

LOS GENOMAS EUCARIOTAS: ASPECTOS GENERALES

Josefina Méndez y Ana M^a González-Tizón
Departamento de Biología Celular y Molecular
Universidad de La Coruña

1. CARACTERISTICAS GENERALES DEL GENOMA EUCARIOTA

Si observamos a los seres vivos, nos encontramos con dos aspectos aparentemente contradictorios, por un lado, la variabilidad entre especies y dentro de una especie, y por otro, la constancia de los caracteres que definen las especies, familias o grupos. Esta paradoja se explica por la presencia, en cada uno de los seres vivos de un material genético que contiene la información necesaria para construir un nuevo individuo.

Históricamente, la extraordinaria diversidad de los seres vivos fue un obstáculo para el descubrimiento de principios unificadores de la Biología en general y de la herencia en particular.

La investigación en genética formal comenzó en el siglo XIX. Se estudió la herencia de los caracteres variables y surgió el concepto abstracto de gen indivisible como la unidad fundamental de la herencia. Hasta la mitad del siglo XX no se comenzó a explorar la identidad química de los genes comenzando el desarrollo de la nueva genética molecular.

Tres clases de moléculas juegan un papel esencial en los procesos genéticos: el ADN, el ARN y las proteínas. Su relación se puede representar como el Dogma de la Biología Molecular, en el que actualmente se describe la dirección de la transferencia de información entre el ADN, ARN y proteínas, señalando la transcripción inversa y la propiedad de autoreplicación del ADN y del ARN.

En un primer período, los grandes descubrimientos se centran en que los genes son ADN, se describe su estructura y como ésta especifica un determinado carácter. La comprensión de la estructura del ADN permitió asegurar las

observaciones de todos los aspectos de la Biología y definió la unidad fundamental para la interpretación de la enorme diversidad de la vida.

En estos avances quedaba implícito el concepto de que los genes son información, que la almacenan y la transmiten de padres a hijos, o lo que es lo mismo, la unidad de herencia que regula la aparición de un carácter particular.

1.1. TAMAÑO DEL GENOMA

En 1920, Winkler, utilizó la palabra genoma para indicar la suma de genes de un organismo; sin embargo, fue preciso llegar hasta 1948 en donde se descubre que la cantidad de ADN por célula de una especie es constante y los gametos estaban constituidos por la mitad.

A esta cantidad constante de ADN de una célula haploide se le denominó tamaño del genoma o valor-C. Hoy en día definimos genoma como toda la información genética contenida en una célula incluyendo a los genes y a otras secuencias de ADN.

Los primeros cálculos estimativos del tamaño del genoma fueron aportados por Mirsky y Ris (1951) midiendo el contenido de ADN celular de una variedad de especies animales y aportando conclusiones fundamentales y sobre todo paradójicas. En la tabla I se muestran ejemplos generales de diversas especies. Datos recogidos del libro *Tratar con Genes* (Berg & Singer, 1994).

TABLA I
Ejemplos de tamaños de genomas

Especies	Tamaño genoma haploide	Nº haploide de cromosomas
<i>Saccharomyces cerevisiae</i>	14	16
<i>Caenorhabditis elegans</i>	80	12
<i>Drosophila melanogaster</i>	170	4
<i>Xenopus laevis</i>	3000	18
<i>Gallus domesticus</i>	1200	39
<i>Mus musculus</i>	3000	18
<i>Homo sapiens</i>	3000	23

Este valor-C, resulta ser de un orden de magnitud mayor que el que podría predecirse a partir del número de genes que codifican para proteínas (Cavalier-Smith, 1985). Por ejemplo se puede estimar que el genoma haploide de la especie humana contiene entre 50.000 a 100.000 genes que codifican para proteínas. Suponiendo que el tamaño medio de las proteínas es de 500 aminoácidos y sabiendo que el genoma haploide humano tiene 2,8 pg de ADN, quiere decir que un 2,5% a un 5% del genoma humano codifica para proteínas. ¿Qué significado genético tiene el 95% restante?.

Desde los primeros estudios se han acumulado gran cantidad de datos cada vez más precisos y de una amplia diversidad de organismos y en ellos, los tamaños de los genomas eucarióticos mostraron que supera con mucho al número de genes precisos para la formación de un organismo y se ha llegado a comprobar que:

- No hay correlación entre valor-C y complejidad genética del organismo.
- Hay una fuerte relación positiva entre valor-C y el tamaño de las células de diferentes especies.
- Existe una fuerte correlación entre tamaño del genoma y volumen nuclear.
- No existe correlación entre el número de genes y el tamaño del genoma.-
- El número cromosómico no refleja la complejidad evolutiva de los organismos.

Las características moleculares de los sistemas genéticos y su organización dentro del ADN eran un misterio y la abundancia de secuencias de ADN muy repetidas, observada en la mayoría de eucariotas, contribuía más a una serie de paradojas.

Era imprescindible un análisis detallado de la anatomía molecular de los genomas eucariotas ya que según se iba conociendo la organización y expresión de la información genética en procariotas, la complejidad de eucariotas impactaba, fue preciso una metodología general que facilitara el análisis molecular de genomas eucarióticos.

Este fue el camino de las nuevas tecnologías o del llamado segundo período que comenzó con el desarrollo de los métodos que se han agrupado bajo la denominación común de técnicas de ADN recombinante o ingeniería genética. Estos métodos revolucionarios permiten el aislamiento y caracterización de genes de cualquier organismo, conocer las diferencias en las estructuras de los genes que dan lugar a variaciones en un carácter o enfermedad.

2. ANATOMÍA MOLECULAR DEL GENOMA EUCARIOTA

La observación de que en todos los genomas de organismos superiores la cantidad de ADN que poseen es muy superior a la necesaria para codificar sus proteínas celulares, hizo pensar que en el genoma existiría una mayoría de ADN no codificante y de función desconocida. Los estudios de la cinética de reasociación molecular han permitido separar distintas fracciones dentro del ADN genómico total, que pueden clasificarse según el número de repeticiones de las secuencias que las componen. Básicamente, existen dos tipos de secuencias: secuencias de ADN de copia única y secuencias de ADN repetitivo. El 50% o más del genoma eucariota está constituido por repeticiones de

segmentos de ADN, mientras que la proporción de segmentos de ADN de copia única es muy escasa.

El ADN de copia única o de bajo número de copias está constituido por la fracción de ADN que renaturaliza lentamente. A partir de los estudios de cinética de renaturalización es muy difícil distinguir entre secuencias presentes en una sola copia de las que lo están en 2 ó 3 copias, por lo que al referirse a ADN de copia única se habla de ADN presente en una o en pocas copias por genoma haploide. Una parte de este ADN de copia única está constituido por genes que se expresan para dar proteínas y parte por secuencias con funciones reguladoras o desconocidas.

2.1. EL ADN REPETITIVO

Hay un extraordinario grado de variación en el tamaño del genoma entre diferentes eucariotas que guarda poca relación con las diferencias en la complejidad del organismo, nivel de ploidía o número de genes que codifican para proteínas. Por ejemplo, el tritón *Triturus cristatus* tiene seis veces más ADN que los humanos. Mucha de esta variación se debe a las secuencias de ADN no codificante repetitivo. En eucariotas el número de unidades de secuencias repetitivas es enormemente variable: desde 2 a millones de pares de bases. El número de copias es diferente según se trate de genomas de especies diferentes.

Una de las teorías que explican el origen de las secuencias repetitivas en un genoma se basa en el proceso genético de amplificación por duplicación de una secuencia. Amplificación en un término que engloba cualquier mecanismo que produzca incrementos a gran escala del número de copias de una secuencia. Uno de los ejemplos más conocidos es el de la amplificación de los genes ADNr en *Xenopus*, el cual es amplificado cerca de 1000 veces durante la oogénesis. La amplificación permite que un oocito acumule aproximadamente 10^{12} ribosomas. El ADN amplificado está en forma de pequeñas moléculas circulares de ADN extracromosómico. En el hombre, existen, aproximadamente, 200 copias del gen principal del ARNr y 2000 copias del gen ARNr 5S.

La amplificación puede causar la generación de copias en tándem o bien de copias dispersas en nuevos loci del genoma. Las secuencias que están relacionadas, dentro de un mismo genoma, podrían tener un origen independiente o podrían originarse a partir de un único segmento de ADN original. La probabilidad de que dos secuencias se originen independientemente es menor cuanto mayor es la similitud entre ellas y el tamaño de la región homóloga. Si una única secuencia "fundadora" es suficiente para el genoma, la acumulación de copias adicionales no ofrece ninguna ventaja adicional. Las copias de ella pueden acumular mutaciones (inserciones y deleciones) que pueden tener efectos diferentes: o bien la copia puede dejar de ser funcional o bien puede adquirir una función nueva. Si estos cambios mutacionales son útiles la selección actuará positivamente sobre ellas; por el contrario, si los cambios no son útiles estas copias son clasificadas de pseudogenes (segmentos genómicos similares estructuralmente a los genes funcionales, pero incapaces de generar

productos génicos funcionales). La amplificación de las secuencias de ADN es la materia prima para que actúe la evolución.

Dentro de las secuencias de ADN repetitivo pueden distinguirse dos grandes categorías: secuencias de ADN moderadamente repetitivo y secuencias de ADN altamente repetitivo.

2.2. EL ADN MODERADAMENTE REPETITIVO

El ADN moderadamente repetitivo está constituido por la fracción de ADN con una velocidad de renaturalización intermedia que incluye una serie muy heterogénea de secuencias repetidas en un moderado número de copias dentro de un genoma (entre 10 y 1000 veces por genoma). En eucariotas, el ADN medianamente repetitivo se encuentra, al menos, en tres clases: disperso (probablemente ADN que no se transcribe, como es el caso de la familia Alu); genes transcritos en muchas copias prácticamente idénticas (ARN ribosómico y genes de las histonas) y genes transcritos en muchas copias que han divergido unas de otras (genes de los anticuerpos, colágeno y genes de las globinas). El término familia génica se refiere a los genes que han surgido por duplicación, con o sin divergencia, a partir de un gen ancestral.

Una gran parte del ADN moderadamente repetitivo disperso (secuencias de ADN repetitivo dispersas entre secuencias de copia única de longitud variable) está representado por los elementos transponibles, que son secuencias capaces de insertar copias de ellas mismas en otras regiones del genoma. En eucariotas, las familias de transposones se integran en dos grupos llamados clase I y clase II. Los elementos móviles de clase I, o retrotransposones, se movilizan a través de ARN, y los de clase II saltan directamente a través del ADN. La mayor parte de las familias conocidas pertenecen al grupo de los retrotransposones. El número de copias de cada familia presente en el genoma varía de unos individuos a otros y de unos tipos de elementos móviles a otros. Las inserciones de elementos transponibles originan una elevada proporción de mutaciones con grandes efectos a nivel de fenotipo. Estas mutaciones que pueden tener importantes efectos a nivel de desarrollo y a nivel evolutivo. La mayor parte de las familias génicas con secuencias de ADN repetitivo dispersas contienen uno o muy pocos genes activos y un gran número de pseudogenes.

En muchos mamíferos, una gran proporción de este ADN está compuesta por muchas copias de una corta secuencia dispersa por el genoma. Por ejemplo, en el hombre, hay aproximadamente 300.000 copias de una secuencia de 300 pares de bases denominada familia Alu. No se conoce con exactitud la función de esta familia, pero posiblemente algunos miembros sean transcritos en ARN 7S pequeños, los cuales están implicados en la secreción de polipéptidos recién sintetizados a través del retículo endoplasmático. También se cree que la secuencia Alu pueda estar implicada en los numerosos orígenes de replicación del ADN a lo largo del cromosoma eucariota o bien en la generación de estructuras secundarias de los ARN mensajeros.

Dentro de los genes que se transcriben en múltiples copias prácticamente idénticas se encuentran los genes cuyo producto final es una molécula de ARN

(las familias multigénicas más repetitivas dentro del genoma), tanto de ARN ribosómico como de ARN transferente. El número y la localización de los genes repetidos se realiza mediante pruebas de hibridación con sondas marcadas. Así, se sabe, por ejemplo, que en el hombre existen más de 1300 copias de los genes ARNt.

Los genes que codifican para los ARNs ribosómicos están presentes en múltiples copias organizadas en tándem dentro del genoma y muestran un elevado contenido en Guanina-Citosina. Por medio de la aplicación de diferentes técnicas moleculares, incluyendo la hibridación *in situ*, se ha demostrado que los genes ADNr se sitúan en las regiones cromosómicas conocidas como NOR (Regiones Organizadoras del Nucleolo). Sin embargo, los genes de las proteínas ribosómicas son genes de bajo número de copias: 1 a 10 copias por genoma de genes funcionales de las proteínas ribosómicas individuales. Los genes de los ARNr 18S, 5'8S y 28S están agrupados en una única unidad de transcripción. El transcrito de mayor longitud es el precursor a partir del cual, por medio de una serie de reacciones intermedias, se generan los tres ARNs. La unidad de transcripción completa contiene dos segmentos de ADN espaciadores, no codificantes, llamados espaciadores transcritos internos (ITS), que separan a las tres regiones codificantes unas de otras. Los ITS difieren muchísimo en secuencia y en organización de una especie a otra. La única característica común de las secuencias ITS entre distintas especies es la presencia de repeticiones directas organizadas en tándem; la secuencia de la unidad de repetición es propia de cada organismo.

Existen, también, espaciadores transcritos externos (ETS), que preceden al primer gen (el de ARNr 18S) y siguen al último (el de ARNr 28S). Ya que el tamaño y la secuencia de los ARNr están muy conservados entre los diferentes eucariotas, las diferencias interespecíficas en los tamaños de las unidades de transcripción son el reflejo directo de los diferentes tamaños de las regiones espaciadoras transcritas. Estas características hacen de los ITS y ETS unos excelentes marcadores moleculares para poder realizar, entre otros, estudios filogenéticos.

Separando las unidades de transcripción de los ADNr de cada una de estas agrupaciones, se encuentran los espaciadores intergénicos (IGS).

Por otra parte, las unidades de transcripción de ADNr aparecen en múltiples copias organizadas en tándem que se repiten, con mucha frecuencia, en varios cromosomas diferentes.

Los genes que codifican para el ARNr 5S no se encuentran ligados a los del 18S, 5'8S y 28S (excepto en los hongos y en los protozoos). Están repetidos muchas veces y organizados en tándem. La región codificante es rica en Guanina-Citosina, y el resto, que es el espaciador no transcrito, es rico en Adenina-Timina. En el caso de *D. melanogaster* existen 160 unidades repetidas agrupadas e ininterrumpidas. En las distintas especies de *Xenopus*, en otros anfibios y en peces, existen 2 familias de genes ARNr de 5S. Una de las familias contiene miles de copias génicas que sólo se expresan durante la oogénesis, y la otra contiene muchas menos copias y se expresa tanto en los oocitos como en las células somáticas a lo largo del proceso de desarrollo. Las unidades de repetición

de 5S de los oocitos son bastante complejas y difieren mucho incluso entre especies relacionadas; las de los ARNr 5S somáticos son sencillas: existe una región espaciadora larga alternando con una unidad de transcripción corta.

Los genes de ARNt muestran una organización desordenada, aparentemente aleatoria: pueden estar agrupados o dispersos en los cromosomas. Por ejemplo, en *Drosophila* se han detectado unos 50 loci diferentes en los que existen varios genes agrupados. Cada uno de estos loci contiene genes que codifican para varios tipos de ARNt.

En el caso de los genes codificadores de histonas las secuencias de los diferentes miembros de la familia multigénica son prácticamente idénticas; las estructuras primarias están enormemente conservadas. Si pensamos en el papel que estas moléculas desempeñan en la estructura y organización del cromosoma eucariota, tal situación no resulta nada sorprendente. Las secuencias de las histonas H1 son las más divergentes, mientras que las de las H3 y H4 son las más conservadas. No obstante, existen algunas variaciones: dentro de una misma especie, las moléculas de histonas que se sintetizan pueden ser distintas (aunque con mínimas variaciones) dependiendo del ciclo celular, del tejido o del estadio de desarrollo. Entre especies diferentes, el número y la organización de los genes que codifican para las histonas son, también, bastante diferentes. Esto confirma el hecho de que la conservación de secuencias codificadoras no está necesariamente asociada a la conservación del número de copias o de la organización del genoma. Los genes de las histonas tienden a estar ligados, si bien se transcriben independientemente unos de otros. Además, el orden de ligamiento no es siempre el mismo de una especie a otra, e incluso hay variaciones dentro de una especie. En aves y mamíferos los genes codificadores de histonas están dispersos por el genoma. En los erizos de mar los genes de histonas pueden ser "tempranos" y "tardíos": antes de la blastulación se expresan unos y después de ella, los otros. En *Drosophila*, durante la embriogénesis, existe una modulación en la expresión de los genes de las histonas: de más alto a más bajo. En *Xenopus*, existen varios grupos ligados de los cinco genes de histonas, con un ordenamiento de los genes diferente. Al menos dos de estos grupos poseen genes de H1 diferentes. En general, los animales de sangre caliente tienen relativamente pocos genes de histonas.

En cuanto a los genes que codifican para polipéptidos, suelen estar representados en el genoma sólo una vez, existiendo un único gen funcional y otros genes homólogos con la secuencia de dicho gen. De esta forma, los genes de copia única suelen formar parte, con bastante frecuencia, de familias génicas, cuyos miembros pueden incluir a genes que codifican para proteínas muy similares, o que se asocien con señales reguladoras diferentes, lo cual permite que los distintos miembros de la familia se expresen en otros tejidos o en otros momentos del desarrollo.

Una extensión al concepto de familia génica es lo que se denomina superfamilia génica, como por ejemplo la superfamilia de las inmunoglobulinas, la cual comprende muchas familias de genes con funciones diferentes pero interrelacionadas unas con otras. En cuanto a la secuencia, los miembros de las superfamilias son más similares entre sí que los de las familias génicas.

Los microsátélites y minisátélites representan secuencias de ADN moderadamente repetitivo que normalmente se repiten en tándem (familias multigénicas en tándem). Genéricamente también se les denomina VNTRs (variable number of tandem repeats). Son secuencias cortas de nucleótidos -de 2 a 5 pares de bases los microsátélites y de, aproximadamente, 15 pares de bases los minisátélites-. Los microsátélites suelen ser repeticiones de Adenina- Timina, y en mucha menor proporción de Guanina-Citosina. En el caso del genoma humano existen, al menos, 30.000 loci de secuencias microsátélites localizados en la eucromatina. Los minisátélites comprenden secuencias altamente hipervariables. Sin embargo, aunque pueden variar en tamaño, presentan un core secuencial común a todas las familias que lo componen. Esta secuencia core está constituida en su mayor parte por repeticiones Guanina-Citosina. En los cromosomas suelen aparecer localizadas en los telómeros y en regiones eucromatínicas próximas a ellos.

En función del número de copias en que estas secuencias aparecen en un genoma, pueden clasificarse como secuencias moderadamente repetitivas o como secuencias altamente repetitivas (ADN satélite). En humanos, por ejemplo, minisátélites y microsátélites se incluyen dentro del ADN altamente repetitivo, ya que aparecen entre $2-3 \times 10^4$ copias (microsatélite) y $3 \times 10^6-10^7$ copias (minisatélite). (También la familia Alu, en humanos, se incluye dentro del ADN satélite). En otros genomas eucariotas estos tipos de secuencias se consideran moderadamente repetidas. Ambos tipos de secuencias presentan herencia mendeliana simple y son extraordinariamente polimórficas (variabilidad de la secuencia en los miembros de una misma familia). Constituyen, por tanto, familias polimórficas. Cuando esto ocurre, si la diferencia es en pocas posiciones, es conveniente identificar a los miembros de la familia mediante una secuencia consenso que representa la base o nucleótido más frecuente de cada posición.

En humanos, los minisátélites poseen elevadas tasas de mutación, mientras que los microsátélites muestran una tasa mutacional media-baja. Estas características hacen que dichas secuencias sean enormemente utilizadas en Medicina legal.

2.3. EL ADN SATÉLITE

Las secuencias de ADN altamente repetitivo constan, por lo general, de secuencias cortas de ADN repetidas muchas miles de veces en el genoma, y que se encuentran dispersas y dispuestas en tándem formando grandes clusters (ADN satélite) a lo largo del ADN. En el genoma eucariota estas secuencias constituyen una parte importante del mismo.

Existe una gran variabilidad para el ADN satélite, variabilidad que se observa en la longitud de las unidades que lo componen, en el número de copias de dichas unidades, en su localización y en las secuencias en sí mismas. El tamaño de las secuencias que se repiten (monómeros) en un ADN satélite oscila entre unos pocos pares de bases (caso de *Drosophila*) a varios miles (cetáceos, bovinos). Sin embargo, el tamaño medio de los monómeros de ADN satélite

suele oscilar entre 100 y 300 pares de bases. La proporción varía según los distintos taxones. Así, por ejemplo, en la levadura el ADN altamente repetitivo constituye el 20% del genoma, en mamíferos es del orden del 60% y en plantas y anfibios está representado por más del 80%.

El ADN satélite está constituido por familias de secuencias cortas que se repiten cientos de miles de veces, e incluso millones de veces, formando clusters o loci cromosómicos en los genomas eucariotas. Las secuencias de ADN satélite dispuestas en tándem se localizan generalmente en las regiones centromérica y telomérica de los cromosomas (zonas de heterocromatina constitutiva), mientras que las secuencias dispersas de ADN altamente repetitivo se localizan en intrones (secuencias intercalares), flanqueando otros genes, formando parte de genes no codificantes o pseudogenes o de genes en copias múltiples. Las unidades de repetición de los ADNs satélites pueden ser similares en longitud a las de los microsátélites y minisátélites (5-10 pares de bases) o mucho más largas (100 pares de bases), pero se diferencian de estas en que se organizan como grandes clusters de hasta 100 megabases en las regiones heterocromáticas de los cromosomas. Aparentemente no son tan variables en tamaño dentro de las poblaciones como los micro- y los minisátélites.

En lo que se refiere a la variabilidad a nivel de secuencia hay que distinguir entre variabilidad intraespecífica y divergencia interespecífica. La variabilidad entre las secuencias de una familia de ADN satélite dentro de una especie, es como mucho del 15%. Sin embargo las diferencias pueden ser mayores, como ocurre con la familia del satélite alfa centromérico humano donde pueden ser superiores al 30%. Por lo general, las variaciones a nivel intraespecífico son menores del 15% y, a veces, mínimas. Si bien cada familia de ADN satélite tiene una unidad monomérica perfectamente definida, cuando se analizan sus secuencias se pueden detectar periodicidades internas, reflejo de la evolución de este ADN satélite. Estas periodicidades consisten en repeticiones directas o inversas, por ejemplo, el ADN satélite menor, centromérico, de *Mus musculus*, está formado por monómeros de 234 pares de bases. Sin embargo, cuando se analiza su secuencia se puede observar que la longitud actual de las unidades monoméricas se ha originado a partir de ciclos alternantes de mutación y amplificación de un nonanucleótido.

Considerando el tamaño de la unidad de repetición, las secuencias dispersas de ADN satélite se clasifican en: SINES, secuencias repetitivas dispersas cortas y LINES, secuencias repetitivas dispersas largas. SINES y LINES fueron descritas originalmente como secuencias que evolucionaron por mecanismos de transposición, intercambio desigual y conversión génica. Generalmente se localizan flanqueando genes, formando parte de intrones, insertadas en las repeticiones del ADN satélite, en lugares intragénicos no codificantes y de secuencias que no se traducen. Las SINES parecen ser pseudogenes procesados derivados de genes que codifican para ARN pequeños, incluyendo los transferentes. La secuencia SINE más conocida es una secuencia Alu de los monos del Viejo Mundo. Dentro de una especie las secuencias Alu pueden divergir entre sí hasta un 15%. Comparando los monos del Viejo Mundo entre sí la diversidad en las secuencias Alu es casi igual a la detectada a nivel de una única especie. En humanos, la familia Alu presenta un alto contenido en

Guanina-Citosina, mientras que la secuencia LINE más estudiada es la L1Hs, presente en, al menos, 100.000 copias, y dispersas en muchos lugares cromosómicos.

SINES y LINES representan el 10% o más de los genomas de mamíferos. No se conoce la función de estos tipos de secuencias en el genoma, pero se cree que estas secuencias o sus transcritos están implicados en la regulación de la expresión génica o bien que no tienen función alguna dentro del genoma. Posiblemente los pseudogenes y genes procesados sean secuencias de este tipo. Este ADN sería el llamado ADN egoísta: su abundancia en el genoma podría ser la consecuencia directa de su capacidad de multiplicarse y dispersarse. No obstante estas características hacen de él un agente potencial para producir mutaciones por inserción en secuencias reguladoras o codificadoras importantes.

2.4. EL ADN DE LOS TELÓMEROS Y CENTRÓMEROS

La mayor concentración de ADN satélite organizado en tándem aparece en los centrómeros y telómeros de los cromosomas animales y vegetales. Debido a su elevado número de copias, se encuentran entre las secuencias que se reasocian muy rápidamente que, incluso, pueden aparecer como bandas individuales tras la centrifugación en gradiente de densidad del ADN. Los satélites pueden diferenciarse unos de otros por el tamaño y por la secuencia nucleotídica de la unidad de repetición. Su estructura es la repetición de bases Adenina-Timina y su tamaño de 2 a varios miles de pares de bases.

2.4.1. TELÓMEROS

Los telómeros son entidades moleculares necesarias para garantizar la replicación y la estabilidad de los extremos cromosómicos. Marcan la terminación lineal del cromosoma y tienen diversas funciones específicas: protegen a los extremos cromosómicos de su tendencia a adherirse y de la degradación por las exonucleasas (durante el proceso de replicación de la molécula lineal de ADN), y permiten que dichos extremos sean perfectamente replicados. El ADN telomérico consiste en repeticiones de secuencias de 5 a 8 pares de bases dispuestas en tándem. En vertebrados la secuencia telomérica es TTAGGG, y está repetida varios cientos de veces. Por pruebas de hibridación *in situ*, esta secuencia no sólo ha sido detectada en los telómeros si no también en alguna otra región cromosómica (fundamentalmente pericentromérica, -intersticial-). Es una secuencia altamente conservada, que además de encontrarse en todos los vertebrados estudiados, aparece también en los tripanosomas unicelulares. Se han hallado secuencias similares en otros eucariotas como, por ejemplo, los ciliados y las levaduras. En numerosas especies de vertebrados, la distribución de las secuencias TTAGGG en regiones no teloméricas es más frecuente que en los telómeros. Se desconoce el origen de las secuencias teloméricas en regiones cromosómicas intersticiales. Algunos autores han sugerido que la mayor parte de estas secuencias teloméricas intersticiales sean originadas por la fijación de alteraciones cromosómicas durante la evolución cariotípica. Ya que las fusiones y fisiones cromosómicas en las zonas de las

unidades de repetición TTAGGG, son los mecanismos más comunes mediante los que tienen lugar los cambios cariotípicos, las regiones pericéntricas son los lugares lógicos para que “surjan” nuevos telómeros.

La conservación de una secuencia repetida indica *a priori* que dicha secuencia pueda tener una función. Por ejemplo, parece estar bastante claro que la conservación de la secuencia telomérica (TTAGGG)_n, en todos los vertebrados, sugiere que esa secuencia está implicada en la función telomérica (estabilidad cromosómica y terminación completa de la replicación sin pérdidas de nucleótidos en el proceso).

2.4.2. CENTRÓMEROS

Cada especie contiene un conjunto distintivo de satélites centroméricos, incluso si se comparan especies muy próximas dentro del mismo género. Existen ADN_s satélites centroméricos cuyas secuencias están conservadas en distintos grupos evolutivos como la secuencia (GGAAT)_n del satélite III de humanos o el satélite dodeca de *Drosophila melanogaster*, aunque se desconoce el papel que estas secuencias podrían desempeñar en la función centromérica. En el caso de *D. melanogaster*, existen cuatro satélites (16% del genoma) centroméricos diferentes. Cada uno de ellos ocupa locus específicos. En bovinos existen ocho satélites centroméricos diferentes (23% del genoma) que se organizan en secuencias moderadamente y altamente repetitivas. Son enormemente divergentes aunque cinco de ellos comparten subsecuencias. Las unidades de repetición tienen generalmente una longitud superior a 1 kpb. Los primates contienen varios tipos de satélites, formados por repeticiones tanto de unidades largas como de unidades cortas.

En humanos existen al menos dos tipos de satélites: el ADN alfoide y el ADN satélite clásico. El ADN alfoide se localiza en todas las regiones centroméricas de los cromosomas y está compuesto por unidades de repetición de unos 170 pares de bases con una homología del 60% al 95%. Estos monómeros de 170 pares de bases se organizan en unidades de repetición de orden superior que están presentes en múltiples copias dentro de las construcciones primarias de un pequeño grupo de cromosomas. Están, además, altamente metilados. El ADN satélite clásico se localiza en las regiones de heterocromatina constitutiva de los cromosomas 1, 9, 16 e Y; están también altamente metilados y se agrupan en tres clases principales: satélite 1 (característico del cromosoma Y y compuesto de una unidad de repetición de 247 kilobases), Y satélites 2 y 3 (compuestos por variaciones de la secuencia ATTC).

2.5. SIGNIFICADO BIOLÓGICO DEL ADN REPETITIVO

Las secuencias de ADN repetitivo constituyen una gran proporción del ADN genómico de eucariotas. Los ADN_s satélites son, generalmente, específicos de especie, pero no está nada claro para qué sirven, cuál es su función en el genoma. Su usual inactividad transcripcional, la ausencia de función

codificadora y la gran divergencia en la secuencia nucleotídica y en el número de copias entre especies (incluso entre especies muy emparentadas), son las razones por las que ha sido considerado como ADN egoísta, parasitario, trivial o basura. No obstante se han propuesto algunas hipótesis para explicar su función en el genoma, así como su localización en los centrómeros y telómeros de los cromosomas eucariotas. Se le han atribuido funciones relacionadas con la organización cromosómica, el metabolismo celular, el apareamiento cromosómico y la evolución cariotípica y especiación; sin embargo, ninguna de estas supuestas funciones han sido demostradas. Lo que sí parece estar claro es la de su papel modificador del entrecruzamiento meiótico. Recientemente se ha comprobado que el ADN satélite humano (localizado en las regiones centroméricas de todos los cromosomas humanos) es una parte funcional del centrómero necesaria para regular la segregación de los cromosomas durante la mitosis.

La hipótesis del ADN egoísta propone que estas secuencias se mantienen dentro del genoma por su habilidad para replicarse dentro de él. El comportamiento de dichas secuencias puede originar mutaciones (causantes de enfermedades genéticas) y conferir una pérdida significativa de fitness a los organismos. Se ha propuesto, con bastante frecuencia, que las secuencias repetitivas son funcionalmente importantes para el organismo hospedador, o son mantenidas porque sus actividades mutagénicas toman parte en el proceso evolutivo de la población. Otros autores sugieren que estas secuencias no reportan beneficio alguno para el genoma hospedador; dentro de estas secuencias se incluirían las secuencias transponibles y las altamente repetidas dispersas. Sin embargo, quizás, lo más factible sea el papel que el ADN satélite pueda jugar a nivel de organización y comportamiento cromosómicos, especialmente en los efectos físico-mecánicos que pueda tener en el ciclo celular. Se asume que el ADN satélite puede jugar un papel importante en la condensación de la cromatina en la metafase, así como en la condensación de la heterocromatina debida a las estructuras secundaria y terciaria.

El intercambio desigual en combinación con la deriva genética y la selección pueden tener un gran efecto en la acumulación en el genoma de secuencias repetitivas organizadas en tándem. En concreto, es de esperar que la acumulación de secuencias altamente repetitivas tenga lugar solo en regiones con una recombinación muy baja y una fuerza selectiva débil. Es el caso de centrómeros y telómeros (regiones cromosómicas donde la recombinación es muy baja y donde suelen localizarse secuencias altamente repetidas). Parece lógico pensar que la supresión de la recombinación origine la acumulación de secuencias repetidas; además, las secuencias que inherentemente no tengan tendencia a recombinar serán más abundantes. Sin embargo, los minisatélites, que constan de secuencias menos repetitivas pero mucho más variables en número de copias que los satélites, aparecen en las regiones eucromatínicas de los cromosomas, donde el intercambio de material genético (crossing over) es más frecuente. Las unidades de repetición tienden a ser eliminadas por deriva genética cuando no existan "fuerzas genéticas" que aumenten el tamaño de las unidades de repetición.

Otra característica diferencial entre los ADNs de los satélites y de los minisatélites es la mayor tendencia a encontrar unidades de repetición multiméricas en los satélites. La formación de unidades de repetición de orden superior (más de 200 kilobases) se ha detectado con mucha frecuencia. El ejemplo más conocido es el de la familia de los ADN satélite alfa, que consiste en una unidad básica de repetición de 171 pares de bases que se combina en varias estructuras de orden superior, que van desde dímeros hasta hexadécámeros. Los estudios de simulaciones por computador incorporando intercambio desigual y selección, sugieren que las unidades de repetición de orden superior se forman en regiones de baja recombinación y fuerza selectiva débil, y que los intercambios entre unidades de repetición dependen más del tamaño de las secuencias (cortas) que de la homología de las mismas.

Gran parte de los ADNs satélites están restringidos a un grupo o a unos pocos grupos de especies relacionadas, lo que implica que pueden tener un origen relativamente reciente. Los ADN satélites muestran frecuentemente cambios rápidos entre las especies, incluso entre especies emparentadas. Pueden perderse o aparecer nuevas familias de ADN satélite en un período de tiempo evolutivo relativamente corto. Es decir, el ADN satélite cambia, aparece y desaparece a un ritmo mucho más elevado que cualquier otra parte del genoma. Así, por ejemplo, existen ADNs satélites que sólo aparecen en una especie, teniendo por tanto un origen muy reciente, como es el caso de los satélites Cgo A y Cgo B de una especie de primates, o ADNs satélites específicos de un grupo de especies relacionadas, como ocurre con el género *Equus*. Otros ADNs satélites muestran un grado de conservación mayor, como es el ADN satélite de la familia *Cebidae*. El ADN satélite alfoide centromérico de primates está ampliamente extendido entre las especies de un orden completo, pudiendo llegar a tener una antigüedad de 40-50 millones de años.

Finalmente, debemos reseñar que la utilidad del ADN satélite en el análisis de las relaciones filogenéticas entre especies emparentadas es muy cuestionado por diferentes autores, debido a la rápida evolución que este tipo de ADN presenta. Sin embargo, una familia de ADN satélite resulta el mejor marcador filogenético cuando se trata de agrupar las especies que la comparten en un clado (grupo monofilético). Así, el simple análisis de presencia/ausencia de un ADN satélite en distintas especies ha resuelto, por ejemplo, sobre la monofilia de los cetáceos o de los pinnípedos; ha permitido discriminar entre los distintos linajes dentro de la familia *Equidae*, entre especies de la familia *Cervidae*, etc. Además, cuando una especie o grupos de especies comparten distintas familias de ADN satélite, se puede aumentar el poder de resolución en el estudio de las relaciones de parentesco entre las mismas, ya que alguna de las familias puede ser común a un grupo completo, mientras que otras son únicas para un linaje determinado. En este sentido, por ejemplo, se ha utilizado la aparición de distintas familias *SINES* a lo largo de los diferentes estadios de la evolución de los salmónidos. Un análisis similar con *LINES* se ha utilizado con éxito en la filogenia de los pinnípedos.

Pueden conocerse aspectos sobre la divergencia genética entre los ADNs satélites de distintas especies, y utilizar estos datos en la inferencia filogenética. Debido a la rápida evolución del ADN satélite, las distintas especies que

comparten una familia del mismo pueden exhibir patrones de restricción específicos. Estos patrones pueden ser utilizados para establecer agrupamientos entre especies de acuerdo a criterios de proximidad entre las mismas. Este método se ha empleado en ciervos, lagartos, aves o murciélagos, por ejemplo.

Los patrones de divergencia entre los ADN_s satélites de un grupo de especies también se ha analizado mediante la comparación de las secuencias de cada uno de ellos: o bien comparando las secuencias consenso del ADN satélite en cada especie o bien un número determinado de secuencias monómero de cada especie (variabilidad intraespecífica). Sin embargo, los datos de secuenciación del ADN hay que manejarlos con precaución, teniendo en cuenta los mecanismos que intervienen en la evolución del ADN satélite. Se puede inferir una filogenia a partir de secuencias consenso, pero puede ocurrir que la divergencia entre secuencias consenso sea menor que la variabilidad intraespecífica. Esto se explica si se tiene en cuenta que los lugares variables en los monómeros de una especie pueden ser sitios en estado de transición en el proceso de fijación de los mismos, mientras que los lugares totalmente homogenizados y fijados en cada especie se verán reflejados en esa secuencia consenso y analizados como lugares divergentes entre las dos especies que se comparan. Cuando se comparan distintas unidades monoméricas de cada especie, la existencia de lugares variables (estados de transición) puede llevar a la paradoja de que secuencias monoméricas de dos especies distintas presenten mayor identidad que cuando se comparan cada una con unidades de la misma especie. Dado que la vida media de un ADN satélite supera la vida media de una especie, puede ocurrir que la variabilidad exista ya en el ancestro de las especies que se analizan. En este caso, si el tiempo que hace que divergieron las especies que se comparan es relativamente reciente, o la homogenización y fijación de las variantes es un proceso relativamente lento, puede darse la paradoja mencionada. Así, las filogenias basadas en ADN satélite han de considerar no sólo las tasas de mutación, sino también las de homogenización y fijación, distinguiendo entre lugares en estado de transición y lugares totalmente homogenizados.

3. EL GENOMA EXTRANUCLEAR

En eucariotas, el particular modo de herencia de algunos genes ha revelado su presencia fuera del núcleo. La existencia del genoma extranuclear se reconoce genéticamente por la existencia de transmisión fenotípica uniparental. A nivel celular, mediante técnicas genéticas y bioquímicas han demostrado que este genoma es ADN y se sitúa en mitocondrias (ADN_{mt}) o en cloroplastos (ADN_{cp}), constituyendo un único cromosoma.

Los genes extranucleares o citoplásmicos son bastante menos que los genes cromosómicos normales pero desempeñan papeles altamente especializados en el control de los fenotipos eucarióticos.

3.1. GENOMA MITOCONDRIAL

Estos tipos de análisis han sido posibles gracias a los estudios del genoma mitocondrial de la levadura *Saccharomyces cerevisiae* mostrándose como un ejemplo del poder de resolución del análisis genético cuando se combina con las modernas técnicas de biología molecular.

El mapa de levaduras se ha analizado con detalle, se ha establecido su secuencia la cual contiene algunas sorpresas. La más llamativa es la presencia de intrones en algunos genes, ya que son raros en el genoma nuclear. Otra sorpresa es la presencia de largas fases de lecturas desconocidas (URF). Son secuencias con codones de iniciación correctos y que no están interrumpidas por codones de terminación. Estos "genes en busca de función" son en la actualidad objeto de intensas investigaciones (podrían cifrar proteínas importantes para la eliminación de los propios intrones en el ARN).

La visión general del genoma mitocondrial pone de manifiesto dos funciones principales:

- 1.- que cifra proteínas implicadas en la cadena de transporte de electrones y otras proteínas,
- 2.- y todos los ARNts y ARNr necesarios para la síntesis de las proteínas mitocondriales.

Lo sorprendente es que el resto de los componentes necesarios para estas funciones sean cifrados por los genes nucleares.

El mecanismo de esta forma tan peculiar de trabajo entre ADN mitocondrial y celular se desconoce. Se indica que quizás en estos orgánulos ha debido existir una transposición evolutiva de información.

Sus patrones de organización y herencia bastante distintos de los cromosomas nucleares.

3.2. GENOMA DE CLOROPLASTOS

Gran parte de lo que hoy sabemos sobre la organización del ADN cloroplástico proviene de los estudios moleculares. De igual forma que el ADNmt, el ADNcp coopera con el ADN nuclear aportando subunidades para formar proteínas funcionales usadas dentro del orgánulo. Recientemente se ha secuenciado este genoma en *Marchantia* cuyo contenido es de 121kb, en la que se han detectado 136 genes, 4 de ARNr, 31 de ARNt y unos 90 de proteínas. De estos últimos, 20 determinan funciones fotosintéticas y de transporte de electrones. La presencia de genes se deduce a partir de las fases de lectura abiertas (ORF), que son largas secuencias que comienzan con un codón de iniciación y que no están interrumpidas por codones de terminación, salvo al final. Además se ha utilizado técnicas de hibridación utilizando sondas de genes conocidos de otros organismos (ARNt y ARNr).

Los genes relacionados con las funciones de traducción ocupan casi la mitad del genoma del cloroplasto, y entre ellos están los de las proteínas y RNA necesarios para que se produzca la traducción en los orgánulos.

3.3. PLÁSMIDOS EXTRAGENÓMICOS EN EUCARIOTAS

La presencia de plásmidos es algo común en bacterias. En eucariotas también se pueden encontrar, aunque son poco frecuentes y no suelen afectar al fenotipo. Su patrón de herencia es no mendeliana, similar a la de los orgánulos. El más conocido es el 2 n (círculo de 2 n) de levaduras cuyo papel en biología molecular ha sido espectacular porque se ha manipulado y se ha convertido en un vehículo genético para transformar células.

Sin embargo, la mayoría de plásmidos eucarióticos son mitocondriales. Un plásmido interesante y bien estudiado está asociado a la esterilidad masculina citoplásmica del maíz.

La existencia y posición evolutiva de los plásmidos eucarióticos es todavía un misterio, pero el estudio de su estructura y comportamiento proporcionará sin duda claves importantes sobre las propiedades fundamentales del material genético.

4. REFERENCIAS

- P. Berg y M. Singer (1994) *Tratar con genes. El lenguaje de la herencia*. Ed. Omega, Barcelona.
- Cavalier-Smith, T. (1985). Cell volume and the evolution of eukaryote genome size. En: *The evolution of genome size*. (T. Cavalier-Smith, ed.). John Wiley and Sons Ltd., London. 105-184.
- Mirsky, E. y Ris, H. (1951). The DNA content of animal cells and its evolutionary significance. *J. Gen. Physiol.* 34: 451-462.

4.1. LECTURAS RECOMENDADAS

ARTICULOS:

- Blackburn, E.H. (1991). Structure and function of DNA of telomeres. *Nature* 321: 209-213.
- Brutlag, D.L. (1980). Molecular arrangement and evolution of heterochromatic DNA. *Ann. Rev. Genet.* 14: 121-144.
- Carracedo, A. (1996). *La variabilidad genética de los micro y minisatélites y su aplicación en medicina legal*. En: *La genética molecular en el diagnóstico de*

- las patologías humanas: estrategias y tecnologías. (Colección: Cursos, Congresos y Simposios, vol. 21). Ed. Universidad de La Coruña. pp. 63-73.
- Cavalier-Smith, T. (1980). How selfish is DNA?. *Nature* 285: 617-618.
- Cavalier-Smith, T. (1985). Cell volume and the evolution of eukaryote genome size. En: *The evolution of genome size*. (T. Cavalier-Smith, ed.). John Wiley and Sons Ltd., London. 105-184.
- Charlesworth, B.; Langley, C.H. y Stephan, W. (1986). The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* 112: 947-962.
- Doolittle, W.F. (1982). Selfish DNA after fourteen months. En: *Genome evolution* (G. Dover y R.B. Flavell, eds.). Academic Press INC., London, New York. pp. 3-28.
- Dover, G. (1978). DNA conservation and speciation: adaptative or accidental? *Nature* 272: 123-124.
- Dover, G. (1993). Evolution of genetic redundancy for advanced players. *Curr. Opin. Genet. Dev.* 3: 902-910.
- Grady, D.L.; Ratliff, R.L.; Robinson, D.L.; McCanlies, E.C.; Meyne, J. y Moyzis, R.K. (1992). Highly conserved repetitive DNA sequences are present at human centromeres. *Proc. Natl. Acad. Sci. USA* 89: 1695-1699.
- John, B. (1988). The biology of heterochromatin. En: *Heterochromatin. Molecular and structural aspects*. R.S. Verma (ed.). Cambridge University Press. pp. 1-147.
- MacGregor, H.C. y Sessions, S.K. (1986). The biological significance of variation in satellite DNA and heterochromatin in newts of the genus *Triturus*: an evolutionary perspective. *Phil. Trans. R. Soc. Lond. B.* 312: 243-259.
- Meyne, J., Ratliff, R.L. y Moyzis, R.K. (1989). Conservation of the human telomere sequence (TTAGGG)_n telomeric sequence in vertebrate chromosomes. *Chromosoma* 99: 3-10.
- Miklos, G.L. (1985). Localized highly repetitive DNA sequences in vertebrate and invertebrate genomes. En: *Molecular Evolutionary Genetics* (J.R. McIntyre, ed.). Plenum Press, New York. pp. 241-321.
- Singer, M.F. (1982). Highly repetitive sequences in mammalian genomes. *Int. Rev. Cytol.* 76: 67-112.

LIBROS:

- Berg, P. & M. Singer, (1994). *Tratar con genes. El lenguaje de la herencia*. Ed. Omega, Barcelona.
- Bradbury, E.M., McJean, N. y Matthews, H. R. (1981). *DNA chromatin and chromosomes*. Blackwell Sci. Pn. Oxford.
- Griffiths, A.J.F.; J. H. Miller, D.T. Suzuki, R.C. Lewontin y W.M. Gelbart. (1995). *Genética* (5ª ed.). Ed. Interamericana. McGraw-Hill, Madrid.
- Lewin. B. (1989). *Genes*. B. Lewin. Ed. Reverté, S.A., Barcelona.
- Strachan, T. (1994). *The human genome*. Ed. A.P. Read y T. Brown. Bios Scientific Publishers.
- Tamarin. R.H. (1996). *Principios de Genética*. R.H. Ed. Reverté, S.A., Barcelona.
- Wagner, R.; M.P. Maguire & R.L. Stallings.(1993). *Chromosomes*. R. Ed. Wiley-Liss, New York.