura propone que la mayoría de estas substituciones son neutras, o casi neutras, a la selección natural. De acuerdo a la teoría neutra, la mayor parte de la variación observada a nivel molecular no es mantenida por la selección natural de tipo darwiniano, sino que es resultado de un balance entre la deriva génica (azar) y las mutaciones neutras o casi neutras. Es decir, que la mayor parte de la variación a nivel molecular ocurre de manera aleatoria y no tiene importancia adaptativa. Otra de las predicciones de la teoría neutra es que las mutaciones no se producirían de manera uniforme en todas las proteínas o al interior de las mismas. Es decir, proteínas que participasen en funciones esenciales serían más restringidas al cambio y sus tasas de cambio serían mucho menores que las proteínas que no tuviesen tanta res-

De acuerdo a la teoría neutra, la mayor parte de la variación observada a nivel molecular no es mantenida por la selección natural de tipo darwiniano, sino que es resultado de un balance entre la deriva génica (azar) y las mutaciones neutras o casi neutras

tricción funcional. De igual manera, dentro de las proteínas encontraríamos regiones que tendrían mayor restricción funcional; por ejemplo, los sitios activos comparados con el resto de la proteína.

Sin embargo, la relación entre la diversidad genética con el tamaño efectivo poblacional por tasa de mutación ($\theta = 4N_e\mu$, donde θ = diversidad genética, N_e = tamaño efectivo poblacional, μ = tasa de mutación), es lo que hace a la teoría neutra tan central a la biología evolutiva moderna. En especies que posean tamaños efectivos poblacionales pequeños, las mutaciones deletéreas se ven incrementadas por causa de la deriva génica; de alguna forma podríamos interpretar a la deriva génica como el error de muestreo, en donde en poblaciones muy grandes es poco evidente pero en poblaciones pequeñas se vuelve relevante.

Conforme comenzó a acumularse información a nivel de ADN, Kimura señaló que, de acuerdo a la teoría neutra, se esperaría que en las secuencias codificantes el número de substituciones sinónimas fuese mayor que el de las substituciones no sinónimas ya que estas últimas al producir un cambio podrían tener un efecto deletéreo en la función de la proteína y serían fácilmente eliminadas por la selección

purificadora, mientras que las sustituciones sinónimas permanecerían neutras a la selección.

Adaptación a nivel molecular

En la actualidad, la mayoría de los trabajos que intentan estudiar la adaptación a nivel molecular lo hacen evaluando los cambios sinónimos (dS) y no sinónimos (dN) a nivel codones (tripletes de bases que codifican para un aminoácido) bajo los supuestos de neutralidad o selección positiva (darwiniana) como hipótesis nulas. La relación más utilizada es el cociente de las tasas de sustitución no sinónimas sobre las sinónimas dN/dS. Este cociente nos indica la relación entre ambas tasas y supone un punto de partida para la evaluación estadística de los supuestos neutros o selectivos. Si su valor es mayor a 1 suponemos que la selección es positiva; si este valor es igual a 1, entonces ambas tasas son similares y nos encontramos bajo el modelo neutro; si el valor es menor de 1 suponemos que la selección es purificadora. Esta relación sencilla ha

tenido un impacto muy grande en los trabajos de detección de selección positiva, neutra o purificadora a nivel molecular. Además, trajo consigo la aparición de nuevas metodologías de estudio como fue el desarrollo de los modelos de sustitución de los codones como unidad mínima evolutiva y funcional dentro de las proteínas. En cada codón, que corresponde a un tipo de aminoácido, se encuentran tres posiciones denominadas primera, segunda y tercera, correspondientes a las tres bases nucleotídicas que lo integran. Cada una de estas posiciones se ve afectada de manera diferencial por la mutación y la selección. De acuerdo al código genético, si un cambio es introducido en la primera posición en su mayoría será un cambio sinónimo y un porcentaje bajo sería no sinónimo; es decir, se mantendría codificando para el mismo aminoácido, mientras que todos los cambios introducidos en la segunda posición corresponden a un cambio no sinónimo (que cambia el tipo de aminoácido). Sin embargo, si un cambio se introduce en la tercera posición, en su mayoría corresponderá a un cambio sinónimo (neutro; codón señalando las tres posiciones [Figura 1]). Por lo tanto, a nivel molecular el segundo nucleótido en el codón es el que nos brinda la posibilidad de evaluar si un cambio podría ser adaptativo y si estará sujeto a la selección natural. Es entonces, que tenemos la posibilidad de evaluar el genotipo y fenotipo al mismo tiempo, donde el genotipo corresponde a la secuencia de ADN y el fenotipo a la proteína, usando como base funcional a los codones. Estos procesos son muy difíciles de evaluar a nivel morfológico en un organismo ya que los cambios morfológicos implican la participación de muchos factores como son los genes, las interacciones de sus productos y el medio ambiente. De esta manera es mucho más fácil evaluar la huella de la adaptación a nivel molecular que a nivel morfológico. Como resultado, esta metodología nos ha brindado un primer puente entre los procesos macro y microevolutivos. Pero en realidad, ¿qué se sabe en la actualidad acerca de la evolución molecular a nivel del gen y del genoma?

El ejemplo clásico de adaptación a nivel molecular es el que involucra al complejo mayor de histocompatibilidad. Este complejo está formado por una familia multigénica (comprende varios genes) que participan en la generación de la respuesta inmune en los vertebrados. Existen dos clases: Clase I, que presentan antígenos a las células T citotóxicas o CD8; Clase II, que presentan antígenos a las células T (ayudadoras) o CD4. Mediante experimentos de trasplantes en ratones fue descubierta una gran cantidad de variación en los genes de la Clase I. Sin embargo, la naturaleza del polimorfismo era total-

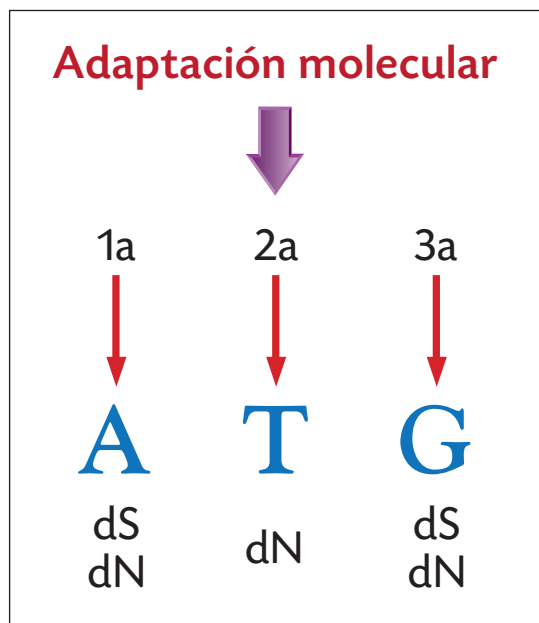


Figura 1. En la actualidad la adaptación a nivel molecular se estudia con base en las tasas de sustitución (sinónimas y no sinónimas) a nivel de nucleótidos, usando al codón como unidad evolutiva.

mente desconocida. Posteriormente fue descubierto que las proteínas de estos genes presentaban antígenos a las células T citotóxicas, activándolas para que éstas a su vez combatan a las células infectadas con patógenos intracelulares. Estas proteínas interactúan directamente con antígenos del patógeno para desencadenar el ataque de la respuesta inmune, lo que resulta en un proceso de co-evolución con el patógeno. Este hallazgo sentaba las bases biológicas para la comprensión del proceso por el cual se mantiene y promueve la diversidad genética de estas proteínas. A continuación se cristalizó la estructura del complejo de clase I y se reveló el sitio exacto de unión del antígeno con el péptido del complejo. Esto brindó una forma de probar que en esta región el polimorfismo es mantenido por selección positiva balanceadora, puesto que se observó un incremento en el cociente dN/dS , en comparación con el resto de la proteína. Este estudio es considerado el primer ejemplo indiscutible y clásico de adaptación a nivel molecular.

A partir de este trabajo, y con la acumulación de grandes bases de datos de secuencias de genes y del desarrollo de nuevos métodos estadísticos de análisis como aquellos basados en máxima verosimilitud, comenzó el estudio masivo de la adaptación a nivel molecular. Esto bajo la premisa de encontrar señales de selección positiva darwiniana a nivel molecular y extrapolarlas a la búsqueda de la función, o los sitios de importancia dentro de las proteínas, de manera experimental. Sin embargo, hasta ahora son pocos los genes donde se ha detectado la participación significativa de la selección positiva de manera clara (Tabla 1). La generalidad es que estos genes pertenecen a familias génicas con funciones definidas, tales como el sistema inmune, involucrados en reproducción y apoptosis o muerte celular. La mayoría de los ejemplos corresponden a eucariontes (animales, hongos, plantas y protistas), mientras que existen pocos ejemplos descritos en procariontes (bacterias). Entre estos últimos destacan las colicinas, que serían el equivalente molecular del sistema inmune en bacterias, y una subunidad de las ATPasas que se ve involucrada en transporte y síntesis de ATP. Sin embargo, recientemente se ha visto que el número de genes que han pasado por selección positiva es grande en bacterias; en el caso

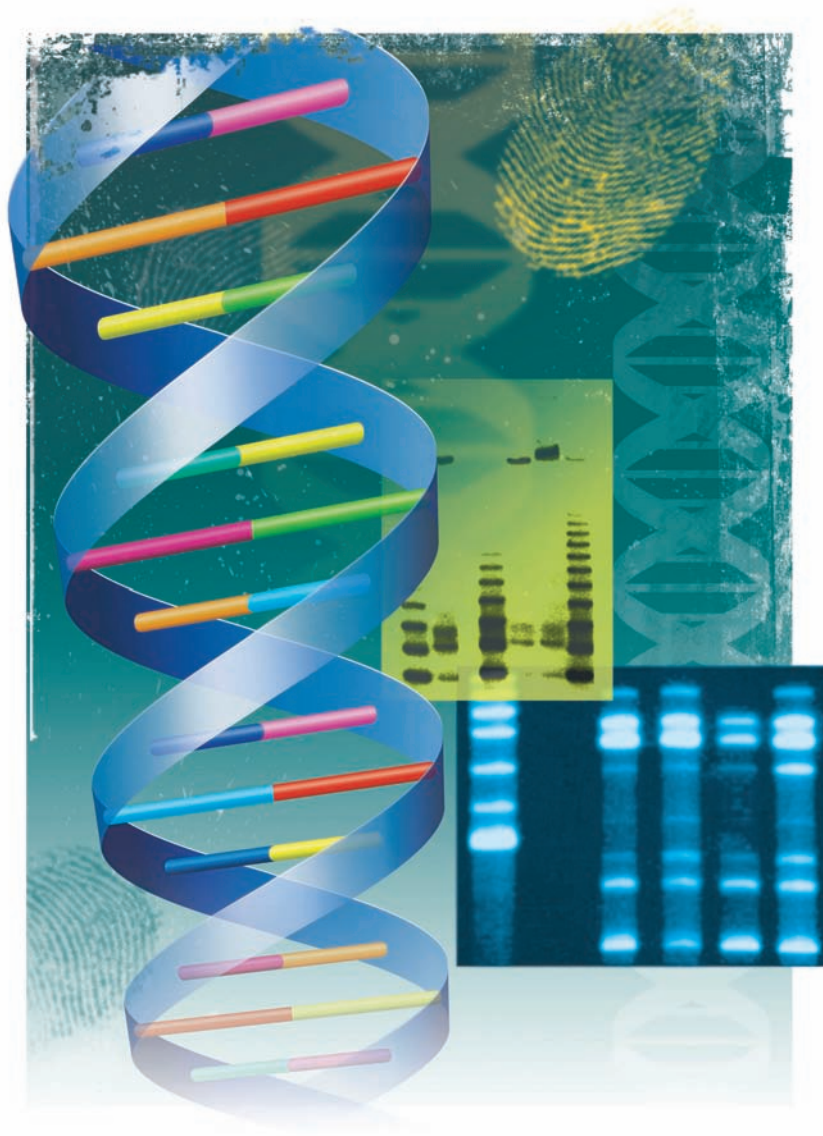


Tabla 1. Genes involucrados en actividades celulares de varias clases que se encuentran bajo selección positiva según diferentes métodos moleculares.

Gen	Organismo
Genes que participan en la respuesta inmune	
Quitinasas clase I	<i>Arabis</i> y <i>Arabidopsis</i>
Colicinas	<i>Escherichia coli</i>
Defensinas	Roedores
Inmunoglobulinas V _H	Mamíferos
Genes del Complejo Mayor de Histocompatibilidad	Mamíferos
Inhibidor de la poligalacturonasa	Dicotiledóneas y legumbres
Genes de grupo sanguíneo RH (RH50)	Primates y roedores
Ribonucleasas	Primates
Gen de la transferrina	Salmones
Interferon- γ tipo I	Mamíferos
Inhibidor de la proteinasa a-1	Roedores
Genes involucrados en la evasión de la respuesta a sistema inmune	
Gen de la cápside	Virus de la fiebre aftosa
Genes CSP, TRAP, MSA-2 y PF83	<i>Plasmodium falciparum</i>
Región codificadora del antígeno delta	Virus de la hepatitis D
Gen de la envoltura de la cápside	Virus HIV
Glicoproteína gH	Virus de la pseudorabia
Genes del antígeno de la invasión del plásmido	<i>Shigella</i>
Antígeno-1 de la superficie del merozoito (MSA-1)	<i>Plasmodium falciparum</i>
Proteína de la membrana externa	<i>Chlamydia</i>
Porina 1 (porB)	<i>Neisseria gonorrhoeae</i> y <i>N. meningitidis</i>
Glicoproteínas S y HE	Coronavirus del ratón
Sigma 1	Reovirus
Gen determinante de la virulencia	<i>Yersinia</i>
Genes involucrados en la reproducción	
Proteína involucrada en la fertilización 18 kDa	<i>Haliotis</i> (Abulón)
Gen Acp26Aa	<i>Drosophila</i>
Proteína que une al andrógeno	Roedores
Hormona de puesta de huevos	<i>Aplysia californica</i>
Gen de caja homeótica Ods	<i>Drosophila</i>
Gen de caja homeótica Pem	Roedores
Protamina P1	Primates
Lisina del esperma	<i>Haliotis</i>
RNAse S	Rosaceae
Gen de la diferenciación testicular Sry	Primates
Genes involucrados en la digestión	
Caseína K	Bovinos
Lisozima	Primates
Toxinas	
Conotoxina	Gastrópodo <i>Conus</i>
Fosfolipasa A ₂	Serpientes crotalinas
Genes relacionados al transporte de electrones y síntesis de ATP	
Subunidad F ₀ de la ATP sintetasa	<i>E. coli</i>
Isoforma COX7A	Primates
COX4	Primates

Tabla 1. Continuación.

Gen	Organismo
Citoquinas	
SF de Granulocito-macrófago	Roedores
Interleucina-3	Primates
Interleucina-4	Primates
Misceláneos	
CDC6	<i>Saccharomyces cerevisiae</i>
Hormona de crecimiento	Vertebrados
Cadena b de la hemoglobina	Peces antárticos
Gen Jingwei	<i>Drosophila</i>
Péptido C3 de la prostateína	Rata

de *Escherichia coli* alrededor de 56% de las substituciones de aminoácidos parecen haber sido fijadas por selección positiva.

Por otra parte, a partir de la comparación de miles de genes de mamíferos se ha estimado que el promedio del cociente dN/dS en este grupo se encuentra en el intervalo de 0.10 a 0.25. Esto significa que aproximadamente 75-90% de las sustituciones no sinónimas son eliminadas por la selección purificadora. En el caso de la población humana la cifra es del 75%. La comparación de los linajes del humano y del ratón común arroja un promedio de dN/dS de 0.115, mientras que entre humanos y chimpancés es de 0.23. Por otra parte, utilizando información derivada de 5 286 genes compartidos entre 5 genomas completos de mamíferos, el promedio del cociente dN/dS fue mayor en los linajes que llevan a los humanos (0.169) y chimpancés (0.175), intermedio en los macacos (0.124) y perros (0.128), y más bajo en los roedores (0.104). Estos resultados señalan la existencia de diferencias significativas en el cociente dN/dS a lo largo de los distintos linajes de mamíferos. Lo que se puede concluir a partir de estos resultados es que en los linajes que poseen tamaños poblacionales efectivos más pequeños (como los primates) el cociente dN/dS será mayor que en los linajes cuyos tamaños poblacionales son mayores (como los roedores). En resumen, lo que estos trabajos demuestran es que al parecer la selección natural positiva darwiniana no es la principal fuerza evolutiva dentro de la evolución a nivel molecular en las regiones codificantes de los genomas.

La selección a nivel genómico

La genómica comparativa nos permite estudiar la participación de la selección natural a gran escala, a nivel de genomas. Con la llegada de la era de la genómica, la acumu-

lación en las bases de datos de secuencias y de genomas completos ha sido exponencial. Esto ha representado un nuevo reto para el desarrollo de herramientas y metodologías con la capacidad de análisis suficiente para el estudio de cientos de genomas con millones de pares de bases nucleotídicas. En seguida presentamos algunos ejemplos de organismos emblemáticos de estudio que muestran lo que se sabe acerca de la participación de la selección natural a nivel genoma, cosa que era impensable de alcanzar hace unos 20 años.

El análisis de doce genomas completos de especies pertenecientes al género *Drosophila* (mosca de la fruta) arrojó como resultado la propuesta de que al menos 33.1% de sus genes compartidos han experimentado algún episodio de selección positiva, dentro de los cuales se destacan los genes que participan en interacciones con el medio ambiente, los sexuales y los involucrados en la reproducción. En el caso de los grandes primates u homínidos (incluyendo al humano) el patrón selectivo no parece ser el mismo que el observado para la mosca de la fruta. El porcentaje de genes que han pasado por algún evento de adaptación a nivel molecular es muy bajo; varios trabajos calculan que se encuentra en el rango de 0.08% hasta el 6%. Tomando en cuenta que el ser humano posee aproximadamente 30 000 genes (se han calculado de 26 000 a 35 000) esto equivaldría sólo de 300 a 1 800 genes



La duplicación genética ocurre en un individuo y puede fijarse o perderse en la población, de manera muy similar a lo que ocurre con las mutaciones puntuales

aproximadamente. Ello mostraría que la selección positiva darwiniana no ha jugado un papel relevante en la evolución de los primates, lo contrario a lo observado en la mosca de la fruta.

En el caso de la planta *Arabidopsis* se ha encontrado también muy poca evidencia de selección positiva. Se ha descrito que en un estudio realizado en 304 genes, alrededor del 5% mostraron alguna señal de selección positiva, lo que es un porcentaje muy bajo; aunque si consideramos que se trata de una especie altamente endogámica sería razonable, ya que se encontraría muy castigada por la selección purificadora.

A partir del cálculo de la proporción de sustituciones potencialmente adaptativas se puede calcular lo que se denomina la tasa genómica de adaptación a nivel molecular. Por ejemplo, en el caso de *Drosophila*, si se ha propuesto que aproximadamente el 45% de sustituciones son del tipo adaptativo, se ha calculado que se fija una sustitución cada 45 años o 450 generaciones, lo que sería alrededor de 22 000 sustituciones de aminoácidos por millón de años. Pero, ¿qué efecto global, en el organismo, tienen estas sustituciones, si pensamos en las distintas especies de la mosca de la fruta que son casi idénticas a nivel morfológico, a pesar de las diferencias adaptativas a nivel molecular? Esta cuestión ha generado muchos debates acerca de si es la variación en las secuencias (sobre todo la de tipo adaptativo) o la variación en la expresión de los genes lo que constituye la base más importante para la evolución a nivel morfológico (fenotípico).

Evolución de secuencias codificantes y regulatorias de genes duplicados

Los genes duplicados juegan un papel primordial como materia prima en la evolución de nuevas funciones moleculares. En 1970 sale a la luz el libro *Evolución mediante duplicación genética* de Susumu Ohno, a partir del cual se populariza esta idea entre los biólogos. Sin embargo, no es hasta finales de la década de los noventa, gracias al análisis de diversas secuencias genómicas, que la omnipresencia e importancia de la duplicación genética queda claramente demostrada. Dichas secuencias genómicas muestran que, en los tres dominios de la vida (eucariontes, bacterias y arqueobacterias), una frac-

ción considerable de los genes detectados se han originado por duplicación genética. Estos números son particularmente notorios en organismos eucariontes: en el ser humano y en la mosca *Drosophila* alrededor del 40% de los genes tienen dos o más copias, mientras que el 65% de los genes de la planta *Arabidopsis* son el producto de eventos de duplicación genética.

La duplicación genética ocurre en un individuo y puede fijarse o perderse en la población, de manera muy similar a lo que ocurre con las mutaciones puntuales. Si la duplicación es selectivamente neutral comparada con el estadio previo a la duplicación, ésta tiene una baja probabilidad de fijarse: $1/2Ne$, donde Ne es el tamaño efectivo de la población. Esto sugiere que la mayor parte de los genes duplicados se pierden en las poblaciones. Para aquellos genes duplicados (parálogos) que logran fijarse en la población, su destino evolutivo estará determinado por las funciones que desempeñan. En general se reconocen cuatro escenarios posibles para las parejas de genes duplicados: la pseudogenización, la conservación funcional, la subfuncionalización y la neofuncionalización.

La pseudogenización ocurre cuando la presencia de uno de los genes duplicados es selectivamente neutral y comienza a adquirir mutaciones que inactivan la expresión o la función, convirtiendo al gen duplicado en un pseudogen. La proporción de pseudogenes varía en diferentes organismos; mientras que se calcula que en el genoma humano existe un pseudogen por cada ocho genes funcionales, en levaduras éstos son prácticamente inexistentes.

En el escenario opuesto, la conservación funcional se da cuando la presencia de dos copias es benéfica per se, debido probablemente al aumento en la cantidad del producto génico (efecto de dosis génica). Dicho fenómeno se ha observado en genes altamente expresados, tales como los genes para ARN ribosomal, proteínas ribosomales o histonas de levaduras. La conversión génica –recombinación entre las copias parálogas o duplicadas– y la selección purificadora son los procesos preponderantes en el mantenimiento de la función ancestral.

En el proceso de subfuncionalización (Figura 2), cada copia hija adopta una parte de la función ancestral, de modo tal que el mantenimiento de ambos parálogos se vuelve indispensable. Una forma de subfuncionalización que tiene relevancia en la evolución del desarrollo es la división de los patrones de expresión posteriores a la duplicación.

Finalmente, uno de los resultados más relevantes de la duplicación genética es el origen de nuevas funciones o neofuncionalización (Figura 2). Ésta puede ser vista como la aparición

Los genes duplicados juegan un papel primordial como materia prima en la evolución de nuevas funciones moleculares. En 1970 sale a la luz el libro *Evolución mediante duplicación genética* de Susumu Ohno, a partir del cual se populariza esta idea entre los biólogos

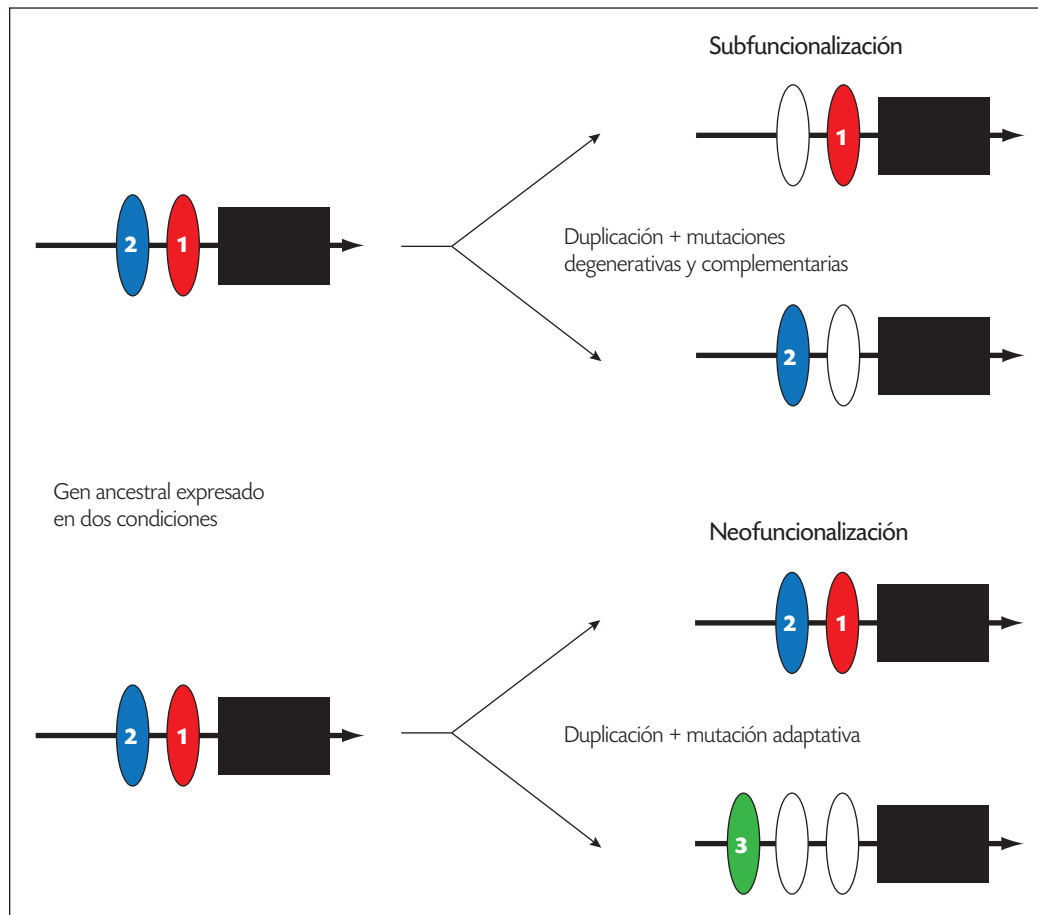


Figura 2. Dos escenarios posteriores a la duplicación genética. Los cuadros representan genes y los óvalos representan los elementos que controlan la transcripción (en este ejemplo, la divergencia funcional es definida en términos de las condiciones en las que se expresan los genes).

de una función completamente original (nueva función bioquímica), o bien el refinamiento de la función preexistente (por ejemplo, la expresión en una nueva condición). Aunque es improbable que una función completamente nueva se origine a partir de un gen duplicado, existen ejemplos documentados. A la fecha no es claro cuál es la proporción de genes cuya divergencia funcional ocurre por subfuncionalización o por neofuncionalización. En términos de los niveles de expresión, se ha sugerido que ambos procesos han determinado los perfiles de expresión de genes duplicados en la levadura. Es probable que lo que ocurre en muchos casos es una combinación de ambos procesos: la subfuncionalización juega un papel importante en el man-

tenimiento de las dos copias inmediatamente después de la duplicación, y ésta es seguida por un proceso constante de neofuncionalización.

La expresión de la divergencia entre genes duplicados es considerado un paso importante en la generación de nuevos genes a partir de una función redundante. De hecho, el mismo Ohno fue quien propuso que la divergencia en la expresión genética sería el primer paso en la divergencia funcional de genes duplicados, aumentando de esta manera la probabilidad de su retención en los genomas. De hecho, se ha encontrado que existen un gran número de genes duplicados que muestran cambios significativos en su expresión, lo cual ha demostrado que existe una gran probabilidad de divergencia en la expresión de estos genes entre especies e incluso dentro de una misma especie. Estas investigaciones computacionales han llevado a plantear la hipótesis que sugiere que en la evolución de genes duplicados, la regulación de la transcripción diverge más rápi-

damente que la función codificada en la proteína. Es decir, la función de la proteína se conserva por más tiempo que su regulación.

Evolución molecular del código regulatorio en los genomas

Por mucho tiempo se consideró que las regiones de ADN en los genomas que no codificaba para genes eran poco relevantes en cuanto a su función e incluso se llegó al extremo de considerar estas zonas como ADN basura. Ahora sabemos que las regiones intergénicas (entre los genes) juegan un papel muy importante para la expresión genética. Esto se debe a que existen secuencias de ADN de 4 a 8 nucleótidos que son regiones de contacto para proteínas reguladoras, también llamadas factores de transcripción. Cada factor de transcripción reconoce una secuencia específica de ADN. A su vez, los factores de transcripción normalmente tienen un dominio (región específica dentro de una proteína) adicional (al de unión al ADN) que reconoce una señal ambiental (por ejemplo, un azúcar o un aminoácido) o que sirve para hacer contacto con otras proteínas, incluyendo otros factores de transcripción. Finalmente, los genes que codifican para factores de transcripción se pueden auto-regular, o bien regular a otros reguladores, lo que formaría cascadas o redes de regulación entre factores de transcripción.

La maquinaria regulatoria está formada por proteínas que reconocen el código regulatorio en el ADN, y como producto de estas interacciones resultan los distintos perfiles de expresión de genes que normalmente son necesarios y suficientes para el desempeño de los organismos en cada condición ambiental donde normalmente se encuentran. En la actualidad aunque contamos con las secuencias genómicas de unos 1 000 organismos (la mayoría de bacterias) aún no conocemos para qué funciona cada uno de los genes de la bacteria más simple, que contiene unos 300 genes. Más aún, de cada nuevo genoma que se secuencía, al menos a una tercera parte de los genes no se les puede asignar una función. Lo que sí se puede hacer, mediante bioinformática, es buscar dominios o regiones conservadas en los productos proteicos de los genes desconocidos y asignarles una posible función. Tomando en cuenta este proceso se puede estimar el número de factores de transcripción codificados en los genomas de los organismos secuenciados. Y si se supieran los sitios de unión en el ADN en cada uno de ellos (o sus familias proteicas) se podría hacer un estimado de la

intensidad de la co-evolución entre las proteínas reguladoras y sus sitios de unión en el ADN. Este punto es relevante, para esta discusión, puesto que desde 1975 King y Wilson se dieron cuenta que el grado de divergencia existente entre proteínas de humanos y chimpancés difícilmente explicaba los cambios morfológicos y de comportamiento tan contrastantes entre ambas especies. De esta manera se propuso que además de la evolución de las zonas codificantes en el genoma (que resulta en divergencia proteica) también debiera contribuir la evolución de las zonas regulatorias (resultando en distintos niveles de expresión de las proteínas).

Lo que conocemos de la regulación transcripcional en los modelos bacterianos mejor estudiados (*Escherichia coli*, *Bacillus subtilis* y sus extrapolaciones hacia otras bacterias) nos enseña que conforme el tamaño de los genomas aumenta (desde unos 300 hasta 8 000 genes) el requerimiento de la maquinaria transcrip-



cional no se incrementa en la misma proporción sino que se cuadruplica, y que los genes reguladores, por lo general no se mantienen a la par que los genes regulados. Es decir, que el cableado de regulación transcripcional en los genomas de bacterias es tremendamente flexible y, al parecer, los genes regulados se conservan más que sus regiones regulatorias (y por ende que sus proteínas reguladoras).

El entender la evolución de los organismos, como un todo, es una tarea compleja y debe involucrar, por una parte, conocer la identidad y saber el arreglo preciso de los elementos genéticos en los genomas: conocer la dinámica de las interacciones de los productos de los genes, conocer la influencia de los factores ambientales en el “hardware” celular (regulación fisiológica) y en el “software” genético (regulación transcripcional). Por otro lado, hace falta entender la influencia de los procesos estocásticos en los procesos bioquímicos y conocer más acerca del grado de robustez y plasticidad evolutiva de los organismos con base en el número y grado de estructuración de sus componentes internos.

Quienes nos dedicamos al estudio de la biología tenemos una oportunidad única y una gran responsabilidad para intentar entender, desde una perspectiva mas integral, los procesos que definen la evolución de la autoorganización material que hizo posible la existencia de los organismos contemporáneos y quizás predecir, e incluso influir en, su comportamiento.

Definiciones de términos empleados

Alelo. Una de las formas alternativas que puede tener un gen.

Mutación. Alteración o cambio en la información genética (ADN o proteína) de un alelo, individuo, especie, y que puede ser heredable.

Migración génica. Transferencia de genes de una población a otra.

Deriva génica. Es una fuerza evolutiva que actúa junto con la selección natural modificando las características de las especies, en el tiempo, mediante el cambio aleatorio de la frecuencia de alelos de una generación a la siguiente.

Estocástico. Concepto matemático que sirve para caracterizar y estudiar todo tipo de fenómenos aleatorios que evolucionan generalmente con el tiempo.

Sitio activo. Es el espacio tridimensional donde ocurre la reacción específica que una proteína realiza (por ejemplo, catálisis en una reacción enzimática).

Tamaño efectivo poblacional. Está definido por los individuos de una población que dejan descendencia, que se reproducen.

Sustituciones sinónimas. Sustituciones a nivel de nucleótidos (codones) que no cambian el aminoácido resultante en la traducción.

Sustituciones no sinónimas. Sustitución de codones que resulta en un cambio del tipo de aminoácido.

Anticuerpo. Son glicoproteínas empleadas por el sistema inmune para identificar y neutralizar elementos extraños tales como bacterias, virus o parásitos.

Antígeno. Sustancia extraña al organismo que desencadena la formación de anticuerpos.

Máxima verosimilitud. Método estadístico basado en distribución de probabilidades; describe el rango de valores de la variable aleatoria, así como la probabilidad de que el valor de la variable aleatoria esté dentro de un subconjunto de dicho rango.

Genómica. Ciencia que estudia el contenido, funcionamiento, origen y evolución de los genomas completos.

Parálogo. Gen producto de la duplicación de otro gen en el mismo organismo.

Dominio proteico. Es un módulo tridimensional de la proteína, el cual puede ser *funcional* si lleva a cabo una función bioquímica, y *estructural* si estabiliza la estructura.

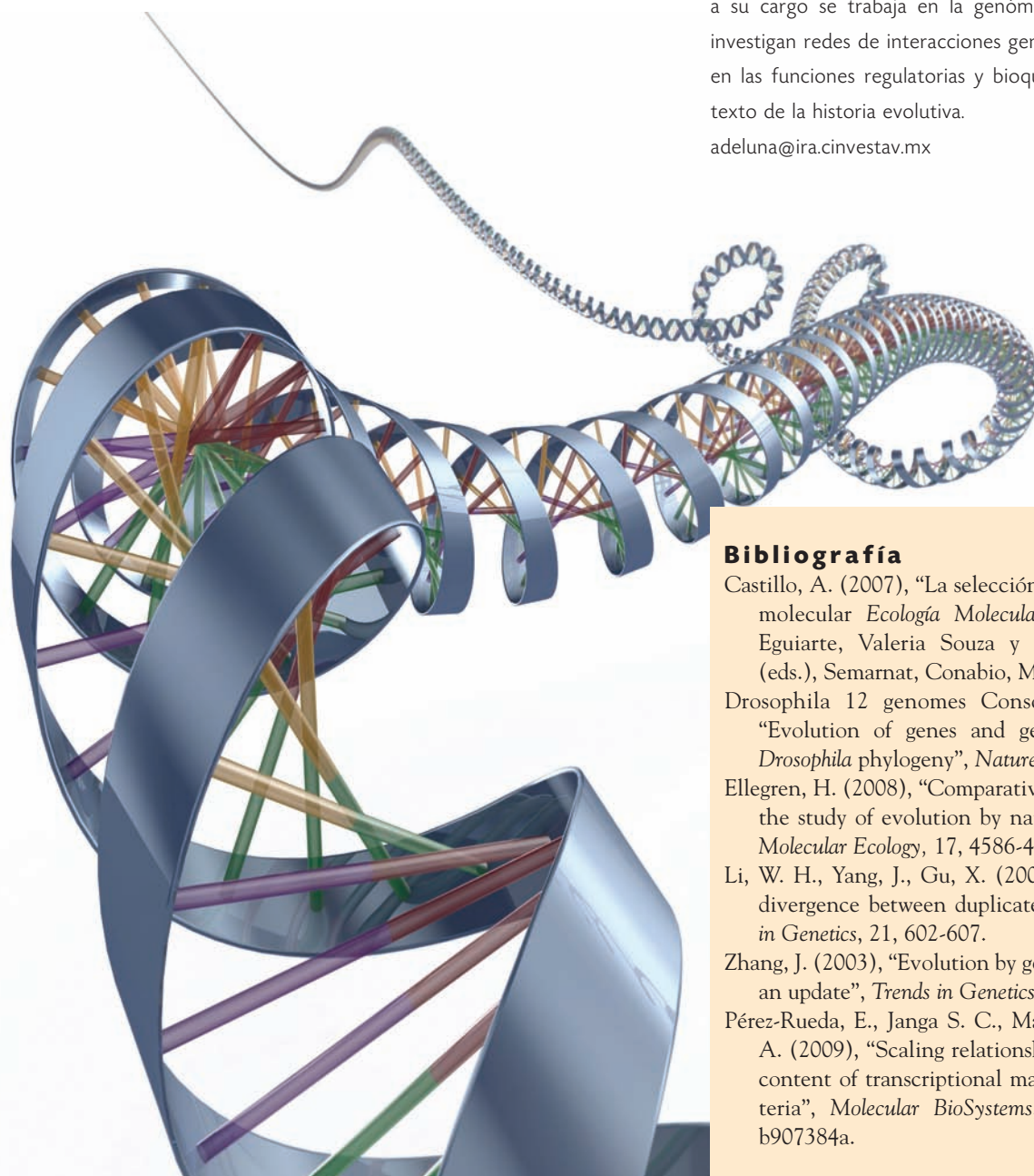
Amanda Castillo Cobián es bióloga egresada de la Facultad de Ciencias de la Universidad Nacional Autónoma de México y doctora en ciencias biomédicas (Instituto de Ecología, UNAM). Ha realizado estancias posdoctorales en el Instituto de Biología y en el de Ecología de la UNAM. Es profesora invitada en el Departamento de Biología del Barnard College, Universidad de Columbia, y ha trabajado para el IBM TJ Watson Research Center en Nueva York. Se interesa en el estudio de la selección y adaptación molecular, la bioinformática y la biología evolutiva. Actualmente realiza una estancia posdoctoral en el Laboratorio de BioSistemas del Centro de Investigación y Estudios Avanzados Irapuato.
amanda.castillo@gmail.com

Agustino Martínez-Antonio es químico-biólogo por la Universidad de Oaxaca (UABJO) y doctor en ciencias bioquímicas por la UNAM (Instituto de Biotecnología). Hizo estudios posdoctorales en el Centro de Ciencias Genómicas (UNAM) y en los Institutos Nacionales de Salud en Francia (INSERM). Actualmente está encargado del Laboratorio de BioSistemas en el Departamento de Ingeniería Genética del Centro de Investigación y Estudios Avanzados Irapuato. Su interés incluye entender los principios que regulan la autoorganización de los sistemas biológicos (sistemas complejos) y su aplicación en el diseño de nuevos sistemas (biología sintética).

amartinez@ira.cinvestav.mx

Alexander de Luna Fors es biólogo egresado de la Universidad de Guadalajara y doctor en ciencias biomédicas por la UNAM (Instituto de Fisiología Celular). Hizo estancias posdoctorales en la Universidad de Harvard (Bauer Center for Genomics Research) y la Harvard Medical School (Department of Systems Biology), donde fue becario del programa Pew Latin American Fellows. Actualmente es investigador titular del Laboratorio Nacional de Genómica para la Biodiversidad (Langebio) del Centro de Investigación y Estudios Avanzados Irapuato. En el Laboratorio de Sistemas Genéticos a su cargo se trabaja en la genómica funcional y se investigan redes de interacciones genéticas, con énfasis en las funciones regulatorias y bioquímicas en el contexto de la historia evolutiva.

adeluna@ira.cinvestav.mx



Bibliografía

- Castillo, A. (2007), "La selección natural a nivel molecular *Ecología Molecular*", en Luis E. Eguiarte, Valeria Souza y Xitlali Aguirre (eds.), Semarnat, Conabio, México, 592pp.
- Drosophila 12 genomes Consortium (2007), "Evolution of genes and genomes on the *Drosophila* phylogeny", *Nature*, 450, 203-218.
- Ellegren, H. (2008), "Comparative genomics and the study of evolution by natural selection", *Molecular Ecology*, 17, 4586-4596.
- Li, W. H., Yang, J., Gu, X. (2005), "Expression divergence between duplicate genes", *Trends in Genetics*, 21, 602-607.
- Zhang, J. (2003), "Evolution by gene duplication: an update", *Trends in Genetics*, 18, 292-297.
- Pérez-Rueda, E., Janga S. C., Martínez-Antonio A. (2009), "Scaling relationship on the gene content of transcriptional machinery in bacteria", *Molecular BioSystems*, DOI:10.1039/b907384a.