

EL GENOMA HUMANO



EL GENOMA HUMANO

Jorge Almarza

VI Escuela Venezolana
para la Enseñanza de la **Química**
Mérida, del 05 al 10 de Diciembre de 2004



Universidad de Los Andes
Facultad de Ciencias
Departamento de Química

VI ESCUELA VENEZOLANA PARA LA ENSEÑANZA DE LA QUÍMICA

Edición 2004

El libro ***El Genoma Humano*** fue escrito especialmente como material de apoyo de uno de los cursos ofrecidos en la *VI Escuela Venezolana para la Enseñanza de la Química*. La *Escuela* es un programa organizado por CELCIEC-ULA, diseñada en base a Cursos de Actualización dirigidos a los docentes de Química de la Educación Básica, Media y Diversificada.

Evaluación de la edición:

Bernardo Fontal, Ricardo Contreras

Comité organizador del VI Encuentro con la Química:

Bernardo Fontal, Fernando Bellandi,
Marisela Reyes, Ricardo Contreras

Autor:

Jorge Almarza

E-mail:

jaimarza@ula.ve

Portada:

Yanelly Gavidia

Diseño y diagramación:

Smart Service C.A.

Se autoriza la reproducción parcial y total de esta obra, únicamente para fines de enseñanza, respetando los créditos del VI Escuela Venezolana para la Enseñanza de la Química y de los autores.

Derechos reservados © 2004, Universidad de Los Andes, Facultad de Ciencias, Departamento de Química, Laboratorio de Organometálicos La Hechicera, Mérida 5101, Venezuela. Tlf.: +58 274 2401380, Fax: +58 274 2401286, E-mail: escueladequimica@hotmail.com

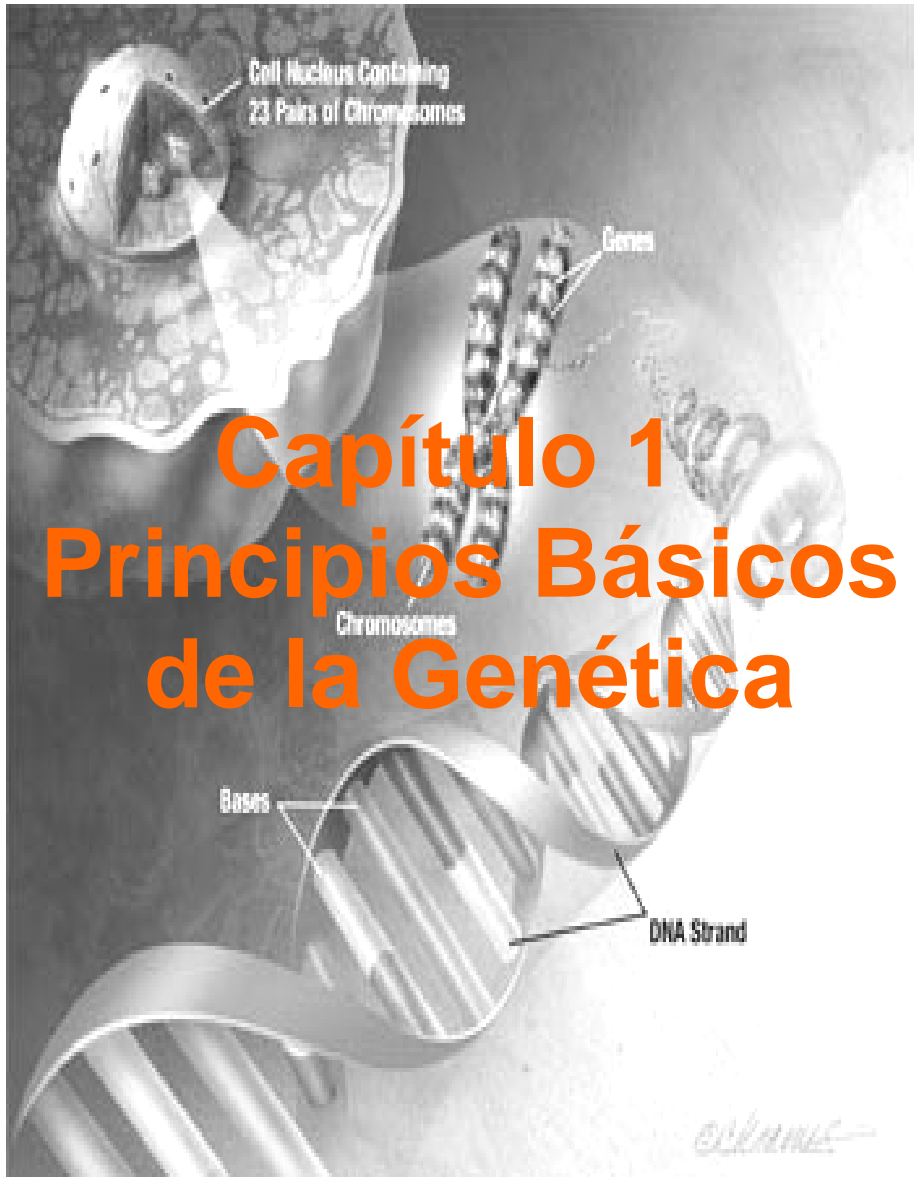
Hecho en Venezuela
Depósito legal:
LF23720045403211

TABLA DE CONTENIDO

Introducción

- El DNA como un polímero lineal
- El modelo de Watson y Crick
- Estabilización del DNA
- El RNA (descripción)
- Tipos de RNA
- Reactividad del RNA
- Procesamiento del RNA
- Intrones y exones
- Enzimas de restricción
- Estudios genómicos
- Proyectos genomas
- Genómica estructural
- Genómica funcional
- El genoma humano
- Secuenciamiento del genoma
- Generación de secuencias del genoma
- Ensamblaje de las secuencias genómicas
- RNAs no codificantes
- Propiedades de los genes conocidos
- Implicaciones éticas del proyecto genoma humano

Referencias



INTRODUCCIÓN

Entre las numerosas propiedades de los organismos vivos, hay una que es esencial para la continuación de la vida: un organismo debe ser capaz de replicarse. Para hacerlo, un organismo debe poseer una descripción completa de sí mismo. Esta descripción es una forma de información, un "conjunto de instrucciones" que especifican cada paso necesario para que la célula construya una réplica idéntica y como cada generación engendra a la siguiente, los descendientes han de recibir una copia del conjunto de instrucciones para que, a su vez, puedan replicarse. En una célula la información necesaria para replicarse se encuentra en el material genético como una molécula llamada *ácido desoxirribonucleico ADN* (o *desoxirribonucleic acid DNA*).

La información sólo es útil si existe un mecanismo para expresarla. En los sistemas biológicos la información contenida en el DNA se copia en una molécula relacionada llamada *ácido ribonucleico ARN* (o *ribonucleic acid RNA*) mediante un proceso llamado transcripción y a continuación, esta información transcrita se *traduce* en forma de una proteína. Así, la información biológica almacenada en el DNA fluye del DNA al RNA y por último, a las proteínas. En este capítulo se estudiará la estructura y las propiedades del DNA y del RNA con el fin de establecer las bases para estudiar posteriormente el flujo de la información biológica y el desarrollo del proyecto del genoma humano (PGH).

El descubrimiento de la sustancia, después denominada DNA fue realizado hace más de un siglo por Friedrich Miescher, médico suizo que se encontraba trabajando en el laboratorio del químico fisiólogo alemán Félix Hoppe-Seyler en la



Universidad de Tübingen. En 1869, Miescher comenzó el estudio con leucocitos obtenidos del pus presente en los vendajes de pacientes en post-operatorio, tratándolos con ácido clorhídrico y obteniendo así el núcleo celular. Al añadir una disolución alcalina a los núcleos purificados se formaba un precipitado cuyo análisis químico mostró que contenía carbono, hidrógeno, oxígeno, nitrógeno y un elevado porcentaje de fósforo. Miescher llamó a esta nueva sustancia *nucleína*, y, más tarde, cuando se descubrió su carácter ácido, se cambió su nombre a *ácido nucleico*.

EL DNA ES UN POLÍMERO LINEAL DE DESOXIRRIBONUCLEÓTIDOS UNIDOS MEDIANTE ENLACES FOSFODIÉSTER.

El largo paso desde el descubrimiento de Miescher de una sustancia desconocida, la nucleína, hasta el conocimiento detallado actual de la estructura tridimensional del DNA ha seguido una pauta similar de estudio, desde la composición química a la composición de las unidades, a la vía de ordenación de las subunidades y finalmente, al ordenamiento tridimensional del DNA en el espacio.

El progreso en la determinación de la estructura de la nucleína fue muy lento y fue en los años 30, más de 60 años después de su descubrimiento, cuando Albrecht Kossel y Phoebus Levene, realizaron los estudios químicos que establecerían que la nucleína era el ácido desoxirribonucleico (DNA). Estos análisis químicos establecieron que la subunidad repetitiva del DNA es un nucleótido y que contiene un azúcar (2'-desoxi-D-ribosa), un grupo fosfato y una de las cuatro bases nitrogenadas heterocíclicas (Figura 1.1). Dos de estas bases, la adenina (A) y la guanina (G), son púricas, mientras que las

otras dos, la citosina (C) y la timina (T), son pirimidínicas (Figura 1.2).

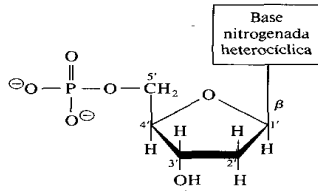


Fig.1.1 Estructura de un nucleótido (la subunidad básica del ADN)

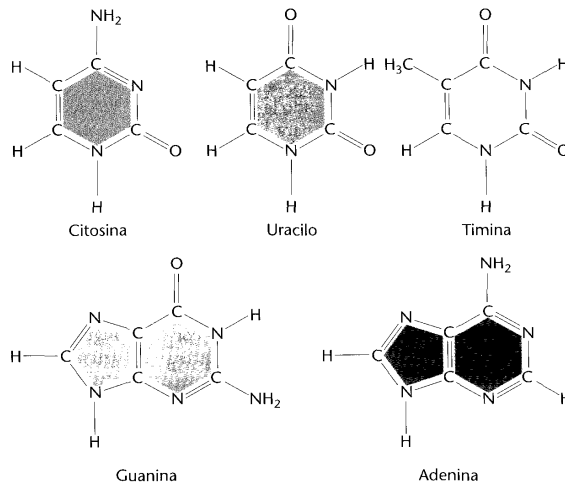


Figura 1.2 Estructura de las bases púricas y pirimidínicas



Los trabajos posteriores de Levene pusieron de manifiesto cómo estas piezas de este rompecabezas estructural encajan entre sí. Levene descubrió que una base (A, G, T o C) está unida al C-1' de la 2'-desoxi-D-ribose mediante un enlace β -N-glicosídico. A esta unidad estructural se la denominó nucleósido. Además, descubrió que el fosfato podía unirse al C-3' o al C-5' del azúcar. El nucleósido fosforilado constituye un nucleótido. Asimismo, Levene descubrió también que dos nucleótidos están unidos por un enlace covalente fosfodiéster entre el grupo hidroxilo de un nucleótido y el grupo fosfato de otro nucleótido. Debido a que los enlaces fosfodiéster en el DNA unen los carbonos 5' y 3' de los residuos de azúcar adyacentes, estos enlaces se denominan enlaces bifosfodiéster 5'-3' y esta unión de los 2'-desoxirribonucleótidos genera una cadena polinucleotídica (Figura 1.3).

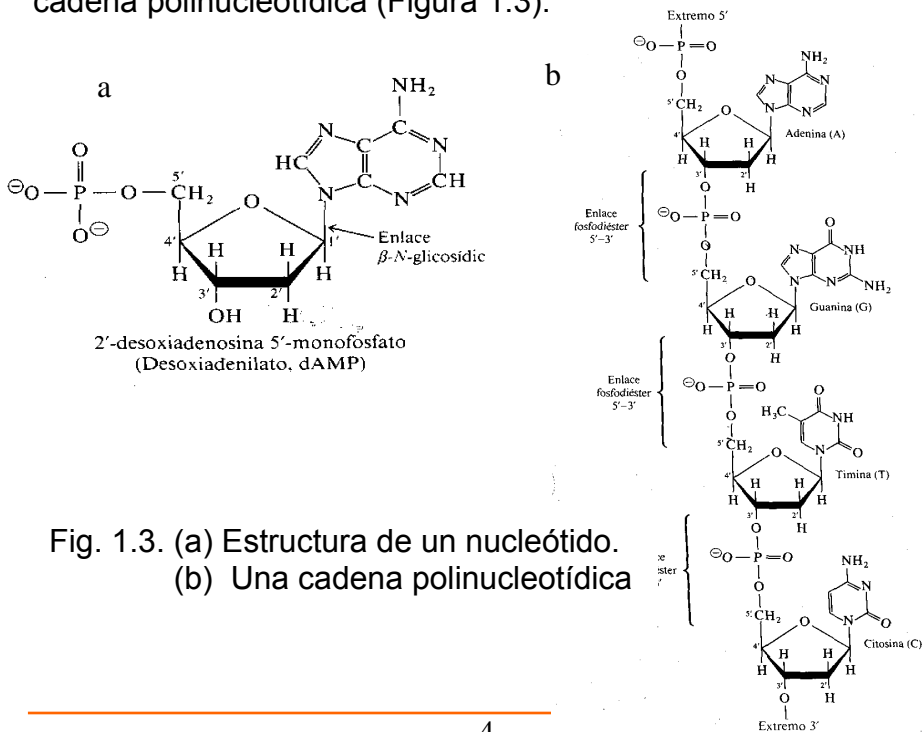


Fig. 1.3. (a) Estructura de un nucleótido.
(b) Una cadena polinucleotídica

Todos los nucleótidos en una cadena polinucleotídica tienen la misma orientación relativa, de modo que si en el primer nucleótido el carbono 5' está por encima del anillo de pentosa y el carbono 3' por debajo, la posición del carbono 5' se mantiene en los nucleótidos restantes. Por tanto, una cadena polinucleotídica es direccional y la dirección de avance se define como 5' a 3', es decir, partiendo del C-5' del azúcar, a través del C-4', al C-3' que conecta con el siguiente fosfato. Los extremos de las cadenas polinucleotídicas se denominan 5' y 3', denotando la dirección de las cadenas.

La fracción molar de cada base en una muestra de DNA, es decir, la composición de bases, puede determinarse hidrolizando la muestra y analizando cuantitativamente los productos. Debido a que la separación y el análisis cuantitativo eran difíciles, la composición de bases del DNA no se determinó hasta 80 años después de su descubrimiento, cuando se inventó la técnica de cromatografía en papel. A finales de los años 40, Erwin Chargaff empleó la cromatografía en papel para determinar la composición de bases de muestras de DNA aislada de diferentes organismos, encontrando que el número de moles de A y T, por un lado, y el número de moles de C y G, por otro, eran iguales. Aunque el significado de esta observación no se comprendió para aquel entonces, pudo explicarse elegantemente más tarde en términos de la estructura del DNA.

Así, hacia 1950 se supo que el DNA era un polímero lineal de residuos de desoxirribonucleótidos unidos por enlaces fosfodiéster 5'-3' y que contenía cuatro residuos de nucleósidos distintos dA, dT, dC y dG, de forma que los pares dA y dT, y dC y dG se encuentran en cantidades equimoleculares. Sin em-



bargo, aún se desconocía la estructura tridimensional del DNA, y la biología molecular se encontraba todavía en su infancia.

EL DNA ES UNA HÉLICE DOBLE

Análisis de difracción de rayos X

Cuando las fibras de una molécula de DNA se someten a un bombardeo de rayos X, los rayos se dispersan en función de la estructura atómica de la molécula. El patrón de dispersión puede registrarse sobre una película fotográfica en forma de manchas, y puede analizarse, especialmente en lo que se refiere a la forma general y las regularidades de la molécula. Este proceso, el análisis de difracción de rayos X, se había usado con éxito por Linus Pauling y otros químicos en el estudio de la estructura proteica. Esta técnica fue probada por William Astbury sobre DNA ya en 1938. En 1947, detectó una periodicidad de 3,4 Å, sugiriéndole que las bases estaban apiladas unas sobre las otras como monedas.

Entre 1950 y 1953, Rosalind Franklin, que trabajaba en el laboratorio de Maurice Wilkins, obtuvo datos más precisos de difracción de rayos X a partir de muestras más purificadas de DNA. Su investigación confirmó la periodicidad de 3,4 Å vista por Astbury y sugirió que la estructura del DNA era algún tipo de hélice. Sin embargo, no propuso ningún modelo definitivo. Pauling había analizado el trabajo de Astbury y otros, y propuso, incorrectamente, que el DNA era una triple hélice.

En 1953, casi 20 años después de que se elucidasen las características de la estructura covalente del DNA, James Watson y Francis Crick determinaron la estructura tridimensional del DNA. Utilizando los estudios de difracción de

rayos X de fibras de DNA obtenidos por Rosalind Franklin y colaboradores; Watson y Crick elucidaron esa estructura mediante una combinación de análisis de datos de difracción de rayos X, de modelos teóricos y de intuiciones.

El modelo de Watson y Crick tiene las características principales que siguen a continuación (figura 1.4):

1. Dos largas cadenas polinucleotídicas están enrolladas alrededor de un eje central, formando una doble hélice enrollada hacia la derecha (**dextrógira**).
2. Las dos cadenas son antiparalelas; es decir, la orientación C-5' - C-3' va en sentidos contrarios.
3. Las bases de las dos cadenas yacen formando estructuras planas y perpendiculares al eje; están «apiladas» unas sobre otras, separadas 3,4 Å (0,34 nm), y se encuentran en el interior de la estructura.
4. Las bases nitrogenadas de las cadenas opuestas están apareadas como resultado de la formación de puentes de hidrógeno. En el DNA sólo se permiten los emparejamientos A-T y G-C.
5. Cada vuelta completa de la hélice tiene una longitud de 34 Å (3,4 nm); de este modo, cada vuelta de la cadena contiene 10 bases.
6. En cualquier segmento de la molécula, se observan un **surco mayor** y un **surco menor** alternados a lo largo del eje.
7. La doble hélice mide 20 Å (2,0 nm) de diámetro.

La característica genéticamente más importante del modelo es la condición del emparejamiento entre bases (Punto 4 anterior). Otras características importantes son: Primero, la condición antiparalela de las dos cadenas es un punto clave del modelo de la doble hélice. Mientras una cadena discurre orientada 5' - 3' (pareciéndonos derecha), la otra cadena está orientada 3' - 5'



(pareciendo así invertida). Dadas las constricciones de los ángulos de enlace entre los distintos componentes de los nucleótidos, no se puede construir fácilmente la doble hélice si ambas cadenas discurren paralelas. Segundo, la característica dextrógira de la hélice se puede apreciar mejor si se compara esta estructura con su copia levógira, que es una imagen especular de la misma.

La clave del modelo propuesto por Watson y Crick es la especificidad en el emparejamiento de bases. Los datos de Chargaff sugieren que las cantidades de A y T son iguales, y que las de G y C también. Watson y Crick se dieron cuenta de que si A se empareja con T y C lo hace con G, justificando estas proporciones, los componentes de estos pares de bases formarían puentes de hidrógeno (Figuras 1.4 y 1.5), proporcionando la estabilidad química necesaria para mantener las dos cadenas juntas. Ordenadas de este modo, se hace aparente a lo largo del eje un surco mayor y uno menor. Más aún, una purina (A o G) está al otro lado de una pirimidina (T o C) en cada "escalón de la escalera de caracol" de la hélice propuesta, justificando los 20 Å (2 nm) de diámetro sugerido por los estudios de difracción de rayos X.

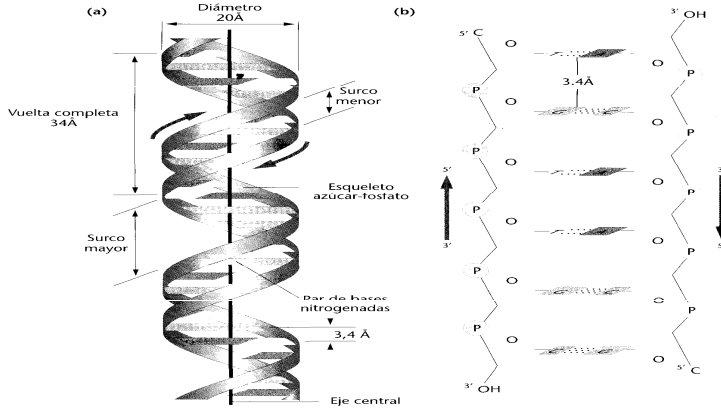


Figura 1.4. (a) Modelo de la doble hélice propuesta por Watson y Crick.
(b) Representación del carácter antiparalelo de las hebras de ADN

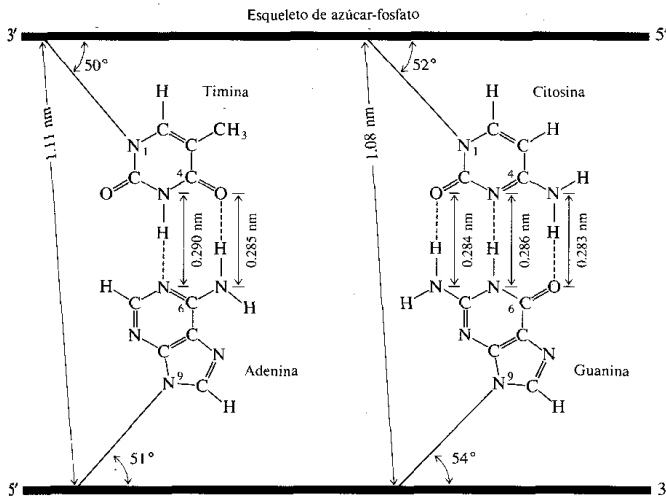


Figura 1.5 Apareamiento y dimensiones de las bases nitrogenadas (AT y GC) entre el esqueleto azúcar fosfato de cada hebra



LA SOLVATACIÓN, EL APILAMIENTO DE LAS BASES Y LOS ENLACES DE HIDRÓGENO ESTABILIZAN LA DOBLE HÉLICE

Las fuerzas responsables de las conformaciones nativas de las estructuras complejas celulares, como los ácidos nucleicos, las proteínas y las membranas, son siempre las mismas. Estas fuerzas son lo suficientemente fuertes como para mantener la integridad estructural, pero lo suficientemente débiles como para permitir una flexibilidad conformacional. Los enlaces covalentes son, por supuesto, muy importantes, ya que suministran el aglutinante que une los átomos en las moléculas; sin embargo, las fuerzas débiles (el efecto hidrofóbico, las fuerzas de Van der Waals, los enlaces de hidrógeno y los efectos electrostáticos) son las que determinan las formas plegadas de las estructuras biológicas. Pueden distinguirse cuatro factores principales que contribuyen a la estabilidad del DNA.

1. Los efectos hidrofóbicos estabilizan los apareamientos de las bases. Los anillos hidrofóbicos de purina y pirimidina de las bases son empujados hacia el centro de la doble hélice en virtud de la elevada cohesión interna de las moléculas de agua, mientras que los sustituyentes hidrofílicos de las bases se encuentran expuestos al disolvente en los surcos.

2. Las bases apiladas forman contactos de Van der Waals. Los pares de bases se encuentran, apilados unos sobre otros, a lo largo del eje central de la doble hélice. La altura por vuelta de hélice en el B-DNA es igual al espesor de los anillos de purina y pirimidina; por tanto, los anillos forman contactos de Van der Waals. Las fuerzas de Van der Waals entre las bases apiladas son débiles pero aditivas, de forma que en una molécula que

contenga más de 10^4 pares de bases, las fuerzas de Van der Waals suponen una importante fuente de estabilidad.

3. Los pares de bases están unidos mediante enlaces de hidrógeno. El par de bases CG es más estable que el par AT por contener un enlace de hidrógeno más.

4. El esqueleto del DNA interacciona con cationes. El esqueleto del DNA, de tipo fosfodíéster, tiene carácter ácido, y a pH 7.0 presenta una gran carga negativa. La repulsión electrostática entre los grupos fosfodíéster negativos es una fuente potencial de inestabilidad de la doble hélice; sin embargo, los cationes celulares, y en particular el magnesio, se unen fuertemente al esqueleto fosfodíéster del DNA, estabilizando la doble hélice.

La doble hélice de DNA se desenrolla localmente durante procesos tales como la replicación del DNA, la transcripción al RNA y la recombinación génica. El desenrollamiento completo del DNA puede ocurrir *in vitro* y se denomina *desnaturalización* del DNA o *transición hélice-cadena*. Este fenómeno tiene lugar al romperse los enlaces de hidrógeno entre las bases con la separación consiguiente de los pares de bases, como cuando se calienta el DNA por encima de una temperatura determinada, o se funde. La desnaturalización del DNA puede seguirse espectroscópicamente midiendo la variación de la absorción de la luz ultravioleta por el DNA. El DNA presenta el máximo de absorción de luz UV a 260 nm. Una vez completamente desenrollado el DNA, las bases desapareadas y no apiladas absorben a 260 nm un 37 % más que cuando se encuentran en la estructura nativa de doble hélice.



EL ÁCIDO RIBONUCLEICO ES UN POLÍMERO LINEAL DE RIBONUCLEÓTIDOS

No mucho después de que Miescher descubriese aquella sustancia que luego resultó ser el DNA, Félix Hoppe-Seyler. Un científico del mismo laboratorio, descubrió otra sustancia muy similar al DNA; esta última sustancia, ahora conocida como RNA, fue aislada en primer lugar a partir de levadura y más tarde de bacterias y plantas. Durante mucho tiempo se pensó que el RNA estaba ausente en los animales; cuando se descubrió su presencia en animales se creía debida al alimento vegetal. Este punto de vista prevaleció hasta 1914. Cuando Robert Feulgen descubrió un colorante que teñía el DNA, pero no el RNA, y otro que teñía sólo el RNA; al teñir células con ambos colorantes descubrió la presencia conjunta de DNA y RNA en todas las células.

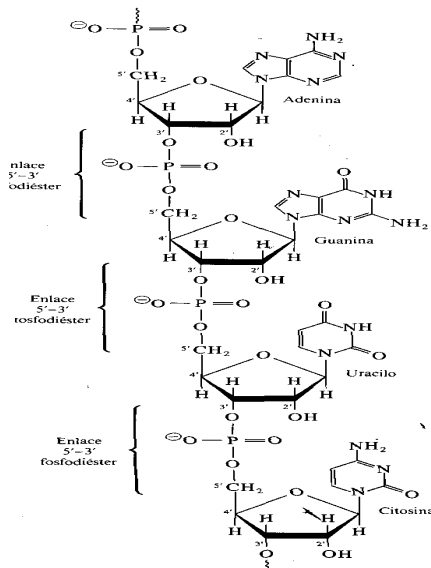


Figura 1.6 Cadena polinucleotídica de ARN

El esqueleto covalente del RNA consiste en un polímero lineal de unidades ribonucleotídicas unidas mediante enlaces fosfodiéster 5'-3' (Figura 1.6). En este aspecto el DNA y el RNA son idénticos; no obstante, el RNA se diferencia estructuralmente del DNA en tres puntos importantes:

1. El grupo azúcar del RNA es la ribosa, y no la 2'-desoxirribosa.
2. La timina se sustituye por uracilo (U) en lo referente a las bases nitrogenadas. Recuérdese que la timina posee un grupo metilo en el C-5 cuya sustitución por un hidrógeno resulta en el uracilo. Las bases nitrogenadas comunes del RNA son la adenina, el uracilo, la guanina y la citosina.
3. Las moléculas de RNA son generalmente monocatenarias. Sin embargo, en una única cadena de RNA el apareamiento de bases de Watson-Crick puede ocurrir entre la adenina y el uracilo y entre la guanina y la citosina, y resultar en toda clase de estructuras secundarias, entre las que cabe citar las estructuras en bucle y en horquilla que participan en el reconocimiento del RNA por proteínas. Unas cuantas moléculas de RNA son bicatenarias y su conformación es análoga a la del A-DNA. En este sentido, la doble hélice de RNA completa una vuelta cada 11 pares de bases; los pares de bases se encuentran inclinados lejos del eje central de la hélice, de forma que permite la solvatación de los grupos hidroxilos de los C-2' de los azúcares. La existencia de una hélice doble de RNA similar al B-DNA no es posible debido a que los grupos 2'-hidroxilo no se encontrarían solvatados.



Estructuras helicoidales bicatenarias en las que una cadena es ADN y la otra RNA existen en las células en varios momentos. Por ejemplo, en la transcripción se forma una molécula de RNA, copia de una cadena de DNA en una reacción catalizada por la RNA polimerasa, de forma que la molécula de RNA sintetizada es complementaria a la de DNA y, por tanto, durante el proceso de elongación se forma un híbrido pequeño gracias al apareamiento entre la dA y el U, la dT y la A, la dC y la G y la dG y la C. Durante el estudio mediante difracción de rayos X de híbridos sintéticos grandes de RNA y DNA se observó que adoptaban la conformación común al RNA y al DNA, es decir, la conformación conocida como A.

EXISTEN VARIOS TIPOS DIFERENTES DE MOLÉCULAS DE ARN

Si bien el DNA es el depositario celular de la información genética, muchas moléculas de RNA participan en el proceso de expresión de tal información. En una célula dada, las moléculas de RNA se encuentran en múltiples copias y formas. Las moléculas de RNA se clasifican atendiendo tanto a su localización celular como a su función. De éste modo, en las células procarióticas se diferencian básicamente tres formas mayoritarias de RNA:

1. El RNA mensajero (mRNA), que transporta la información genética del DNA a los ribosomas, orgánulos celulares responsables de la síntesis de proteínas.
2. El RNA ribosómico (rRNA), que es una parte constitutiva de los ribosomas y que da cuenta de aproximadamente el 75 % del RNA total celular.

3. El RNA de transferencia (tRNA), que transporta los residuos de aminoácidos que son adicionados a las cadenas polipeptídicas crecientes durante la síntesis de proteínas. Estos tres tipos de RNA fueron los primeros para los que se describió una función bioquímica. Las células eucarióticas contienen, además de estos tres tipos de moléculas de RNA, una población de moléculas grandes de RNA nuclear de peso molecular muy variable. Estas moléculas de RNA heterogéneo nuclear (hnRNA) son los precursores del mRNA maduro. Otro grupo de moléculas de RNA, pequeñas y también ubicadas en el núcleo de las células eucarióticas, se encuentran unidas a proteínas formando unos complejos conocidos como partículas de ribonucleoproteínas pequeñas nucleares (snRNPs o "snurps"), y que desempeñan un papel importante en la síntesis del mRNA. Por otra parte, en el citoplasma también hay "RNAs pequeños" asociados con proteínas específicas, algunos de los cuales desempeñan un papel estructural, mientras que otros son necesarios para la catálisis de ciertas reacciones. Además, en el proceso de replicación del DNA se sintetizan varios RNAs pequeños. Finalmente, el RNA, y no el DNA, constituye la dotación del material genético de algunas bacterias y virus animales.

EL RNA ES QUÍMICAMENTE MÁS REACTIVO QUE EL DNA

El cambio de la 2'-desoxirribosa en el DNA por la ribosa en el RNA puede parecer insignificante y, sin embargo, afecta en gran medida a las propiedades del RNA. El grupo 2'-hidroxilo: (1) impide a las moléculas de RNA la adopción de la conformación B, (2) permite que en las moléculas de RNA ocurra un número mayor de interacciones terciarias y (3) promueve la reactividad química, de hecho, las moléculas de RNA son componentes imprescindibles en muchos procesos



enzimáticos. Aunque en la mayoría de estas reacciones el RNA tiene una función estructural, también se conocen ejemplos de moléculas de RNA con función catalítica.

Un ejemplo que ilustra la importancia del grupo 2'-hidroxilo es el comportamiento del RNA frente a las disoluciones alcalinas. En este sentido, el tratamiento del RNA con un álcali 0.1 M a 25 °C produce, al cabo de pocas horas, una mezcla de nucleósidos 2'- y 3'-monofosfato. Sin embargo, el DNA bajo estas mismas condiciones es bastante estable. En presencia de un ión hidróxido el grupo 2'-hidroxilo de la ribosa del RNA se convierte en su base conjugada, un anión alcóxido, que es un nucleófilo poderoso. El 2'-alcóxido ataca al grupo 3'-fosfodiéster adyacente, originando un nucleósido 2',3'-monofosfato cíclico. Durante este proceso el enlace 5'-3' fosfodiéster se rompe y, por consiguiente, la molécula de RNA se escinde. En esta etapa, un enlace 5'-3' fosfodiéster se convierte en un enlace 2'-3' fosfodiéster. Luego la primera etapa de la hidrólisis alcalina del RNA es una reacción de transesterificación. La segunda parte de la reacción implica la hidrólisis de un fosfodiéster cíclico que ocurre por ataque de un segundo anión hidróxido a este derivado de cinco eslabones tenso y, por consiguiente, termodinámicamente inestable. La apertura del fosfodiéster cíclico origina una mezcla de nucleósidos 2'- y 3'-monofosfato. El DNA es estable en disoluciones alcalinas por carecer de un grupo 2'-hidroxilo que lleve a cabo la catálisis intramolecular.

RNA NUCLEAR HETEROGÉNEO Y SU PROCESAMIENTO: CAPERUZAS Y COLAS

A partir del estudio del RNA se han obtenido ideas sobre otras regiones del DNA que no codifican directamente proteínas. Esta investigación ha proporcionado un conocimiento detallado de la estructura génica de los eucariotas. El código genético

está escrito en la secuencia ribonucleotídica del mRNA. Por supuesto, esta información se originó en la cadena con sentido del DNA, en la que hay secuencias complementarias de desoxirribonucleótidos. En bacterias, la relación entre DNA y RNA parece ser muy directa. La secuencia de bases del DNA se transcribe a una secuencia de mRNA, que luego se traduce a una secuencia de aminoácidos de acuerdo con el código genético. Sin embargo, la situación es mucho más compleja en eucariotas que en bacterias. Se ha encontrado que muchas secuencias de bases internas de un gen no aparecen nunca en el mRNA maduro que se traduce. También se producen otras modificaciones al inicio y al final del mRNA antes de la traducción. Estos descubrimientos han puesto en evidencia que en los eucariotas se produce un complejo procesamiento del mRNA antes de que sea transportado al citoplasma para participar en la traducción.

En 1970, las pruebas acumuladas mostraban que el mRNA eucariótico se transcribe inicialmente como una molécula precursora más larga que la que se traduce. Esta noción se basaba en la observación de James Darnell y sus colaboradores del **RNA nuclear** heterogéneo (hnRNA) en células de mamífero. El RNA heterogéneo, acompañado con una abundante variedad de proteínas (formando los **hnRNP - ribonucleoproteínas nucleares heterogéneas**), es grande y de tamaño variable (hasta 10^7 daltons). Se encuentra sólo en el núcleo. Es importante decir que se encontró que el hnRNA contenía secuencias nucleotídicas comunes con las moléculas de mRNA más pequeñas que hay en el citoplasma. Debido a esta observación, se propuso que el transcrito inicial de un gen era una larga molécula de RNA que primero debía procesarse en el núcleo, antes de aparecer en el citoplasma como una molécula de mRNA madura. Los distintos pasos del pro-



cesamiento, comentados, están esquematizados en la Figura 1.7.

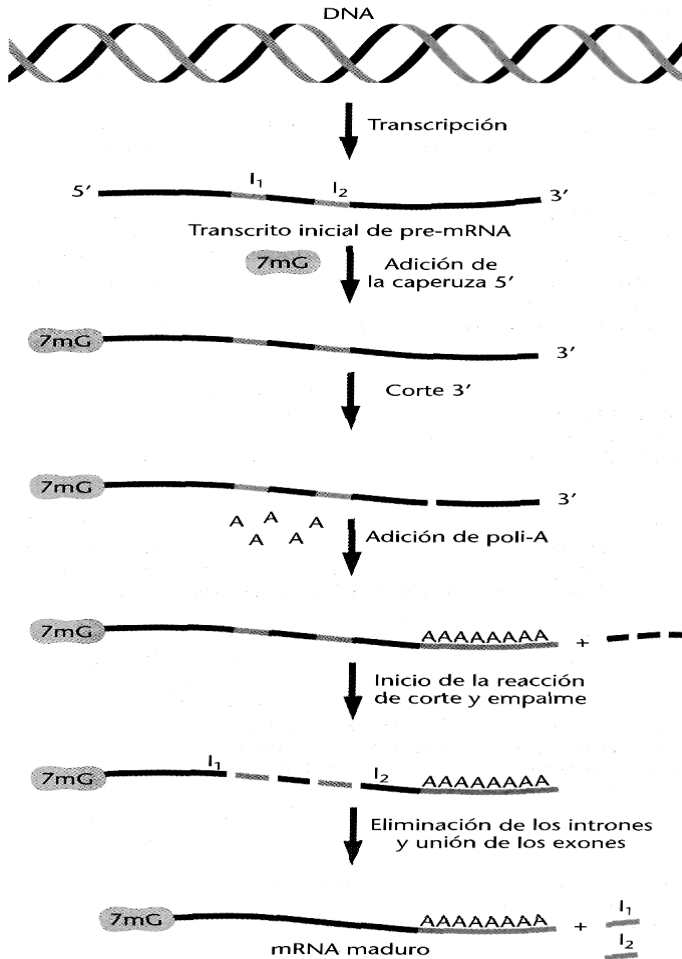


Figura 1.7 Conversión del ARN nuclear heterogéneo (hnRNA) mensajero en eucariotas

Las **modificaciones postranscripcionales** de los transcritos de RNA eucariótico destinados a convertirse en mRNA abarcan el extremo 5' de estas moléculas, donde se añade una caperuza de 7-metil-guanosina (7mG). La caperuza parece ser importante para el procesamiento posterior en el núcleo, quizá protegiendo el extremo 5' de la molécula del ataque de nucleasas. Posteriormente, esta caperuza puede estar implicada en el transporte del mRNA, maduro por la membrana nuclear hacia el citoplasma; la caperuza es bastante compleja y se caracteriza por la unión especial 5'-5' entre la caperuza y el ribonucleótido inicial del mRNA. Algunos eucariotas contienen también un grupo metilo ($-\text{CH}_3$) en el carbono 2' del azúcar ribosa de los dos primeros ribonucleótidos del RNA.

Un descubrimiento posterior proporcionó más ideas sobre el procesamiento de los transcritos de RNA durante la maduración del mRNA. Se ha encontrado que tanto los hnRNA como los mRNA tienen en su extremo 3' un alargamiento de hasta 250 residuos de ácido adenílico. Estas secuencias de poli-A se añaden después de que se haya añadido la caperuza 5'-7mG. Primero se corta enzimáticamente el extremo 3' del transcrito inicial en un punto situado a unos 10-35 ribonucleótidos de la secuencia AAUAAA, altamente conservada. Luego se produce la poliadenilación mediante la adición secuencial de residuos de ácido adenílico. Actualmente se ha encontrado poli-A en el extremo 3' de casi todos los mRNA estudiados en diversos organismos eucarióticos. Los productos de los genes (proteínas) de las histonas parecen ser la excepción.



La importancia de la secuencia AAUAAA y de la cola 3' se hace evidente cuando se investigan mutaciones de esta secuencia. Las células que llevan estas mutaciones no pueden añadir la secuencia de poli-A, y en ausencia de esta cola, los transcritos de RNA se degradan rápidamente. Por lo tanto, la cola de poli-A es esencial si un transcrito de RNA se ha de seguir procesando y ha de ser transportado al citoplasma.

SECUENCIAS INTERCALADAS Y GENES FRAGMENTADOS (intrones y exones)

Uno de los descubrimientos más excitantes en la historia de la genética molecular sucedió en 1977 cuando Susan Berget, Philip Sharp y Richard Roberts presentaron pruebas directas de que los genes de los virus animales contenían secuencias nucleotídicas *internas* que no se expresaban en la secuencia de aminoácidos de la proteína que codifican. Esto se debe a que ciertas secuencias internas presentes en el DNA no aparecen en el mRNA maduro que se traduce a proteína.

Se ha denominado a estos segmentos nucleotídicos secuencias intercaladas, y están contenidas en genes fragmentados. Estas secuencias de DNA que no están representadas en el producto final del mRNA también se denominan intrones («int» por intercaladas), y las conservadas y expresadas, exones («ex» por expresadas). La eliminación de las secuencias presentes en los intrones produce como resultado de un proceso de escisión y unión del RNA, denominado corte y empalme (*splicing*). Las secuencias de los intrones están presentes en los transcritos iniciales del RNA, pero son eliminadas antes de que se traduzca el mRNA maduro (Figura 1.7).

Pronto se hicieron descubrimientos similares en diversos eucariotas. Dos enfoques han sido los más fructíferos. El primero comprende la hibridación molecular de mRNA maduro, purificado y funcional, con el DNA que contiene los genes que especifican el mensaje. Cuando se produce hibridación entre ácidos nucleicos que no son exactamente complementarios, se producen hetero-dúplex que pueden visualizarse en el microscopio electrónico.

El segundo enfoque proporciona una información más específica. Comprende la comparación de las secuencias nucleotídicas del DNA con las del mRNA, y su correlación con las secuencias de aminoácidos. Este enfoque permite la identificación precisa de las secuencias intercaladas.

Hasta ahora, se ha demostrado que un gran número de genes de diversos eucariotas contienen intrones, Uno de los primeros en identificarse fue el gen de la beta-globina de ratón y de conejo, estudiado independientemente por Philip Leder y Richard Elavell. El gen del ratón tiene un intrón de 550 nucleótidos, que empieza inmediatamente después del codón que especifica el aminoácido 104. En el conejo hay un intrón de 580 pares de bases cerca del codón del aminoácido 110. Además, en ambos genes hay un intrón anterior de unos 120 nucleótidos. Se han encontrado intrones similares en el gen de la beta-globina en todos los mamíferos examinados (Figura 1.8).

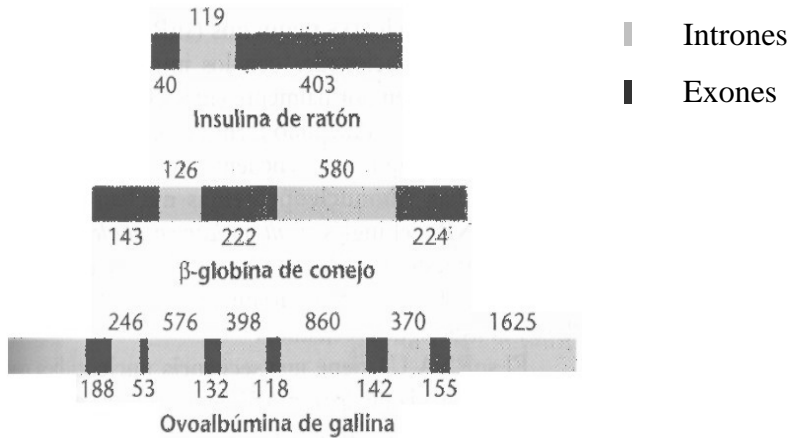


Figura 1.8 Intrones y exones de diversos genes de eucariotas. Los números indican la cantidad de nucleótidos de las distintas regiones

Descubrimientos como éstos en mamíferos han probado que la edición de RNA proporciona otro importante mecanismo de modificación postranscripcional, y que este proceso no está restringido a los sistemas genéticos menos evolucionados como las mitocondrias. Así, parece probable que, al continuar investigando, se encuentre que la edición de RNA esté más extendida. Además, este proceso tiene implicaciones importantes en la regulación de la expresión genética.

ENZIMAS DE RESTRICCIÓN

Las enzimas de restricción son endonucleasas que reconocen una secuencia de entre 4-8 bp en una molécula de ADN de doble hebra. El sitio de reconocimiento se llama sitio de restricción y la enzima rompe un enlace fosfodiéster en la hebra

guía y otro enlace fosfodiéster en la hebra complementaria. Son extraídas de organismos procarióticos (bacterias) y se ha visto que ciertas cepas de *E. coli* degradan (cortan) el DNA de ciertos virus fagos infectivos, a menos que los ácidos nucleicos presenten metilación (adición de grupos metilo) en algunos residuos de adenina o citosina del sitio de corte. La función de las enzimas de restricción es la proteger al organismo de DNA extraño. Cuando una porción de DNA foráneo ingresa a la célula, las enzimas de restricción se encargan de degradarlo cortándolo en pequeños fragmentos, siempre y cuando éste no esté modificado.

Existen 3 tipos de enzimas de restricción:

1. Tipo I y Tipo III:

- a. Tienen actividad de restricción (cortan) y modificación (metilan).
- b. Cortan a cierta distancia de la secuencia de reconocimiento, las Tipo I cortan lejos de la secuencia de reconocimiento, ya sea río arriba o río abajo. Las Tipo III cortan de 5-8 bases antes o después de la secuencia que reconocen.
- c. Necesitan ATP para moverse a través de la molécula de DNA, desde el lugar de reconocimiento hasta el sitio del corte.

Tipo II:

- a. Sólo tienen actividad de restricción.
- b. Cortan de manera consistente y predecible dentro de la secuencia que reconocen.
- c. Sólo requieren Mg^{++} como cofactor.
- d. No necesitan ATP.

Las endonucleasas se nombran a partir de las bacterias de las que son extraídas, su nombre está dado según el género y la



especie de la bacteria de donde se aisló por primera vez esta enzima. La primera letra representa el género de la bacteria, las próximas dos indican la especie, una cuarta letra indica la cepa, y un número al final indica la cantidad de enzimas que se han aislado de esa cepa. Ej:

Eco RI → E = género Escherichia
co = especie coli
R = cepa RV 13
I = primera endonucleasa aislada de esta cepa

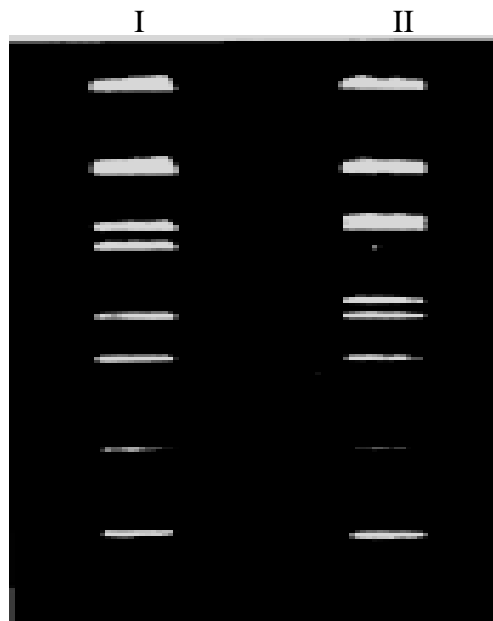


Figura 1.9 Digestión de una muestra de DNA con dos enzimas de restricción distintas

Las enzimas de restricción son una herramienta muy importante en la ingeniería genética, ya que permiten la digestión de secuencias específicas del DNA. Las secuencias digeridas se separan en bandas individuales de distintos tamaños con la ayuda de electroforesis en geles de agarosa (figura 1.9); estas digestiones (que generan bandas de DNA) son fundamentales para la extracción de un gen de interés (presente en un genoma o plásmido, por ejemplo) o para estudios de polimorfismo de poblaciones, entre otras aplicaciones que son imprescindibles para el análisis genético.



Capítulo 2

Proyecto Genoma



ESTUDIOS GENÓMICOS

El objetivo de la Genómica es comprender la organización molecular y la información contenida en el genoma completo y en los productos génicos cifrados por éste. Esta subdisciplina de la genética emplea muchas de las técnicas analíticas que el genetista aplica a genes individuales o a regiones cromosómicas pequeñas y las extiende de forma global a todo el genoma. De tal modo, resulta posible el estudio directo de la organización a gran escala de los genes y los cromosomas, y el de la regulación global de los genes.

A pesar de los considerables obstáculos técnicos, podemos estar seguros de que, en unos pocos años, tendremos un catálogo completo con las secuencias de nucleótidos y aminoácidos de todos los genes y los productos génicos cifrados en los genomas de muchos organismos complejos, incluida la especie humana. El disponer de dichos catálogos proporcionará la materia prima que puede servir como origen para la comprensión de cualquier proceso, desde los asuntos prácticos de las enfermedades humanas y la genética relacionada con la agricultura a los fenómenos biológicos básicos tales como los que constituyen el fundamento de la fisiología celular, el desarrollo, el comportamiento, la ecología y la evolución. La disponibilidad de estos catálogos es un suceso científico excitante y digno de la transición del milenio, y promete tener efectos dramáticos en el proceso de la investigación científica de la Biología.



GENÓMICA. UNA VISION GENERAL

La genómica se divide en dos áreas: la GENÓMICA ESTRUCTURAL, cuya finalidad es la caracterización de la naturaleza física de genomas enteros; y la GENÓMICA FUNCIONAL, que se encarga de caracterizar el transcriptoma (conjunto completo de los transcritos producidos por un organismo concreto) y el proteoma (colección completa de las proteínas cifradas en el genoma)

El principal objetivo del análisis genómico estructural es la determinación completa y precisa de la secuencia de DNA de un genoma haploide representativo de una determinada especie. El conocimiento de esta secuencia abre la puerta a numerosas posibilidades. A través del análisis computarizado de la secuencia, empleando los principios desarrollados por la genética y la biología molecular para el análisis de los transcritos y proteínas, podemos predecir todas las proteínas cifradas. Por otro lado, podemos analizar otros genomas haploides de la misma especie y desarrollar una descripción estadística de la variación genética existente entre las poblaciones de dicha especie; y podemos comparar las secuencias genómicas de especies distintas para, de esta manera, comprender cómo se han reestructurado los genomas en el curso de la evolución.

Los estudios de GENÓMICA COMPARADA han avanzado ya lo suficiente como para revelar que, en especies relacionadas (por ejemplo, entre los mamíferos), se mantiene una considerable sintenia (localización conservada de genes dentro de grandes bloques del genoma). Los estudios de Genómica comparada ofrecen también una herramienta poderosa para la identificación de motivos muy conservados, y por lo tanto funcionalmente importantes dentro de las secuencias

genómicas codificantes y no codificantes. Esto ayuda a los investigadores a confirmar las predicciones sobre las regiones del genoma que cifran proteínas y a identificar elementos reguladores importantes del DNA.

Aunque la genómica estructural tuvo su inicio hace tan solo una década y ya se han obtenido las secuencias de muchos genomas, el recorrido desde los mapas genéticos clásicos hasta llegar a completar los mapas de secuencias de DNA no se hizo en un único paso. Más bien, de manera similar a como se va incrementando progresivamente el aumento en un microscopio óptico, hubo un avance paso a paso en la resolución de los mapas genómicos durante el desarrollo de las tecnologías de la genómica.

En este capítulo, centraremos nuestra atención en el desarrollo de la genética de alta resolución y de las tecnologías de cartografía física que permitieron, en última instancia, la secuenciación de genomas complejos. Estas tecnologías no representaron tan solo pasos valiosos en la forma de realizar mapas hasta el nivel de la secuencia, sino que también suministraron herramientas que por sí mismas eran de gran importancia para la identificación de enfermedades genéticas y para la clonación posicional.

Pronto, se evidenció que la disponibilidad de genomas completamente secuenciados estimulaba el apetito científico hacia la obtención de información global adicional. En particular, la transformación de la “piedra Roseta” de la secuencia genómica en predicciones rigurosas de las secuencias de los transcritos y las proteínas resultó ser un desafío en si mismo, y ello condujo a que evolucionaran los



proyectos para la caracterización directa de las estructuras y las secuencias de todos los RNA y los polipéptidos.

Estos proyectos han supuesto la base de la creación de la Genómica funcional, en la que las estructuras de los transcritos se han caracterizado mediante la secuenciación de los cDNA completos y la comparación de las secuencias obtenidas con las de los correspondientes DNA genómicos. Como veremos al final de éste capítulo, la disponibilidad de estas secuencias de DNA ha permitido el desarrollo de los micromatrices extraordinariamente densas en las que cada posición de la matriz representa un RNA diferente. Estas micromatrices (o Chips), que constituyen un *transcriptoma* completo, pueden colocarse en un solo portaobjeto y exponerse a sondas para la caracterización de las concentraciones de los transcritos en un tipo celular concreto y bajo unas determinadas condiciones ambientales.

Estos experimentos de hibridación permiten el análisis de, literalmente, cientos de miles de datos en una sola tarde y proporcionan una información global sobre como una determinada condición influye en las actividades de los genes. Asimismo, de la misma manera que las estrategias empleadas para la transcriptoma, están en desarrollo vías para la identificación sistemática y global del *proteoma* (esto es, todas las proteínas que puede producir una especie). Muchos procesos biológicos que implican la toma de decisiones requieren modificaciones de las proteínas y cambios en las interacciones proteína-proteína, comprender el proteoma (y el transcriptoma) es tan importante como entender el genoma.

PROYECTOS GENOMA. CONSIDERACIONES PRÁCTICAS

Hay en marcha proyectos genoma en varios organismos distintos, incluyendo la especie humana y varios organismos modelos. Los sistemas modelos son los mismos que se han explotado profundamente en el análisis genético tradicional. Estos incluyen a *Mus musculus* (el ratón), *Drosophila melanogaster* (la mosca de la fruta), *Saccharomyces cerevisiae* (levadura del pan), *Caenorhabditis elegans* (un nematodo), *Arabidopsos thaliana* (una planta) y varias bacterias. Los primeros genomas en secuenciarse por completo fueron los mas pequeños. Primero se completaron las secuencias de los genomas víricos, a las que siguieron las de los genomas de los cloroplastos y las mitocondrias. A continuación se secuenció el primero de una serie de genomas bacterianos. En estos casos, algunos de los genomas se escogieron por su interés genético, otros para el análisis de la diversidad evolutiva entre los procariotas y otros porque los organismos son patógenos importantes de los humanos.

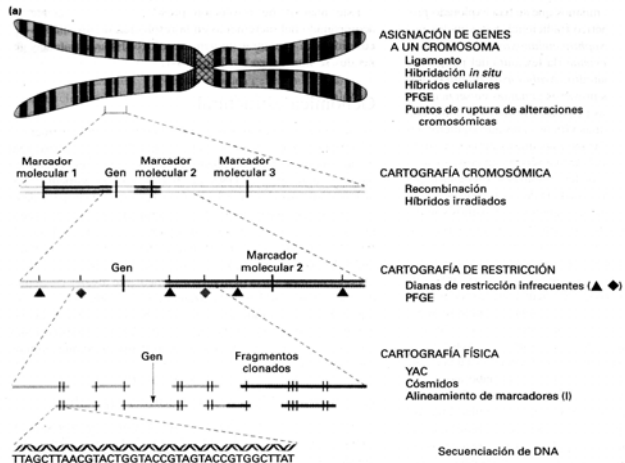


Figura 2.1
Resumen de las estrategias generales de la Genómica estructural. Esquema general de la elaboración de un mapa genómico mediante métodos analíticos de resolución creciente.



En 1996, se publicó la primera secuencia completa de un genoma eucariótico, el de la levadura *Saccharomyces cerevisiae*. Debido a la magnitud de estas tareas, muchos de los proyectos se llevan a cabo a través de acuerdos internacionales, en los que participan cientos de investigadores que colaboran y comparten datos sobre distintas regiones del genoma. A menudo, los grupos o naciones enteras, incluso, se especializan en el análisis de ciertos cromosomas concretos. Puesto que estos esfuerzos requieren una experimentación a una escala superior a la que puede llevar a cabo un único laboratorio, los proyectos genoma han tenido éxito gracias a la unión de diversos profesionales (genetistas, biólogos moleculares, químicos, físicos, ingenieros e informáticos) para el desarrollo de la tecnología necesaria, incluyendo la automatización de los diversos pasos del proceso. Este esfuerzo interdisciplinario del análisis genómico es una continuación de la historia científica de la genética, que se ha beneficiado en muchos aspectos de la interacción intelectual con otras disciplinas.

Antes de la disponibilidad del análisis genómico, los fundamentos genéticos de nuestro conocimiento a cerca de un organismo se debían a los mapas cromosómicos, de resolución relativamente bajo, de los genes que producían fenotipos mutantes conocidos y de algunos marcadores moleculares. A partir de este punto, el análisis genómico se realiza, por regla general, siguiendo varios pasos de resolución creciente:

- 1.- Situar genes y marcadores moleculares en mapas genéticos de alta resolución de cada cromosoma.
- 2.- Caracterizar y localizar físicamente, unos respecto a otros fragmentos de ADN clonados, con objeto de crear mapas

físicos de cada cromosoma. El mapa genético del genoma puede compararse entonces con el mapa físico.

3.- Llevar a cabo análisis a gran escala de la secuencias del genoma para producir un mapa con la secuencia completa de cada cromosoma. Los mapas físicos y genéticos pueden entonces compararse con el mapa de secuencia.

Este análisis de resolución progresivamente creciente va acompañado del incremento de la resolución de los análisis necesarios para encontrar un gen específico (ver fig. 2.1)

GENÓMICA ESTRUCTURAL

Como sugiere el nombre, el objetivo de la genómica estructural es caracterizar la estructura del genoma. Conocer la estructura de un genoma individual puede ser útil para la manipulación de genes y segmentos de ADN de una especie en particular. Por ejemplo, los genes se pueden clonar en base a el conocimiento de su posición en el genoma. Cuando se han caracterizado varios genomas a nivel de su estructura es de esperar que, mediante la genómica comparada, resultara posible deducir las reglas generales que gobiernan la organización estructural de todos los genomas.

La genómica estructural progresa mediante niveles crecientes de resolución analítica, comenzando con la asignación de genes y marcadores a cromosomas individuales, siguiendo con la cartografía de estos genes y marcadores dentro de los cromosomas y finalizando con la preparación de un mapa físico que culmine con la secuenciación.



ASIGNACIÓN DE LOCCI A CROMOSOMAS ESPECÍFICOS

Existen varios métodos diferentes que permiten asignar genes o marcadores a cromosomas individuales entre los que se encuentran: los microsátélites, minisátélites, hibridación *in situ* y otros que se discutirán a continuación.

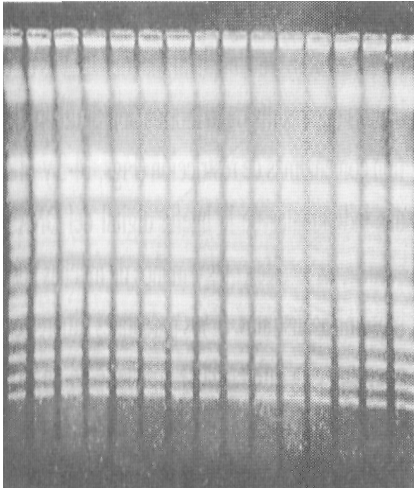
LIGAMENTOS A LOCI CONOCIDOS

En los organismos bien estudiados, resulta sencillo cruzar una estirpe que lleva el alelo “nuevo”, de posición desconocida, con una serie de estirpes que tienen marcadores dispersos a lo largo del genoma cada uno en una posición cromosómica conocida. La frecuencia de recombinación meiótica menor del 50 % indica que el alelo no cartografiado y un marcador completo están ligados y, por lo tanto, se encuentran en el mismo cromosoma. A menudo, estos datos de ligamiento dan una idea aproximada de la posición cromosómica, quizás el brazo cromosómico o la banda, incluso en la que está.

ELECTROFORESIS DE CAMPO PULSANTE (ECP)

Si los cromosomas son lo suficientemente pequeños para que se puedan separar mediante electroforesis de campo pulsante (fig. 2.2), las bandas de ADN se pueden utilizar para localizar los nuevos genes mediante hibridación. En primer lugar, se debe establecer que banda de ADN corresponde a cada cromosoma. El tamaño de los cromosomas, las translocaciones entre cromosomas y la hibridación con sondas cuyas posiciones en los cromosomas se conocen son útiles para este propósito. A continuación un nuevo gen clonado puede utilizarse como sonda en un Southern del gel de ECP; y de esta forma, se puede determinar cual es el locus cromosómico del gen.

a



b

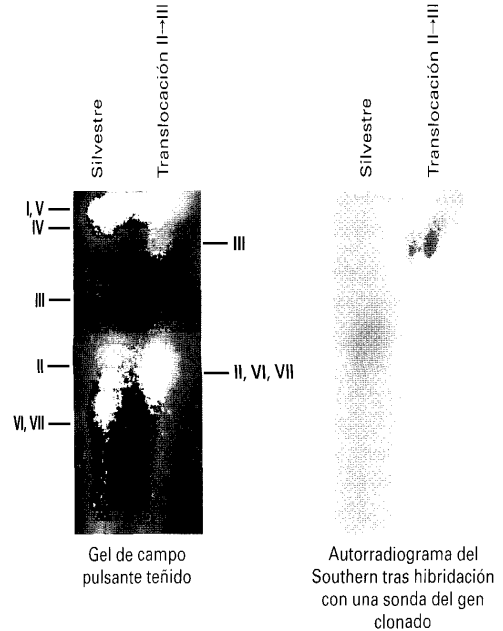


Figura 2.2 (a) Electroforesis en campo pulsante de DNA cromosómico no digerido de diferentes estirpes de levadura. Se distinguen 16 bandas. Al haber 22 cromosomas en levadura, algunos de ellos han debido migrar juntos en el gel. (Todas los canales se cargaron de forma idéntica), (b) La electroforesis en campo pulsante separa los cromosomas de *Neurospora* y facilita la localización de un gen clonado en un segmento cromosómico: (a la izquierda) se observan los siete cromosomas silvestres y una translocación en la que una porción de 100 kb del cromosoma II se ha insertado en el cromosoma III; (a la derecha) autorradiograma realizado después de hibridar con una sonda de un gen clonado de localización desconocida. Los resultados demuestran que el gen está situado en la región translocada del cromosoma II. (Parte a tomada de Bio-Rad Laboratories; parte b tomada de Myron Smith.)



HIBRIDOS DE CELULAS SOMÁTICAS HOMBRE-RATON

La técnica de hibridación de células somáticas ha sido ampliamente utilizada en la cartografía genética humana, aunque se puede emplear en principio en muchos sistemas animales distintos.

MAPAS CROMOSOMICOS DE ALTA RESOLUCIÓN

El siguiente nivel de resolución consiste en determinar la posición de gen o marcador molecular en el cromosoma. Este paso es importante, ya que los mapas genéticos que se generen se alinearán con los mapas físicos que veremos a continuación y se utilizarán para verificar dichos mapas físicos. Además, los clones producidos como parte del mapa físico pueden ayudar a identificar el ADN genómico correspondiente a los genes del mapa genético. Se utilizan varios métodos diferentes para localizar genes y marcadores.

CARTOGRAFÍA POR RECOMBINACIÓN MEIÓTICA

El análisis de las frecuencias de recombinación en cruzamientos dihíbridos y multihíbridos. En organismos experimentales, como la levadura, Neurospora, Drosophila y Arabidopsis, los genes que determinan diferencias en características fenotípicas cualitativas se pueden cartografiar en forma directa, gracias a la facilidad con que se pueden realizar cruzamientos controlados (como los cruzamientos de prueba). Por lo tanto, se dispone de estos organismos con mapas cromosómicos elaborados a lo largo de los años, que están llenos de genes de efecto fenotípico conocido, todos ellos asignados a sus respectivos loci.

Por varias razones éste no es el caso de la especie humana. En primer lugar, hay una carencia de cruzamientos informativos. En segundo lugar, los tamaños de las muestras

de los descendientes son demasiados pequeños como para hacer análisis estadísticos fiables que permiten determinar la existencia de ligamiento. En tercer lugar, el genoma humano es enorme. De hecho, incluso la asignación de un gen responsable de una enfermedad humana a un autosoma concreto por análisis de ligamiento constituyó una tarea difícil (la mayoría de los genes con fenotipos conocidos se asignaron mediante cartografía de híbridos de células somáticas hombre-ratón y no por análisis de la RF).

Incluso en los organismos cuyos mapas aparecen “llenos” de loci de efectos fenotípicos conocidos, se demostró que los intervalos cromosómicos entre los genes conocidos debían contener cantidades enormes de ADN. Estos intervalos, o “huecos”, no se podían cartografiar por análisis de ligamientos, puesto que no había marcadores en dichas regiones.

Para la obtención de mapas de mayor resolución, era preciso conseguir un gran número de marcadores genéticos adicionales que se pudieran utilizar para rellenar los huecos. Esto se resolvió con el descubrimiento de varias clases de marcadores moleculares. Un marcador molecular es un sitio en el que hay heterocigosis para algún tipo de variación neutra de ADN. La variación neutra es aquella que no está asociada a ninguna variación fenotípica observable.

Cuando está en heterocigosis, “este locus de DNA” se puede utilizar en cartografía genética como cualquier par de alelos convencionales en heterocigosis. Dado que los marcadores moleculares pueden detectarse fácilmente y son tan numerosos en los genomas, una vez cartografiados mediante análisis de ligamiento, rellenan los huecos entre los genes de fenotipo conocido.



Observe que, en la cartografía, el significado biológico del marcador no es importante en sí mismo; el sitio heterocigótico es simplemente un punto de referencia convencional cuya utilidad es la de permitir “orientarse” en el genoma.

En este sentido, los marcadores se utilizan como los puntos estratégicos utilizados por los viajeros en los siglos anteriores. Viajeros que no estaban interesados en las propias señales (marcadores) y, sin embargo, se habrían desorientados sin ellos.

Polimorfismo en la longitud de los fragmentos de restricción. Los RFLP fueron los primeros marcadores neutros del ADN que se aplicaron en la cartografía genómica por frecuencia de recombinación.

Marcadores de ADN basados en la variabilidad y en el número de repeticiones de secuencias cortas. Aunque los RFLP fueron los primeros marcadores de ADN en utilizarse en forma generalizada en la caracterización genómica, tanto en los análisis de los genomas de los animales como en las plantas, hoy día han sido reemplazados por marcadores que se basan en la variación en el número de repeticiones en tándem de secuencias cortas. Estos marcadores reciben colectivamente la denominación de polimorfismos en la longitud de secuencias sencillas (SSLP) que ofrecen ventajas básicas sobre los RFLP.

Primero, respecto a los RFLP, usualmente aparecen sólo uno o dos “alelos”, o morfos, en un pedigrí o una población en estudio. Esto limita su utilidad; sería mejor contar con gran número de alelos que servirían como marcas específicas de una gran variedad de regiones cromosómicas homologas.

Los SSLP cubren esta necesidad, dado que el alelismo múltiple es mucho más común habiéndose encontrado hasta 15 alelos de un locus. Segundo, el grado de heterocigosis de los RFLP puede ser bajo; en otras palabras si un alelo de un locus es relativamente raro en relación con el otro, la proporción de heterocigotos (los individuos cruciales en la cartografía) serán bajos.

Los SSLP, sin embargo, además de ofrecer más alelos muestran niveles de heterocigosis mucho mayores, lo que los hace más útiles en la cartografía genética que los RFLP, puesto que los heterocigotos son la base del análisis por recombinación. En la actualidad, se utilizan rutinariamente dos tipos de SSLP en Genómica.

1.- *Marcadores minisatélites*. Estos marcadores se basan en la variación en el número de repeticiones en tándem de los minisatélites (VNTR Repeticiones en Tándem de Número Variable). Una huella digital de ADN es un conjunto de bandas que resulta de la hibridación Southern de una digestión con enzimas de restricción (fig. 2.3) las bandas individuales de la huella digital del ADN representan secuencias de diferentes tamaños, de muchas posiciones cromosómicas distintas. Si los parentales difieren en una banda concreta, esta diferencia se convierte en un locus heterocigótico (mas/menos) que se puede utilizar en cartografía.

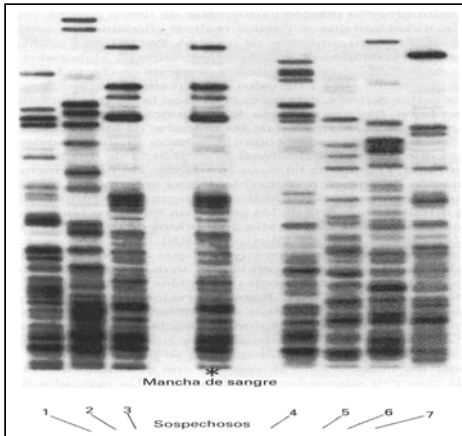
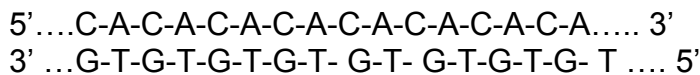


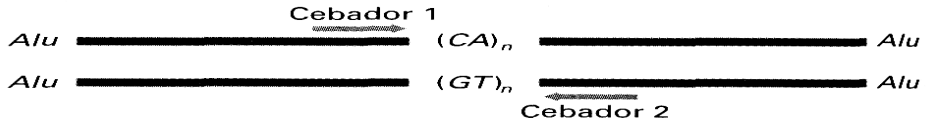
Figura 2.3 Huellas digitales de DNA de una mancha de sangre encontrada en la escena de un crimen y de la sangre de siete sospechosos.

(Cellmark Diagnostics, Germantown, MD.)

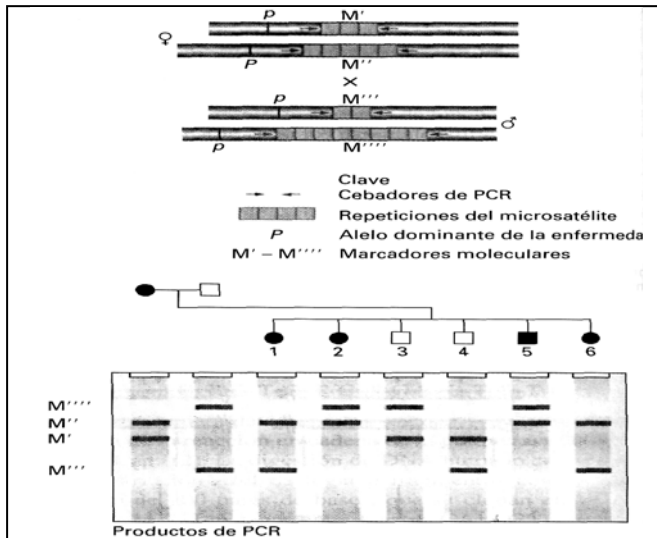
2.-*Marcadores microsátélites*. Recuerde que ADN microsátélite es una clase de ADN repetido que se basa en repeticiones de dinucleótidos; el tipo mas común es la repetición del dinucleótido C-A y su complementario G-T como en el ejemplo, siguiente:



Las sondas para detectar estos segmentos se generan con la ayuda de la reacción en cadena de la polimerasa. La digestión de ADN humano con la enzima de restricción AluI da lugar a fragmentos con un tamaño medio de 400 pares de bases que se clonan en un vector fágico M13. Los fagos con insertos $(CA)_n/(GT)_n$ se identifican mediante hibridación con una sonda $(CA)_n/(GT)_n$. Se secuencian los clones positivos y se diseñan las parejas de cebadores en base a las secuencias que flanquean el tramo repetido.



Los cebadores se emplean para la reacción de amplificación, utilizando ADN genómico como sustrato. Una pareja de cebadores concreta amplificará su propio tramo repetido y cualquier variante en el tamaño del mismo que aparezca en el ADN de distintos individuos. Una elevada proporción de las parejas de las parejas de cebadores revela la presencia de al menos tres alelos marcadores, que se distinguen por las diferencias en los tamaños de los productos de amplificación. La figura 2.4 muestra un ejemplo de la técnica de cartografía con microsatélites. Pueden fabricarse miles de parejas de cebadores que igualmente detectan miles de loci marcadores.





Hay que tomar en cuenta algunas diferencias en cuanto a las ventajas de los análisis de RFLP y SSLP. El análisis de RFLP requiere tener a mano en el laboratorio una sonda completa clonada para la detección de cada locus marcador individual.

El análisis de microsatélite requiere una pareja de cebadores para cada locus marcador, pero estas secuencias cebadoras pueden ser fácilmente compartidas por todo el mundo, distribuidas por correo electrónico y fabricadas rápidamente por un sintetizador de ADN. El análisis de minisatélites requiere solo una sonda que detecte la secuencia central del elemento repetido en cualquiera de los loci distribuidos por todo el genoma.

El descubrimiento de los marcadores de RFLP y SSLP ha permitido la construcción de un mapa genético humano con una resolución de 1 centimorgan [cM o unidad de mapa (u.m.)] Aunque esta resolución constituye un logro notable, un cM es todavía un tramo enorme de ADN, que en la especie humana se estima que equivale a una megabase (1 Mb = 1 millón de pares de bases).

En la actualidad, se están elaborando mapas genéticos con una resolución mayor, que se basan en los polimorfismos de un solo nucleótido (SNP; Polimorfismo de Un Nucleótido). Un SNP es un sitio del genoma formado por un único par de bases para el que, en las poblaciones naturales aparecen normalmente más de uno de los cuatro posibles pares de bases. Se han identificado y cartografiado en el genoma más de varios cientos de miles de sitios SNP, lo que proporciona el mapa con mayor densidad de marcadores.

DNA polimórficos amplificados al azar (RAPD). Un único cebador para PCR diseñado al azar amplificará, a menudo, por casualidad varias regiones distintas del genoma. La secuencia única se encuentra flanqueada por dos copias invertidas de la secuencia del cebador, el resultado es un conjunto de bandas amplificadas de ADN con diferentes tamaños (Fig. 2.5). En un cruzamiento, algunas de las bandas amplificadas pueden ser únicas de un parental, en cuyo caso pueden tratarse como loci heterocigóticos (+/-) y utilizarse como marcadores moleculares en el análisis cartográfico.

Observe también que el conjunto de fragmentos de ADN amplificados, llamados ADN polimórficos amplificados al azar (RAPD), ofrece otro tipo de huella digital de ADN que puede utilizarse para caracterizar un organismo individual. Tales marcas de identidad pueden ser muy útiles para el análisis genético rutinario o en los estudios de las poblaciones.

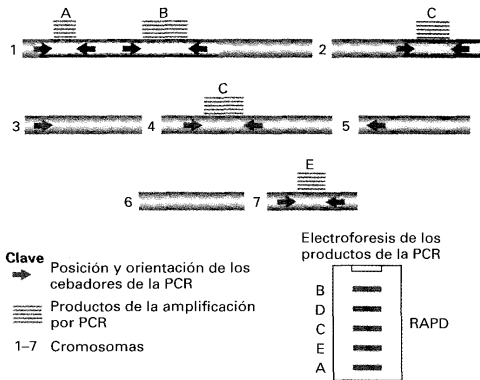


Figura 2.5 (a) El análisis de DNA amplificado al azar (análisis de RAPD) proporciona marcadores moleculares cromosómicos. Si otra estirpe carece de una de las bandas, se puede considerar que tal banda pertenece a un locus marcador heterocigótico para su uso en la cartografía, como se muestra en el ejemplo de la parte b.

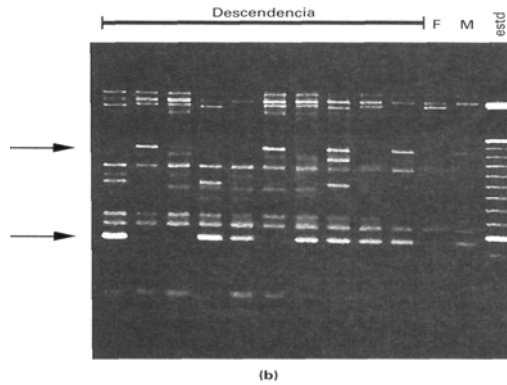


Figura 2.5 (b) Análisis de RAPD en un cruzamiento de una especie de árbol. Se utilizó un cebador de 10 nucleótidos para amplificar regiones de DNA genómico en los árboles parentales (M = masculino, F = femenino) y de 10 descendientes. En la columna de la derecha, marcada «estd», aparecen muestras de DNA estándar para calibrar los tamaños. Las flechas apuntan a dos bandas que representan dos loci de un cruzamiento de prueba «dihíbrido». Los dos alelos de ambos loci se manifiestan como presencia o ausencia de una banda de RAPD. De los dos parentales, únicamente el masculino mostró estas bandas, pero debía ser heterocigótico para la presencia (alelo +) o ausencia (alelo -) de las bandas de ambos loci. El parental masculino podría designarse $1^+/1^- \cdot 2V2^-$, y el femenino $1^-/1^- \cdot 2^-/2^-$. Los descendientes muestran varias combinaciones parentales y no parentales de estos alelos. (John E. Carlson)

Hibridación *in situ*. Un gen clonado puede utilizarse para fabricar una sonda marcada con la que poder hibridar cromosomas *in situ*. Si se puede reconocer cada cromosoma de una dotación genómica por su patrón de bandas, tamaño, relación de tamaños entre los brazos u otra característica citológica, entonces podemos asignar el nuevo gen al cromosoma con el que hibrida. Además, el sitio de hibridación revela la posición cromosómica aproximada del gen.

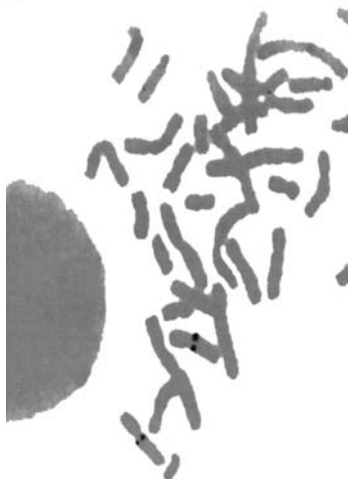


Figura 2.6 Análisis FISH. Los cromosomas se han hibridado *in situ* con una sonda fluorescente específica de un gen presente en una sola copia por dotación cromosómica, en este caso, el gen de una proteína del músculo. Sólo un locus muestra una mancha fluorescente, que corresponde a la sonda unida al gen de la proteína del músculo. (Tomado de P. Lichter et al., «High-Resolution Mapping of Human Chromosome 11 by *in Situ* Hybridization with Cosmid Clones», *Science* 247, 1990, 64.)

La sonda se marca comúnmente con radiactividad o fluorescencia. En el procedimiento de hibridación *in situ* fluorescente (FISH), el clon se marca con un compuesto fluorescente, y se baña con el una preparación cromosómica parcialmente desnaturizada. La sonda se une al cromosoma *in situ*, y la localización del fragmento clonado se reconoce por una mancha brillante de fluorescencia (Fig. 2.6). El coloreado de cromosomas es una extensión de la técnica FISH. En este caso un conjunto de fragmentos de ADN clonados, que sabemos pertenecen a cromosomas o regiones cromosómicas concretas, se marcan con compuestos fluorescentes distintos.



Estos compuestos “colorean” así regiones específicas y permiten distinguirlas al microscopio. Si un clon de un gen cuya posición se desconoce se marca con otro compuesto fluorescente se puede establecer su posición en la serie de cromosomas coloreados.

Puntos de ruptura y reorganización cromosómicas. Normalmente las mutaciones cromosómicas se atribuyen a que las rupturas cromosómicas rompen los genes en dos interrumpiendo secuencias estructurales o reguladoras esenciales. Si el punto de ruptura se puede ver o cartografiar con marcadores conocidos, aplicando el análisis de recombinación, esta información se puede emplear para asignar un gen a una posición en el mapa citogenético de un cromosoma. Una característica útil de los puntos de ruptura de las reorganizaciones es que también sirven como marcadores moleculares. Cuando se ha identificado un clon de ADN que cubre el punto de ruptura este se detecta en un experimento de Southern como una banda esperada que desaparece para dar lugar a la aparición de dos nuevas bandas.

Cartografía con híbridos irradiados. La técnica que se utiliza para localizar genes en cromosomas individuales se puede extender a la obtención de un mapa genético.

Una extensión importante es la cartografía con híbridos irradiados. Esta técnica se diseñó para generar un mapa de alta resolución de marcadores moleculares distribuidos por todo el cromosoma. El procedimiento consiste en irradiar células humanas con rayos X para fragmentar los cromosomas y, a continuación, fusionar las células irradiadas con células de ratón para formar un panel de híbridos distintos.

En este caso, los híbridos tienen un conjunto de fragmentos de los cromosomas humanos, como se representa en la Figura 2.7. La mayoría de los fragmentos están embebidos en los cromosomas de ratón, pero se pueden encontrar también cromosomas humanos truncados. Primero se calcula la frecuencia de aparición de varios marcadores moleculares humanos en los híbridos. El paso siguiente es calcular la frecuencia con que aparecen simultáneamente parejas de marcadores moleculares humanos. Hay que asumir que los marcadores estrechamente ligados entre si se incorporaran juntos con mucha frecuencia, ya que la probabilidad de que ocurra una rotura inducida por la radiación entre los dos loci es baja. La aparición simultanea de marcadores alejados entre si, así como de los que están en cromosomas diferentes, debería suceder con una frecuencia cercana al producto de las frecuencias individuales. Se utiliza una unidad de cartografía, cR_{3000} , que se ha calibrado a un valor aproximado de 0.1 cM(u.m.).

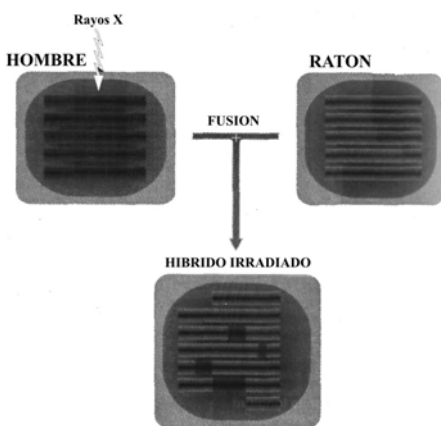


Figura 2.7 Generación de híbridos irradiados con rayos X. Los fragmentos de los cromosomas humanos se integran en los del ratón. El análisis de la presencia simultánea de marcadores humanos en un panel de varios híbridos irradiados puede desvelar la existencia de ligamiento.



La obtención de un panel estándar con unos 100 a 200 híbridos es bastante sencilla. Este panel sería suficiente para obtener un mapa cR_{3000} de alta resolución del genoma humano, que tendría una resolución diez veces mayor que el mapa genético actual en centimorgans. Una desventaja de la técnica es que esta limitada a los marcadores para los que existen diferencias entre el genoma humano y el del ratón.

Cartografía física de los genomas. Un incremento adicional en el grado de resolución cartográfica se consigue mediante la manipulación directa de fragmentos de DNA clonados. Como el DNA es la materia física del genoma, los procedimientos empleados se denominan, de forma general, cartografía física.

Un objetivo de la cartografía física es generar un conjunto de clones con fragmentos solapados, que cubran entre todos un cromosoma entero o un genoma completo.

El mapa físico resultante presenta tres características útiles. En primer lugar, permite ordenar los marcadores genéticos presentes en los clones, contribuyendo así al proceso general de cartografía del genoma.

En segundo lugar, una vez obtenidos, los clones continuos representan una serie ordenada de secuencias de DNA que se puede explotar en futuros análisis genéticos, como la búsqueda de correlaciones entre fenotipos mutantes y alteraciones en determinadas regiones moleculares. En tercer lugar, estos clones constituyen la materia prima para la obtención de la secuencia de nucleótidos en los proyectos genoma a gran escala.

En la preparación de los mapas físicos de los genomas, los vectores que pueden incorporar insertos muy grandes son, na-

turalmente, los más útiles. Los más utilizados son los cósmidos, YAC (cromosomas artificiales de levadura; del inglés, *yeast artificial chromosomes*), BAC (cromosomas artificiales de bacterias; del inglés, *bacterial artificial chromosomes*) y PAC (cromosomas artificiales derivados del fago λ ; del inglés, *λ -based artificial chromosomes*). Los BAC (Fig. 2.8) se basan en el plásmido F de 7 kb de *E. coli*. Recuerde que este plásmido puede llevar grandes fragmentos de DNA de *E. coli* como derivado F'. De forma similar, como vector de clonación, puede llevar también fragmentos de DNA foráneo de gran longitud, hasta de 300 kb, aunque el tamaño medio está alrededor de 100 kb. Los PAC se construyen mediante una manipulación similar del fago λ , y llevan insertos comparables a los de los BAC.

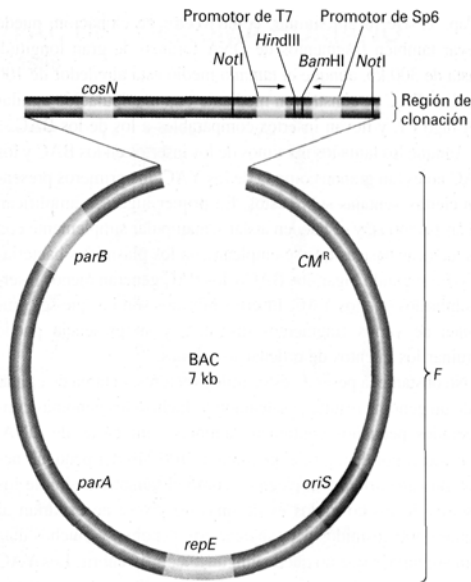


Figura 2.8 Estructura de un cromosoma artificial bacteriano (BAC), empleado para clonar fragmentos largos de DNA foráneo. CM^R es un marcador de selección de resistencia a cloranfenicol. *oriS*, *repE*, *parA* y *parB* son genes del plásmido F necesarios para su replicación y la regulación del número de copias. *cosN* es el sitio eos del fago λ . *HindIII* y *BamHI* son sitios de clonación en los que se inserta el DNA foráneo. Los promotores de T7 y Sp6 permiten transcribir el fragmento insertado. Los sitios *NotI* se emplean para retirar el fragmento insertado.



Aunque los tamaños máximos de los insertos en los BAC y los PAC no es tan grande como el de los YAC, los primeros presentan ciertas ventajas sobre éstos. En primer lugar, se amplifican en las bacterias, y se pueden aislar y manipular simplemente con las técnicas básicas que se emplean con los plásmidos bacterianos. En segundo lugar, los BAC y los PAC generan menos insertos híbridos que los YAC. Insertos híbridos son los que se componen de varios fragmentos distintos, y su presencia puede arruinar los intentos de ordenar los clones.

No obstante, a pesar de estos útiles vectores, la tarea de clonar todo un genoma resulta desalentadora. Incluso los genomas considerados pequeños contienen cantidades inmensas de DNA. Considere, por ejemplo, el genoma de 100 Mb del pequeño nematodo *Caenorhabditis elegans*, como el tamaño medio de los insertos de los cósmidos es de unas 40 kb, se necesitarían al menos 2500 cósmidos para abarcar este genoma, y muchos más para asegurarse que no quede ningún hueco sin cubrir. Los YAC pueden llevar fragmentos del orden de 1 Mb, así que con ellos la tarea resulta algo más sencilla.

La clonación de un genoma completo comienza con la obtención de un gran número de insertos clonados aleatoriamente.

El contenido de los clones se debe caracterizar de alguna forma que permita establecer los solapamientos. Un conjunto de clones solapados se denomina “*contig*”. En las fases iniciales de los proyectos genoma, los “*contig*” son numerosos y representan «islas» clonadas del genoma. Sin embargo, conforme se caracterizan más clones, los “*contigs*” se hacen más largos y se solapan con otros, hasta que el proyecto acaba generando un “*contig*” para cada cromosoma.

Genotecas específicas de cromosomas. Si la genoteca de clones se prepara a partir de DNA genómico total, la elaboración de los “*contigs*” es muy lenta. Sin embargo, si se utiliza un cromosoma concreto para hacer la genoteca, los “*contigs*” se elaboran más rápidamente. Puede emplearse la PFGE para aislar cromosomas individuales (si son pequeños) o segmentos cromosómicos resultantes del corte con enzimas de restricción que generan fragmentos muy largos, como *NotI*.

Otra opción para preparar DNA de un cromosoma concreto es la citometría de flujo. Los cromosomas (como los humanos) se pueden separar por citometría de flujo mediante la técnica de separación de cromosomas activada por fluorescencia (FACS; del inglés, *Fluorescence-Activated Chromosome Sorting* (Fig.2.9). En esta técnica, los cromosomas metafásicos se tiñen con dos colorantes, uno que se une a regiones ricas en AT y otro que lo hace a regiones ricas en GC. Se rompen las células para liberar los cromosomas a una suspensión líquida.

Esta suspensión se pulveriza de forma que la concentración de cromosomas es tal que cada gota contiene un solo cromosoma.

La suspensión pulverizada se pasa a través de un rayo láser calibrado para excitar la fluorescencia. Cada cromosoma produce su propia señal fluorescente característica, que es reconocida electrónicamente para que dos placas deflectoras dirijan las gotas que contienen el cromosoma deseado a un tubo colector.

Se utilizan varias técnicas distintas para ordenar los clones genómicos en “*contigs*”. A continuación, trataremos algunas de las más importantes.



Ordenamiento mediante FISH. Si se dispone de buenas marcas cromosómicas, se puede utilizar el análisis FISH para localizar las posiciones aproximadas de los insertos grandes.

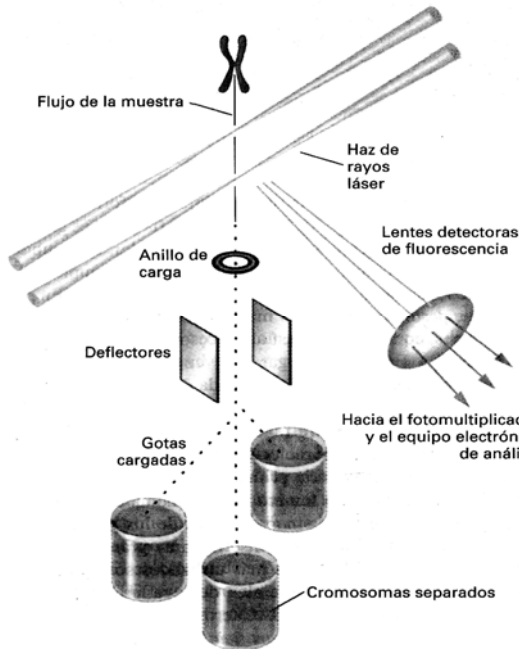


Figura 2.9 Purificación de cromosomas mediante citometría de flujo. Los cromosomas se unen con un compuesto fluorescente y se pasan a través de un haz de rayos láser. Se mide la intensidad de la fluorescencia de cada cromosoma que pasa y, de acuerdo con ella, es desviado al tubo colector correspondiente. Los cromosomas se recogen en pequeñas gotas.

El fragmento genómico insertado en un vector posee su propia secuencia única, que se puede utilizar para producir una huella digital de DNA. Por ejemplo, una digestión múltiple con enzimas de restricción genera una serie de bandas cuyo número y posición constituye una huella digital única, que es específica para ese clon.

Las distintas bandas producidas por clones diferentes se pueden alinear, ya sea visualmente o empleando un programa de ordenador, para determinar si hay algún solapamiento entre los DNA insertados. De esta forma, se puede ensamblar el "contig".



Ordenamiento mediante sitios marcados por su secuencia. Se pueden utilizar secuencias únicas de pequeño tamaño, contenidas en insertos grandes, como marcas para alinear varios clones en “*contigs*”. Por ejemplo, si un clon A tiene las marcas 1 y 2, y el clon B tiene las marcas 2 y 3, los dos clones deben solapar en la región de la marca 2. En la práctica, se acumula un conjunto numeroso de clones al azar con pequeños insertos genómicos (por ejemplo, clones en el fago λ) y se secuencian pequeñas regiones de cada clon.

A partir de estas secuencias, se diseñan cebadores de PCR que amplificarán pequeñas secuencias concretas del DNA flanqueado por los cebadores. Estas secuencias cortas de DNA se conocen como sitios marcados por su secuencia (STS; del inglés, *sequence-tagged sites*).

Aunque se desconoce inicialmente la posición de estos STS en el genoma, se puede utilizar un panel con muchos STS para caracterizar clones con grandes insertos genómicos (como los YAC). Los clones que tienen determinados STS en común deben presentar tramos que solapan y, por lo tanto, se pueden alinear en “*contigs*”. La Figura 2.10 muestra un ejemplo de este proceso.

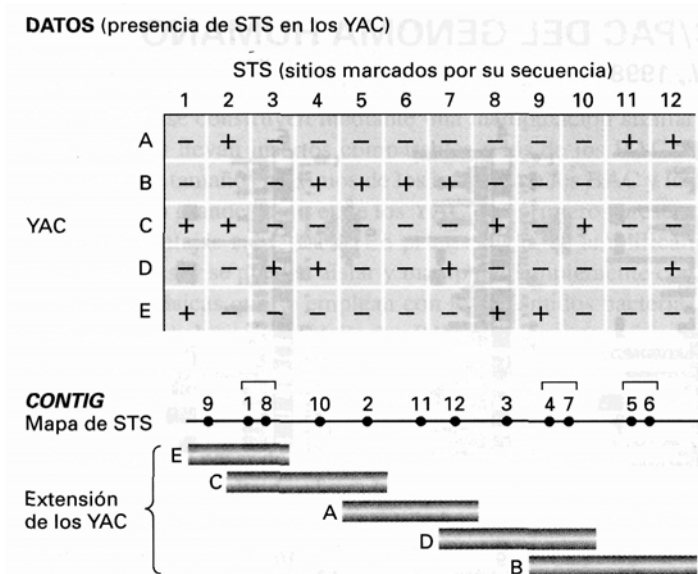


Figura 2.10 Utilización de sitios marcados por su secuencia (STS) para ordenar clones solapados (en este ejemplo, YAC) en un **contig**. Se analiza la presencia de los STS en cinco YAC distintos (arriba), y se emplean estos datos para generar un mapa físico (abajo).

Estas secuencias cortas se obtienen a veces de clones de cDNA, recibiendo entonces la denominación de marcadores de secuencias expresadas (EST; del inglés, *expressed sequence tags*). Los EST se obtienen mediante la secuenciación del inserto de cDNA, empleando para ello un cebador diseñado a partir de la secuencia del vector.

Se pueden utilizar para alinear los EST en los “*contigs*”, superponiendo así el mapa genético al mapa físico. Además, si en el EST está parte de la secuencia de lectura abierta (ORF; del inglés, *open reading frame*) del transcrito, la traducción «virtual» del ORF puede suministrar un «anticipo» de la función



de la proteína cifrada por el mRNA a partir del cual se obtuvo el cDNA.

La combinación de estos métodos de cartografía física ha derivado en la clonación de los genomas completos de varios organismos. Por ejemplo, el genoma de *C. elegans* está ahora disponible en series de “contigs” de cósmidos o de YAC.

Además, el DNA de los “contigs” se ha dispuesto de forma ordenada sobre filtros de nitrocelulosa; de manera que, para saber dónde se localiza en el genoma un fragmento concreto de DNA que sea de interés, basta con utilizar ese DNA como sonda sobre los filtros, y la detección de una señal positiva de hibridación indicará su posición precisa.

Un ejemplo: clonación y cartografía del cromosoma Y humano. Varios de los cromosomas humanos más pequeños se han clonado por completo como series solapadas de YAC (“contigs” de YAC). Examinaremos la clonación del cromosoma

Y como ejemplo, ya que ilustra varias de las técnicas de cartografía física. De hecho, el mapa de STS del cromosoma Y se obtuvo mediante dos métodos distintos, el alineamiento de YAC y el análisis de deleciones.

Alineamiento de YAC. La citometría de flujo permitió la obtención de una muestra de cromosomas Y, a partir de la cual se generaron clones de λ . Utilizando los que no contenían DNA repetido, se diseñaron cebadores de STS. En total, se fabricaron 160 cebadores. Se obtuvo una genoteca de YAC con 10.368 clones en los que el tamaño medio de los insertos era de unas 650 kb. Con estos números, se estimó que cada punto del cromosoma

Y debía estar representado en la genoteca una media de cuatro veces. Los clones YAC se dividieron en 18 series de 576 YAC cada una, y se analizó cada serie con los cebadores STS. La subdivisión de las series positivas permitió asignar rápidamente cada STS particular a YAC concretos. Se determinó el contenido global de STS en cada YAC, y se establecieron los solapamientos entre los YAC de la misma forma que se muestra en el ejemplo general de la Figura 2.10.

Análisis de deleciones. Hay varios tipos de deleciones del cromosoma Y que ocurren de forma natural. Por ejemplo, algunos varones XX poseen fragmentos truncados del Y, mientras que algunas mujeres XY llevan deleciones de la región que contiene el gen de la masculinidad (determinante de testículos).

Estas deleciones del Y se mantuvieron en cultivos celulares y fueron la base del alineamiento de STS del cromosoma Y. Se probó cada deleción respecto de su contenido en STS. Dado que, de manera natural, las deleciones constituían una serie ordenada, se pudo utilizar su contenido en STS no sólo para elaborar un mapa de STS, sino también para cartografiar la extensión de las deleciones. Los mapas de STS generados mediante el alineamiento de YAC y mediante el análisis de deleciones resultaron idénticos.

Secuenciación de los genomas. Se han empleado con éxito varias estrategias distintas en los proyectos genoma. Sus ventajas y desventajas dependen del tamaño y la complejidad del genoma, siendo de particular importancia la frecuencia de DNA repetido.

Secuenciación aleatoria de clones. El primer genoma en clonarse fue el de la bacteria *Haemophilus influenzae*. El DNA



genómico se fragmentó mecánicamente y se utilizó para obtener una colección numerosa de clones aleatorios que se presumía debían solapar unos con otros de muchas formas. Se emplearon cebadores complementarios al DNA adyacente del vector para secuenciar regiones cortas de los extremos de los insertos de *Haemophilus clonados*. Luego, estas secuencias cortas se utilizaron (de manera parecida a los STS) para alinear los clones genómicos. Como se obtuvieron muchas secuencias aleatorias de pequeño tamaño, juntas abarcaban la mayor parte del genoma de *Haemophilus*.

Los huecos se rellenaron mediante «paseo con cebadores»; esto es, empleando el extremo de una secuencia clonada como cebador para secuenciar tramos adyacentes no clonados.

Secuenciación de clones ordenados. La mayoría de los programas de Secuenciación genómica comienzan con una serie ordenada de clones. Previamente, vimos cómo se generó una serie ordenada de clones YAC para el cromosoma Y, y para otros cromosomas humanos.

Sin embargo, los clones YAC no son adecuados para su Secuenciación directa. Los YAC se subclonan en una serie de BAC o PAC solapados que, de nuevo, se alinean en *contigs* mediante STS o sus huellas digitales de DNA. A continuación, se subclonan los clones BAC o PAC en insertos más pequeños, que serán los que se secuencien. A este nivel, ya no se ordenan los clones, sino que se secuencian al azar múltiples insertos, de manera que cada clon BAC o PAC se secuencie por completo hasta un total de cinco veces.

Secuenciación de clones desordenados. Una estrategia empleada actualmente consiste en secuenciar los dos bordes de

los fragmentos genómicos clonados con cebadores complementarios a los extremos del vector. Si la longitud de los tramos secuenciados y las de los fragmentos clonados son suficientemente grandes, se pueden reunir estas secuencias para formar largos tramos continuos de secuencia que pueden abarcar a los DNA repetidos contenidos en el genoma.

La ventaja de esta estrategia es que se elimina el tiempo y el esfuerzo del proceso de cartografía de los clones. Este tipo de secuenciación se está probando en la actualidad en el genoma de *Drosophila* y en el humano.

Automatización. Se pueden acelerar todos los pasos del análisis genómico mediante automatización. La preparación de los clones, el aislamiento de DNA, la electroforesis y los protocolos de secuenciación han sido todos ellos adaptados a máquinas.

Utilización de los mapas genómicos en el análisis genético. Los mapas genéticos y físicos son un magnífico punto de partida para varios tipos de análisis genético, como el aislamiento de genes (incluidos los implicados en enfermedades genéticas humanas) y los estudios de Genómica funcional.

Aislamiento de genes de enfermedades humanas mediante clonación posicional. Consideraremos, como ejemplo, los métodos empleados para identificar la secuencia genómica del gen de la fibrosis quística (CF). Cuando se aisló el gen, no se conocía el defecto bioquímico causante de la enfermedad, de manera que se trataba de un gen a la búsqueda de una función. El ligamiento a marcadores moleculares había permitido localizar el gen en el brazo largo del cromosoma 7, entre las bandas 7q22 y 7q31.1.



Los datos indicaban que el gen *CF* estaba dentro de esta región, flanqueado a un lado por el gen *met* (un protooncogén) y por el marcador molecular D788 al otro lado. No obstante, entre ambos marcadores hay 1.5 centimorgans (unidades de mapa) de DNA, un enorme tramo no cartografiado de 1.5 millones de bases. Se obtuvieron marcadores adicionales dentro de la región, utilizando sondas nuevas derivadas de una genoteca del cromosoma 7 elaborada empleando citometría de flujo.

Sin embargo, las dos técnicas claves que se utilizaron para recorrer grandes distancias genéticas fueron el paseo cromosómico y una técnica relacionada denominada salto cromosómico. La última técnica ofrece la posibilidad de saltar por encima de regiones inclonables de DNA, y genera marcas muy separadas a lo largo de la secuencia, que pueden emplearse como puntos de partida de paseos cromosómicos en ambas direcciones.

El salto cromosómico se ilustra en la Figura 2.11. En este método, se generan grandes fragmentos de DNA mediante la restricción parcial de la región en la que se piensa está el gen de interés. A continuación, se circulariza cada fragmento, quedando así unidos sus dos extremos entre sí. Se corta la zona de unión y se clona en un vector fágico que, junto con otras zonas de unión, constituyen una *genoteca de saltos*. Se utiliza una sonda del inicio del tramo de DNA en investigación para analizar la genoteca de saltos y encontrar el clon que contiene el extremo inicial. Cuando se encuentra este clon, se corta el otro extremo de la región de unión y se utiliza para inspeccionar de nuevo la genoteca y realizar un segundo salto.

A partir de cada punto de salto pueden hacerse paseos cromosómicos en ambas direcciones, en búsqueda de secuencias que parezcan genes.

Se obtuvo un mapa de restricción de toda la región que debía contener el gen *CF*, utilizando enzimas de corte poco frecuente, y los sitios de restricción se utilizaron para establecer la posición y orientación de las secuencias obtenidas en los saltos y paseos cromosómicos. Una vez obtenida la secuencia de un número de segmentos suficiente para cubrir zonas representativas de toda la región, comenzó la caza de cualquier gen situado dentro de la misma. Los genes se buscaron mediante varias técnicas. En primer lugar, sabiendo que los genes humanos están normalmente precedidos en su extremo 5' por ristas de citosinas y guaninas, denominadas *islas CpG*, se buscaron y encontraron varias de ellas.

En segundo lugar, se razonó que un gen debía mostrar homología con el DNA de otros animales, debido a la conservación evolutiva, de manera que las secuencias candidatas se emplearon como sondas en lo que se denominó *zooblots*, análisis de hibridación de las sondas con DNA genómico de varios animales distintos.

En tercer lugar, los genes deben tener señales apropiadas de inicio y fin de la traducción. En cuarto lugar, los genes se transcriben, por lo que deben detectarse los transcritos correspondientes. Finalmente, se encontró un gen candidato que se extendía a lo largo de 250 kb de la región. Algunos síntomas de la CF se manifiestan en las glándulas sudoríparas; así que se preparó cDNA de células cultivadas de dichas glándulas, detectándose un cDNA de 6500 nucleótidos que era homólogo del gen candidato. Cuando se secuenció el cDNA de



personas sanas y de enfermos de CF, se observó que el cDNA de éstos contenía una deleción de tres pares de bases que eliminaba una fenilalanina de la proteína.

Por lo tanto, lo más probable era que esta región determinara la función afectada en la enfermedad. Así pues, se había encontrado el gen *CF*. De su secuencia nucleotídica se infirió la de aminoácidos; y a partir de ésta se predijo la estructura tridimensional de la proteína. Ésta resultó ser similar en estructura a las proteínas transportadoras de iones de otros sistemas, lo cual sugería que la causa principal de la enfermedad era un defecto en el transporte. Cuando se empleó el gen silvestre para transformar líneas celulares mutantes de enfermos de CF, se restableció la función normal. Siendo este «rescate» fenotípico la confirmación definitiva de que la secuencia aislada correspondía en realidad al gen *CF*.

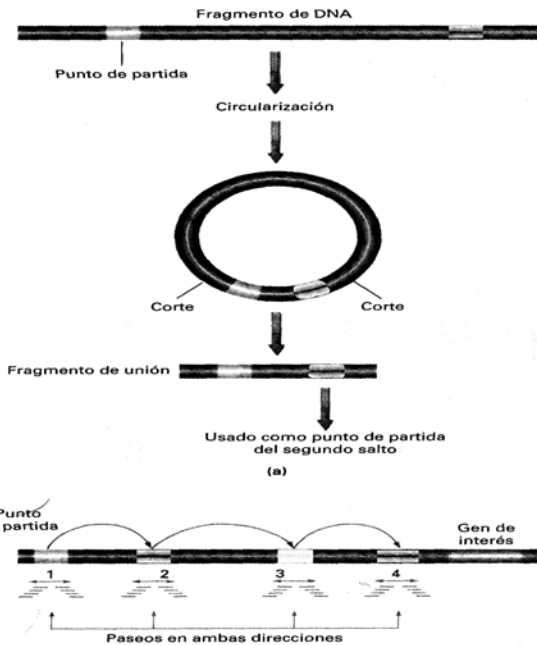


Figura 2.11 Manipulación de fragmentos genómicos clonados para realizar un salto cromosómico; esto permite evitar regiones de difícil clonación, como las que contienen DNA repetido.

Método del gen candidato. La clonación y caracterización intensiva de una región cromosómica revela inevitablemente la presencia de genes de función desconocida. Si un gen de interés, como el responsable de una enfermedad, se cartografía en dicha región, estas secuencias que parecen genes se convierten en genes candidatos a ser el implicado en la enfermedad. Esta estrategia para aislar un gen recibe la denominación de método del gen candidato. La información sobre datos fenotípicos, como un defecto bioquímico y el patrón de expresión tisular de la enfermedad, se puede confrontar con la presencia de ciertos dominios en la secuencia de la proteína y la expresión tisular del gen candidato.

GENÓMICA FUNCIONAL

Los datos de la secuenciación a gran escala son el comienzo de la Genómica funcional. Las siguientes secciones muestran algunos de los análisis que se pueden realizar para investigar la función.

Caracterización del proteoma mediante análisis de ORF. La secuencia de DNA genómico se analiza con programas de ordenador capaces de predecir la existencia de los genes.

Entre otras cosas, estos programas examinan cada una de las seis fases de lectura de todas las secuencias y buscan segmentos que comiencen con el codón AUG, de inicio de la traducción, y terminen con uno de los codones fin de mensaje. Cualquier secuencia de lectura abierta con al menos 100 codones es considerado como un posible gen. La mayoría de los ORF son completamente novedosos, sin que correspondan a algún gen conocido cuyos alélos generen fenotipos identificables. Inicialmente, la función de un ORF se puede



analizar empleando el ordenador para explorar las bases de datos en búsqueda de homología total o parcial con genes caracterizados en otros organismos.

La localización, orientación y agrupamiento de los ORF constituye también una información genómica de interés; de tales análisis se puede deducir una distribución provisional de los genes que determinan proteínas. En los eucariotas superiores, en los que los intrones son una característica común de los genes, la predicción de los ORF a partir del DNA genómico es más difícil.

Interrupción de genes: *knockouts*. La función de los ORF se puede investigar mediante la interrupción sistemática de los genes en los *knockouts*. Esto se consigue con mutagénesis *in vitro* y la búsqueda de cualquier fenotipo mutante que pudiera ofrecer pistas sobre la función.

Este procedimiento se está empleando en genomas secuenciados por completo. Resulta interesante señalar que, cuando se interrumpen, muchos ORF no muestran efectos fenotípicos. Más de la mitad de los ORF predichos se incluyen en esta categoría.

Estudio de interacciones génicas mediante el sistema del doble híbrido de levadura.

Este método investiga interacciones utilizando un sistema basado en dos plásmidos de levadura. La prueba se basa en el activador transcripcional GAL4 de levadura. Esta proteína consta de dos dominios, uno de unión al DNA y otro activador, que deben estar en estrecha conexión para que este activador inicie la transcripción.

En un plásmido que actúa como «cebo», se une el gen de la proteína en estudio al tramo de GAL4 que determina el dominio de unión al DNA. En otro plásmido, el gen de otra proteína en estudio se coloca junto al tramo de GAL4 que determina el dominio activador; esta proteína recibe la denominación de «presa» (Fig. 2.12). A continuación, los dos plásmidos se introducen en la misma célula. Una forma de conseguirlo es cruzando dos células haploides, una que contiene el cebo y la otra la presa. La única forma de que los dominios de unión a DNA y de activación entren en contacto es que las dos proteínas, cebo y presa, se unan entre sí, poniendo de manifiesto una interacción física. El sistema del doble híbrido puede automatizarse para facilitar la búsqueda a gran escala de interacciones entre proteínas de todo el proteoma.

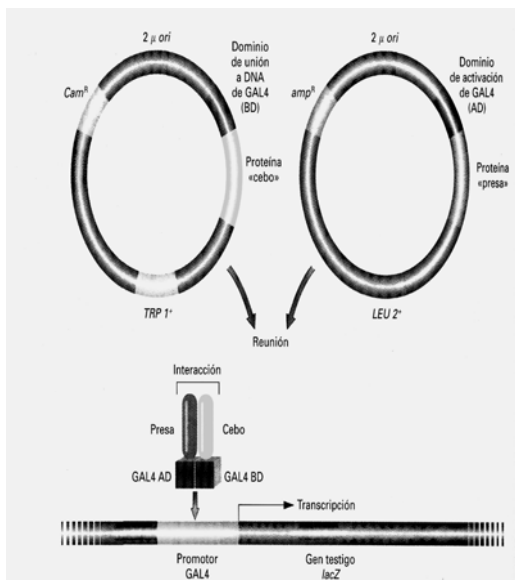


Figura 2.12 Sistema del doble híbrido de levadura para la detección de interacción entre genes. El sistema aprovecha la unión de dos proteínas para detectar la restauración de la actividad de la proteína GAL4, que activa un gen testigo.



EL GENOMA HUMANO

El Genoma Humano es el número total de cromosomas del cuerpo. Los cromosomas contienen aproximadamente 80.000 genes, los responsables de la herencia.

La información contenida en los genes ha sido decodificada y permite a la ciencia conocer mediante evaluaciones genéticas, qué enfermedades podrá sufrir una persona en su vida. También con ese conocimiento se podrán tratar enfermedades hasta ahora incurables. Pero el conocimiento del código de un genoma abre las puertas para nuevos conflictos ético-morales, por ejemplo, seleccionar que bebés van a nacer, o clonar seres por su perfección. Esto atentaría contra la diversidad biológica y reinstalaría entre otras la cultura de una raza superior, dejando marginados a los demás.

Quienes tengan desventaja genética quedarían excluidos de los trabajos, compañías de seguro, seguro social, etc. similar a la discriminación que existe en los trabajos con las mujeres respecto del embarazo y los hijos.

Un genoma es el número total de cromosomas, de un organismo, incluido sus genes, los cuales llevan la información para la elaboración de todas las proteínas requeridas por el organismo, y las que determinan el aspecto, el funcionamiento, el metabolismo, la resistencia a infecciones y otras enfermedades.

En otras palabras, es el código que hace que seamos como somos. Un gen es la unidad física, funcional y fundamental de la herencia. Es una secuencia de nucleótidos ordenada y ubicada en una posición especial de un cromosoma. Un gen contiene el código específico de un producto funcional. La

importancia de conocer acabadamente el genoma es que todas las enfermedades tienen un componente genético, tanto las hereditarias como las resultantes de respuestas corporales al medio ambiente.

El Proyecto Genoma Humano es una investigación internacional que busca seleccionar un modelo de organismo humano por medio del mapeo de la secuencia de su DNA. Se inició oficialmente en 1990 como un programa de quince años con el que se pretendía registrar los 80.000 genes que codifican la información necesaria para construir y mantener la vida. Los rápidos avances tecnológicos han acelerado los tiempos esperándose que se termine la investigación completa en el 2003.

Los objetivos del Proyecto son:

- Identificar los aproximadamente 100.000 genes humanos en el DNA.
- Determinar la secuencia de 3 billones de bases químicas que conforman el DNA.
- Acumular la información en bases de datos.
- Desarrollar de modo rápido y eficiente tecnologías de secuenciación.
- Desarrollar herramientas para análisis de datos.
- Dirigir las cuestiones éticas, legales y sociales que se derivan del proyecto.

Este proyecto ha suscitado análisis éticos, legales, sociales y humanos que han ido más allá de la investigación científica propiamente dicha. (Declaración sobre Dignidad y Genoma Humanos, UNESCO)



El propósito inicial fue el de dotar al mundo de herramientas trascendentales e innovadoras para el tratamiento y prevención de enfermedades. Como se expresó, el genoma es el conjunto de instrucciones completas para construir un organismo, humano o cualquiera. El genoma contiene el diseño de las estructuras celulares y las actividades de las células del organismo.

El núcleo de cada célula contiene el genoma que está conformado por 23 pares de cromosomas, los que a su vez contienen alrededor de 80.000 a 100.000 genes, los que están formados por 3 billones de pares de bases, cuya secuencia hace la diferencia entre los organismos.

El DNA que conforma el genoma, contiene toda la información necesaria para construir y mantener la vida desde una simple bacteria hasta el organismo humano. Comprender como el DNA realiza la función requiere de conocimiento de su estructura y organización.

El tamaño del genoma es usualmente basado en el total de pares de bases. En la especie humana, contiene aproximadamente 3 billones de pares de bases. Otros organismos estudiados con motivo de éste estudio fueron la bacteria *Escherichia coli*, la mosca de la fruta, y las ratas de laboratorio.

Todos los genes están dispuestos linealmente a lo largo de los cromosomas. EL núcleo de muchas células humanas contiene dos tipos de cromosomas, uno por cada padre. Cada set, tiene 23 cromosomas simples, 22 de tipo autosómico y uno que puede ser X o Y que es el cromosoma sexual. Una mujer normal tendrá un par de cromosomas X (XX), y un hombre normal tendrá un cromosoma X y otro Y (XY).

Los cromosomas contienen proximadamente igual cantidad de partes de proteína y DNA. El DNA cromosómico contiene un promedio de 150 millones de bases.

Los cromosomas pueden ser evidenciables mediante microscopio óptico y cuando son teñidos revelan patrones de luz y bandas oscuras con variaciones regionales. Las diferencias en tamaño y de patrón de bandas permite que se distinguan los 24 cromosomas uno de otro, el análisis se llama cariotipo.

Desde un punto de vista científico, el mapa del genoma humano es una herramienta genética que permite estudiar la evolución del hombre y que cambiará drásticamente la medicina actual tal como la conocemos. Permitirá el tratamiento de enfermedades hasta ahora sin cura. Las investigaciones estuvieron a cargo fundamentalmente de Estados Unidos (Instituto Nacional de Investigación del Genoma Humano-NHGRI- de Maryland) y Gran Bretaña (Centro Sanger en Cambridge), pero también acompañaron Francia, Alemania, Japón y China.

Hoy el mapa del genoma está casi completado. Se abre también el camino para la manipulación genética, motivo por el cual se han dictado documentos tendientes a acotar ese aspecto. La empresa privada Celera Genomics de Rockville (EEUU), es la que lidera los procesos. La investigación duró diez años e insumió cerca de 2.000 millones de costo.

La fiabilidad del mapa de 3.000 millones de pares de bases llegará a un 99,99%. Además se conocerá el número preciso de genes del organismo calculado entre 60.000 y 100.000.



Actualmente el 85% del genoma está detalladamente mapeado.

El mito del ser humano inmortal y perfecto se asocia a la aplicación práctica de los conocimientos del mapa del genoma humano. Como se puede apreciar, la búsqueda de la raza perfecta buscada hace años por Hitler resulta ser una aspiración de la raza humana ahora encarnada en el proyecto del genoma humano. Además, el conocimiento del genoma permitirá que se creen nuevas drogas terapéuticas que desplazarán a las anteriores en la medida que los presupuestos permitan comprarlas. De este modo se podrá polarizar la industria farmacéutica. Las nuevas drogas prometen tener menores efectos colaterales que las actuales.

Se le podrá informar a una persona, que puede comer alimentos grasos porque carece de predisposición genética a la obesidad y a enfermedades cardíacas, pero que debe huir del alcohol porque es genéticamente propenso al alcoholismo.

Además el grado de certidumbre que otorga el conocimiento del código genético resultaría más creíble para la persona en cuestión, ya que sabe que lo que se le informa será absolutamente cierto. Es una predicción absoluta, de su futuro.

Podríamos hablar de genomancia o sea la adivinación del futuro mediante el código genético; si una persona carece de un determinado tipo de célula que le produce una enfermedad, la misma se podrá cultivar y luego colocar al sujeto. Claro que esto debería en principio ser realizado periódicamente, ya que el sujeto carecería de la habilidad propia para restaurar la función.

Pero la terapia de línea germinal, apuntaría a solucionar ese inconveniente, ya que afectaría las futuras generaciones celulares. Esto es impredecible y éticamente intolerable, pero de no serlo o de permitirse se borrarían del planeta el síndrome de Down o el sida.

Hasta ahora, el médico ha tenido muy clara su tarea: devolver al paciente al estado natural de salud. Pero cuando pueda manipular el programa vital, ¿resistirá la tentación de mejorar el modelo? Dentro de los llamados beneficios anticipados del Proyecto figuran a nivel de Medicina molecular, la posibilidad de mejorar el diagnóstico de enfermedades, detección temprana de predisposiciones genéticas a ciertas enfermedades, el diseño racional de drogas, terapia génica, sistemas de control para drogas y farmacogenomas.

Se ha estudiado un gen que determina la producción de la proteína llamada SPARC, la que normalmente impide al organismo atacar y anular células cancerígenas. La terapia génica en estos casos actúa permitiendo que las células cancerosas sean atacadas por el organismo.

A nivel de genomas microbianos, sirvió para explorar nuevas fuentes de energía (bioenergía), monitoreo del medio ambiente para detección de poluciones, protección contra guerra Química y biológica y eficiente limpiado de residuos tóxicos.

También es útil para estimar el daño y riesgo por exposición a la radiación, agentes mutagénicos, toxinas cancerígenas y reducción de probabilidad de mutaciones hereditarias. La identificación de oncogenes (genes que permiten que un sujeto que se exponga a ciertas sustancias desarrolle un determinado tumor, ejemplo, quien posea el oncogen para el cáncer de



pulmón y fume cigarrillos desarrollará cáncer de pulmón a diferencia de quien no tenga dicho oncogen).

En bioarqueología, evolucionismo y migración humana tiene su utilidad en las mutaciones de linaje, migraciones de diferentes grupos poblacionales basados en el DNA mitocondrial, mutaciones del cromosoma Y, además de comparar los cambios evolutivos con eventos históricos.

En identificación forense, para potenciales sospechosos en los cuales el DNA puede conducir a liberar a personas que fueran acusadas de crímenes injustamente, para identificar víctimas de catástrofes, paternidad y otras relaciones familiares, identificar y proteger especies en peligro, detectar bacterias que pueden polucionar agua, aire, alimentos, determinar compatibilidad de órganos donantes en programas de trasplante, determinar el pedigree en ganados y para autenticar productos de consumo como caviar, vinos.

En agricultura, ganadería y bioprocesamientos, se utiliza para mejorar la resistencia de cultivos ante insectos, sequías, para hacerlos más productivos y saludables igualmente para producir animales más saludables y nutritivos, elaborar biopesticidas, vacunas comestibles y nueva limpieza del medio ambiente de plantas como tabaco.

Los problemas derivados de la investigación genética son la equidad en su uso por parte de aseguradoras, seguro social, escuelas, agencias de adopción, cumplimiento de la ley, instituciones militares. A quien pertenece la potestad del control? Otro problema es el impacto psicológico y la estigmatización debido a diferencias individuales y acerca de cómo influirá a la sociedad el determinismo genético. El

personal que cuida de la salud aconsejará a los padres acerca de los riesgos y limitaciones de la tecnología genética. Qué tan confiable será, además de útil, el testeo genético fetal?

Respecto de la terapia génica usada para tratar o curar trastornos genéticos plantea la pregunta acerca de qué es una discapacidad o trastorno y quién decide acerca del mismo. Las dishabilidades son enfermedades? Deben ser curadas o prevenidas? El mejoramiento génico incluye el uso de terapia genética para suplir características como la altura que un padre podría querer en sus hijos, pero que no significa la prevención de una enfermedad, sino la búsqueda de un ser perfecto acorde a un ideal.

Si esto se vuelve una práctica común, como podría afectar la diversidad genética? Finalmente, que consecuencias sociales traería a la humanidad?

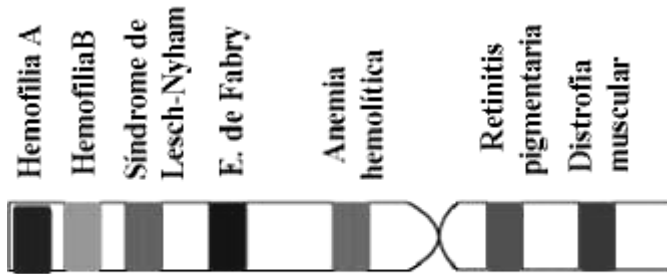
La equidad en el uso de las tecnologías génicas, plantea quién tendrá acceso a la misma y quien pagará por su uso. Los estudios clínicos incluyen educación de proveedores de servicios de salud, pacientes y público, acerca de cómo se implementarán los exámenes genéticos. Este proyecto supone la realización de dos tipos de mapas:

Mapas genéticos: Estos mapas simplemente indican la posición relativa de los diferentes genes. Para esta confección se están estudiando la transmisión de caracteres hereditarios, capaces de ser objetivados de una generación a otra en grandes familias.

Por ejemplo, en Estados Unidos se han localizado muchos genes gracias a estudios realizados en comunidades mormonas, cuya endogamia es notoria.



En 1994 se terminó el primer mapa genético de todo el genoma humano.



*Esquema del cromosoma X
y localización de enfermedades genéticas*

Mapas físicos: De mayor resolución, pues muestra la secuencia de nucleótidos en la molécula de ADN que constituye el cromosoma. Se obtiene la secuencia de nucleótidos de un gen. Se realiza fundamentalmente mediante la electroforesis en geles de distintos fragmentos de ADN y la ayuda de ordenadores, el completar este mapa se ha conseguido cinco años antes de lo que se esperaba.

COMO SE SECUENCIÓ EL GENOMA?

1.- Los cromosomas tienen un rango de tamaño comprendido entre 50 millones y 250 millones de bases; lo primero que se hace es romper o fragmentar los cromosomas en piezas más pequeñas.

2.- Cada pieza pequeña es usada como un templado para generar un grupo de fragmentos que difieren en longitud una del otro en una simple base que será identificado en un paso posterior (preparación del templado y pasos para la reacción de secuenciamiento)

- 3.- Los fragmentos son separados en grupos por gel de electroforesis (paso de separación)
- 4.- La base final del extremo de cada fragmento es identificada. Este proceso recrea la secuencia original As, Ts, Cs y Gs para cada pieza corta generada en el primer paso.

El límite en tamaño de la electroforesis está entre 500 a 700 pb por lectura secuenciada.

Secuenciadores automáticos analizan los resultados de los electroferogramas y el resultado es un cromatograma de cuatro colores que muestran picos que representan cada uno de las cuatro colores bases de ADN.

Después de que muchas bases son leídas; programas computacionales son usados para ensamblar secuencias cortas (en bloques de 500 bases) en longitudes continuas que se irán extendiendo; así se observarán los genes de regiones codificantes y otras características.

Obtenidas las secuencias son suministradas a las bases de datos como el GenBank

QUE SECUENCIAS DEL GENOMA HUMANO SERAN PUBLICADAS?

Por los números. El genoma humano contiene 3164.7 millones de bases nucleotídicas. El número total de genes estimado es de 30 a 35.000; mucho mas bajo de lo que se estimaba (80.000 – 140.000), que era una extrapolación de áreas ricas en genes en oposición a los compuestas por áreas pobres en genes.

Todas las personas tienen exactamente (99,9%) la misma secuencia de sus bases. Las funciones del 50% de los genes descubiertos son desconocidas. Menos del 2% del genoma



codifica proteínas y por lo menos un 50% del genoma humano esta constituido por secuencias repetidas que no codifican proteínas.

Métodos de secuenciamiento y programas para el análisis de datos

- Método de secuenciamiento Hierarchical Shotgun sequencing, clones (100- 200kb)
- Desarrollo tecnológico para el análisis de datos:
PHRED software package
PHRAP computer package

GENERACIÓN DE SECUENCIAS DEL GENOMA

La generación de una secuencia determinada del genoma involucra tres pasos:

- 1.- Selección de los clones BAC para ser secuenciados
- 2.- Secuenciamiento de los mismos
- 3.- Ensamblaje de la secuencia individual.

1.- SELECCIÓN DE LOS CLONES.

El método “hierarchical shorgun” involucra el secuenciamiento de clonación con insertos grandes de genoma. Para el proyecto genoma humano, los clones fueron escogidos de insertos grandes contenidos en una librería, presentes en cromosomas artificiales derivados de clones BAC o PAC. Las librerías fueron hechas por digestión parcial de ADN genómico con enzimas de restricción.

Las librerías fueron preparadas de ADN obtenido de donadores humanos anónimos en concordancia con “US Federal Regulations for the Protection of Human Subjects in Research (45CFR46)” Se escogieron muestras al azar para el procesamiento y se prepararon líneas celulares inmortalizadas, de las cuales se hacían las extracciones de ADN.

El ADN de cada clon (BAC) fue digerido con la enzima de restricción HindIII y el tamaño de los fragmentos resultantes fueron medidos por electroforesis en geles de agarosa. Los patrones de restricción generan un “fingerprint” (huella dactilar) para cada BAC los cuales permiten diferenciarlos y distinguir el grado de solapamiento con respecto a otros clones y así poder ensamblar los BACs (en fingerprint clone contig)

Los clones con secuencias “fingerprint contigs” fueron posicionadas a lo largo del cromosoma por la relación con marcadores STS de existencia genética y mapas físicos. Estos clones fueron unidos para ubicar específicamente los STSs por pruebas de hibridación y después por búsqueda directa de las secuencias de los clones. Para localizar “fingerprint clone contigs” que no contengan marcadores conocidos, se generaron nuevas “STSs” y fijadas en el cromosoma. Los clones representativos fueron posicionados por “FISH.”



Los solapamientos entre clones BAC proveen una rica colección de SNPs. Mas de 1.4 millones de SNPs han sido identificados por el solapamiento de clones y comparación con otras secuencias.

Debido a que el proyecto de secuenciamiento del genoma fue compartido por veinte centros en seis países, fue importante coordinar la selección de clones en cada uno de los centros.

La mayoría de los centros se enfocó sobre cromosomas particulares y el Internacional Human Genome Sequencing Consortium, mantuvo el registro de clones para seguir la pista de los clones seleccionados y sus progresos. En fases posteriores el mapa global provee una visión integrada de los datos de todos los centros, facilitando la distribución del esfuerzo para así, maximizar la cobertura de todo el genoma.

Antes de comenzar el secuenciamiento extensivo sobre un clon centros examinaron una muestra de 96 muestras de secuencias leídas de cada subclase de la librería para evaluar el posible solapamiento con clones secuenciados previamente.

2.- SECUENCIAMIENTO

Los clones seleccionados fueron sujetos al secuenciamiento por el sistema "Shotgun". La aproximación básica del secuenciamiento "Shotgun" estuvo bien establecida y los detalles de implementación variaron en los centros. Los centros difieren en el empleo del marcaje fluorescente empleados y el "cebador-coloreado" que usan.

La automatización también varía en cada centro, los cuales se esforzaron para generar un sistema que procesaron mas de 100.000 reacciones de secuenciamiento en 12 horas (Fig 3. pg.

867. Nature) en otras palabras los centros difieren en la cantidad de datos (“secuencias en bruto”) obtenidas por cada clon. La información de las secuencias de los diferentes centros pueden ser directamente integradas a pesar de tal diversidad; por esta razón los datos fueron analizadas y ensambladas con el PHRED y PHRAD software packages.

Todos los contig ensamblados de mas de 2 Kb fueron colocados en bases de datos públicas en el espacio de 24 horas después de ensambladas.

NOTA:

Los centros alcanzaron la velocidad de 1000 nucleótidos/seg, trabajando las 24 horas del día y los 7 días de la semana. Esto generó un incremento de las secuencias disponibles en las bases de datos públicas.

3.- ENSAMBLAJE DE LAS SECUENCIAS GENÓMICAS.

Los procesos de solapamiento tienden a resolver problemas generados con la naturaleza tan compleja de muchas secuencias; que van desde la gran diversidad de clones y la alta cantidad de secuencias repetidas en el genoma humano. Estos procesos involucran tres pasos: filtrar, seleccionar y unir.

Los datos obtenidos fueron filtrados uniformemente para eliminar contaminación de secuencias no humanas y otros artefactos que no han sido removidos por los centros individuales.

Los clones secuenciados fueron asociados con clones específicos sobre mapas físicos para producir un “Layout” (selección). En principio los clones secuenciados que corresponden a un “fingerprinted BAC” pueden ser



directamente asignados por un nombre para ese clon en particular sobre el “fingerprint” basado en el mapa físico. Sin embargo, en la práctica el laboratorio “mixups” ocasionalmente realizó asignaciones incorrectas. Para eliminar cualquier problema los clones secuenciados fueron asociados en el “fingerprint clone contigs” en el mapa físico por uso de la secuencia de datos para calcular una lista parcial de fragmentos de restricción “in silico” y comparando con la base de datos experimental de los BAC fingerprints. La comparación fue factible debido a que el tamaño experimental de los fragmentos de restricción fue altamente exacta (entre 0.5 – 1.5 % del tamaño real, para el 95% de los fragmentos de 6.000 a 12000 pb).

CONTENIDO DE GENES DEL GENOMA HUMANO.

Los genes (o regiones codificantes) comprenden solamente una fracción muy reducida del ADN humano, pero ellos representan la función biológica más importante del genoma y punto central del interés para los biólogos. Ellos son el mayor desafío–futuro para identificar la secuencia del genoma humano.

El último objetivo será obtener una lista completa de todos los genes humanos y sus proteínas codificantes, que servirán como una “tabla periódica” para las investigaciones biomédicas. Pero esto es un desafío muy difícil. En organismos con genoma pequeños; la identificación de mayor número de genes es más rápida por la presencia de gran número de ORFs. En contraste, los genes humanos tienen pequeños exones (algunos exceden 10 Kb); esto crea un problema (señal de ruido) en la predicción directa de los genes para los programas de computadoras limitando con esto la exactitud.

En cambio, la predicción computacional de genes humanos debe ser altamente confiable sobre la disponibilidad de secuencias de cDNA o sobre la conservación de secuencias con genes y proteínas de otros organismos. Esta aproximación es adecuada para genes fuertemente conservados (como histonas o ubiquitina), pero puede tener baja sensibilidad para genes que cambian rápidamente (incluyendo genes cruciales para la especiación, determinación sexual y fertilización).

ARNs NO CODIFICANTES.

Los biólogos a menudo hablan de una estrecha relación entre genes y su producto codificante (la proteína), esto es importante para recordar que gran cantidad de genes humanos producen ARNs no codificantes (ncRNAs) como su último producto. Son algunas clases de ncRNAs:

- ARN de transferencia (ARNt)
- ARN ribosomal (ARNr)
- ARN nucleolar pequeño (ARNsno)
- ARN nuclear pequeño (ARNsn)

Es muy común encontrar ncRNAs en el genoma, aún sabiendo la secuencia completa. Sin embargo, se pueden identificar secuencias genómicas que sean homólogos a genes conocidos (ncRNAs) usando el programa BLASTN o en algunos casos métodos más especializados. Para los ARNt hay suficiente información acerca de la estructura secundaria (de genes verdaderos y pseudogenes) detallada para ser distinguidos con alta sensibilidad.

Para otros ncRNAs, hay poca información estructural y se emplean criterios operacionales de alta similitud de



secuencias (> 95% de semejanza entre secuencias) para distinguir desde genes verdaderos a pseudogenes. Estos asignamientos son eventualmente necesarios para reconciliarse con los datos experimentales.

GENES QUE CODIFICAN PROTEÍNAS

La identificación de genes que codifican proteínas, en el genoma humano es una de las aplicaciones más importantes de los datos secuenciados, pero también uno de los retos más difíciles. Actualmente se están haciendo muchos esfuerzos para crear un “Index” de proteínas.

EXPLORANDO LAS PROPIEDADES DE LOS GENES CONOCIDOS.

Los alineamientos genómicos permiten un estudio de las estructuras intrón – exón y el contenido local GC, que son valorados para estudios biomédicos, a causa de la conexión de ellos con genes del mapa genético y citogenética; además de su relación con secuencias regulatorias.

Previos esfuerzos para estudiar la estructura de genes humanos han sido interrumpidos o bloqueados por el tamaño limitado de las muestras y las fuertes tendencias a favor de genes compactos.

Hay considerables variaciones en el tamaño de los genes y el de los intrones. Algunos genes tienen una longitud de 100 Kb; el ejemplo más largo conocido es el gen de “dystrophin” (DMD) de 2.4 MB. El gen “titin” tiene una secuencia codificando de 80.780 pb; también tiene el número de exones más grande y un exón simple puede alcanzar un tamaño de 17.106 pb.

Es instructivo comparar las propiedades de genes humanos con el de una oruga y una mosca. Para estos tres organismos la típica longitud de una secuencia codificante es parecida (1311 pb para la oruga, 1497 pb para la mosca y 1340 para el caso del humano) y los exones mas internos constan entre 50 a 200 pb (Fig. 2.13)

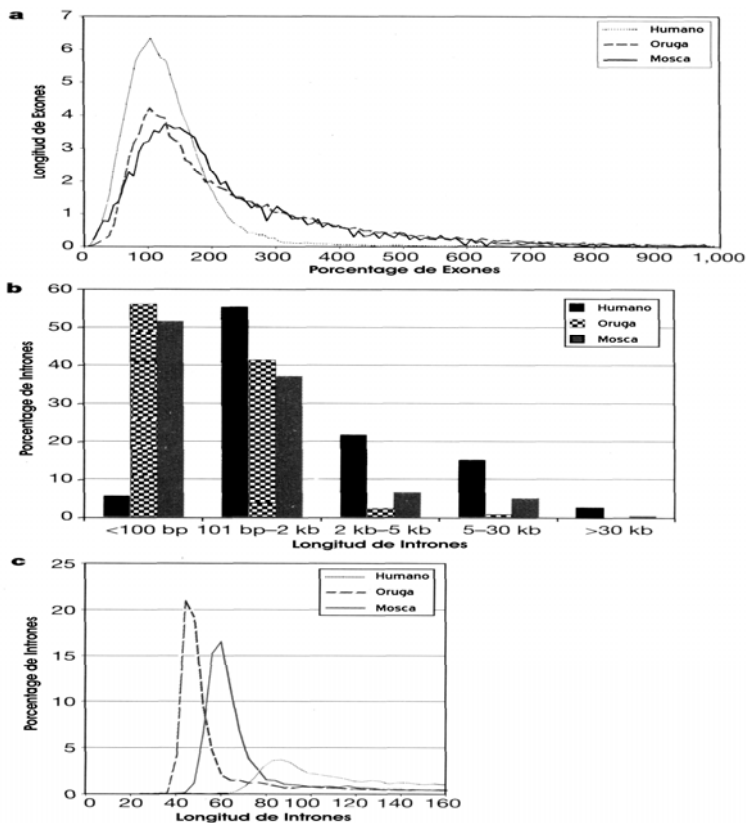


Figura 2.13 Distribución de los exones, intrones en genomas secuenciados. a) Exones, b) Intrones, c) Intrones cortos



La evidente conservación del tamaño de los exones entre estas tres especies sugiere una semejanza en la maquinaria de “splicing”. Las secuencias ricas en purinas pueden potenciar el splicing y es posible que estas secuencias sean requeridas para asegurar un “splicing” correcto de exones muy cortos (como los 42 exones humanos detectados que están constituidos por secuencias de 19 pb o menos).

En contraste con los exones, el tamaño de distribución de los intrones difiere sustancialmente en estas tres especies (fig. 2.13b, c). En la oruga y la mosca tienen una distribución estrechamente parecida en cuanto a la longitud de los intrones (47 pb en la oruga y 59 pb para la mosca). El tamaño de los intrones en humanos es más variable, cuyo máximo se encuentra en 87 pb, pero una longitud muy larga implica un incremento significativo en el tamaño del gen.

La variación en el tamaño del gen y el tamaño de los intrones puede estar en parte correlacionada con regiones ricas en secuencia GC las cuales tienden a ser regiones con una alta densidad de genes (con algunos genes compactos); mientras que las regiones ricas en AT tienden a ser pobres en genes con algunos genes extendidos que contienen intrones de gran tamaño.

La correlación de la densidad de genes con el contenido GC es mostrado en la fig. 360a, b; la densidad relativa incrementa significativamente con el aumento del contenido GC desde 30% a 50%. La correlación parece ser debida principalmente al tamaño de los intrones; los cuales disminuyen con el incremento del contenido de GC (Fig.2.14c). En contraste la longitud o número de exones varía muy poco (Fig.2.14c).

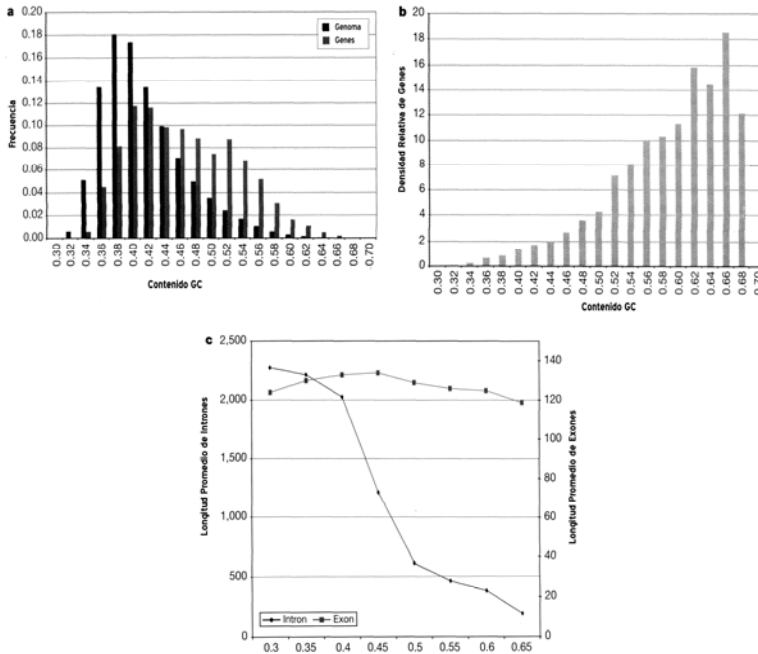


Figura 2.14 Contenido GC a) Distribución del contenido GC en genes y en el genoma. b) Densidad de genes como una función del contenido GC. c) Dependencia de la longitud media de intrones y exones sobre el contenido GC

El gran número de intrones en los humanos permiten analizar sitios variantes de “splicing”, muy importantes debido a que algunas proteínas pueden ser generadas a partir de un gen simple, que puede ser utilizada para distintos niveles de regulación genético, lo que parece ser prevalente en humanos.

Para profundizar mas los avances particulares del PGH consultar la bibliografía.

APLICACIONES A LA MEDICINA Y A LA BIOLOGÍA:

Las aplicaciones del PGH (algunas ya comentadas) son infinitas sobre todo en el campo de la medicina y biología:



GENES CAUSANTES DE ENFERMEDAD:

Una aplicación clave en las investigaciones del genoma humano ha sido la habilidad para encontrar genes de enfermedades, cuya función bioquímica es desconocida por clonación posicional. Estos métodos involucran, mapeo de la región cromosomal que contiene el gen por análisis asociado de familias afectadas y entonces, exploran la región para encontrar el gen mismo (el clonación posicional aunque tedioso es una técnica poderosa para este tipo de trabajo). Por lo menos 30 genes causantes de enfermedades (Tabla 26. Pag. 912) han sido clonados debido a los esfuerzos que dependen directamente de las secuencias genómicas publicadas.

Las secuencias genómicas han permitido revelar los mecanismos de algunos síndromes originadas por delección cromosomal (pérdida o arreglo de información genética en el cromosoma). En algunos casos, las delecciones recurrentes han sido encontradas como resultado de recombinación homóloga o entrecruzamiento desequilibrado entre duplicaciones intracromosomales muy cercanas. Ejemplos de este hecho, incluyen al síndrome velocardiofacial /DiGeorge, que ocurre sobre el cromosoma 22 y el síndrome de Williams-Benren que es una delección recurrente sobre el cromosoma 7.

Las mutaciones en genes parálogos pueden originar una enfermedad genética relacionada. El gen CNGA3, codifica la subunidad α del fotorreceptor cíclico de la compuerta -GMP del canal, el cual ha sido blanco de mutaciones en algunas familias con achromatopsia. Investigación computacional de la secuencia genómica revelaron que el gen CNGB3 corresponde a la subunidad β del canal comentado [este gen no se encontraba anteriormente en las bases de datos (EST)]. Este gen CNGB3 también es causante de achromatopsia en otras

familias. Otro ejemplo, conocido es el de los genes presenilin-1 y presenilin-2; su mutación puede causar enfermedades tempranas de Alzheimer's.

El conocimiento de estos genes, puede proveer una oportunidad para intervenciones terapéuticas para intentar reactivar genes no funcionales producto de alteraciones (mutación) en individuos que padecen de enfermedades genéticas hasta ahora incurables, desde el punto de vista de la medicina convencional.

BIOLOGÍA BÁSICA:

Además de la aplicación directa comentada sobre la medicina, algunas aplicaciones similares son importantes para la biología celular y fisiología básica. Para citar un ejemplo satisfactorio, las secuencias públicas disponibles fueron usadas para resolver un misterio que mantuvo a muchos investigadores inquietos por algunas décadas: Las bases moleculares del sabor amargo. Los humanos y otros animales son polimórficos para ciertos sabores amargos. Recientemente, encontraron que esto se trata de una región del genoma que representa secuencias para el acoplamiento proteína G-Receptores. Estos estudios son importantes para el descubrimiento de nuevas familias de proteínas implicadas en los sabores y la confirmación experimental es la respuesta que se genera por parte de los receptores en los cultivos celulares frente a sustancias específicas de sabor amargo.

Considerables progresos, se han logrado en el secuenciamiento del genoma humano, pero falta mucho trabajo por hacer para desentrañar los mecanismos moleculares del desarrollo humano, de las enfermedades, entre otros aspectos, que permitirán entender los mecanismos adaptivos y evolutivos en bienestar de la supervivencia.



PROYECTO GENOMA HUMANO. SUS IMPLICACIONES Y DESAFÍOS ÉTICO-JURÍDICOS.

Los resultados emergentes del desarrollo del Proyecto no se hicieron esperar y sus implicaciones sociales no tardaron en plantear la necesidad de reflexionar sobre las consecuencias y su encuadramiento ético-jurídico.

Ante el nuevo conocimiento se abrieron posibilidades - literalmente- insospechadas, para las que no existían normas de conducta previstas. Las fuentes: filosóficas, religiosas, legales exhibían una diversidad de posturas ante lo que aparecía como una panacea para enfermedades y debilidades humanas a la vez que como una amenaza hacia la integridad del ser humano tal y como hasta hoy se lo pensó.

Por un lado, el dominio adquirido sobre las relaciones químicas del material cromosomático permitió su manipulación para la creación de nuevos productos y procesos de interés no sólo terapéutico sino industrial en sus más variadas aplicaciones, y con ello la posibilidad y el deseo de apropiarse comercialmente de tales contribuciones. El sistema de patentes que ya había protegido la propiedad de material viviente¹[1] recibió un aluvión de solicitudes reivindicando novedades biotecnológicas y en 1980 -después de casi una década de lucha judicial- en el caso Chakrabarty 2. la Suprema Corte de los Estados Unidos de Norteamérica señaló un camino sin retorno: la materia viva es apropiable, no la natural pero sí la artificial, la lograda en el laboratorio y con aplicación o uso industrial. La patente lograda

para la bacteria de Chakrabarty muy pronto justificó intentos que involucraron al ser humano³ y hoy existen patentes que protegen métodos de diagnóstico y productos terapéuticos logrados a partir de las investigaciones genéticas en la especie humana y métodos para modificar el genoma de mamíferos (incluidos los hombres, a los que está decididamente destinado). Estas nuevas posibilidades: diagnosticar la predisposición o presencia genética de enfermedades dio pie a una nueva cuestión: la divulgación de dicha información.

Los ordenamientos jurídicos más modernos ya habían comenzado a consagrar la confidencialidad de la “información personal” a partir de cuestiones de orden informático y médico, se le agregó, entonces, el genético como nuevo aspecto a cubrir o delimitar. Los interesados en dicha “información” no son sólo los propios portadores de ella, sino sus empleadores, aseguradores y familiares. Paralelamente, se discute el derecho a “no conocer” esa información, especialmente en los casos en que el diagnóstico no trae -aún- la consecuente terapia. Ésta dependerá del tiempo, pues se promete hallarla en breve lapso, pero la investigación que lleve a ella depende de la investigación no ya sólo con cobayos de laboratorio sino con los mismos seres humanos. La experimentación en ellos también ha de ser reglamentada y sus protocolos de investigación médico-científica sometidos a control no sólo científico sino social.

Por último una de las más vapuleadas pesadillas la constituye la idea de la posibilidad de replicar seres humanos asexualmente, a través de la copia de material celular humano tal como lo han hecho los científicos del Instituto Roslin de



Escocia. En nuestro país el decreto 200/96 del Poder Ejecutivo prohibió la clonación en seres humanos y el Consejo de Europa: "la creación de individuos idénticos a seres humanos vivos o muertos". Estos intentos legislativos chocan con la imprecisión de los términos científicos adoptados respecto de la voluntad que intentan plasmar, por un lado; y por otro, la necesidad de no detener el promisorio y provechoso desarrollo de la investigación científica en el campo biológico molecular.

La diversidad filosófica y religiosa ya apuntada, la imprecisa legislación y las poderosas pretensiones empresariales hacen difícil el desarrollo jurídico en esta materia. El debate entre todos los miembros de los diferentes estamentos sociales es imprescindible para definir el curso a adoptar: ¿qué hombre, qué sociedad deseamos legar a las futuras generaciones?

El hombre ha atravesado momento como el presente cuando hubo de decidir sobre la esclavitud, sobre los derechos humanos, sobre la discriminación. Es de esperar que hoy lo haga orientado, como entonces, no sólo por sus intereses actuales sino por los de los hombres porvenir.

No hay que olvidar que lo que entendemos por Proyecto Genoma consiste en principio en la obtención de *información* más o menos en bruto, pero lo realmente importante empieza después (en realidad, simultáneamente): dar *sentido biológico* a tal cúmulo de información, es decir, extraer auténtico *conocimiento*.

La "orgía de datos" que se nos viene encima habrá de ser "digerida" adecuadamente, impulsando nuevos avances a base de sugerir nuevos enfoques, nuevos experimentos, renovadas hipótesis de trabajo, todo ello retroalimentándose en un "circulo

virtuoso" que abrirá las puertas de una nueva era en las Ciencias Biológicas. Se habla por ello de una "Era Postgenómica", en la que se irán integrando los conocimientos acumulados en diversos "Atlas" del ser humano y de otros seres vivos, en los que se podrán interrelacionar de modo funcionalmente significativo diversos niveles de comprensión de la materia viva: génico, genómico, regulación, biología celular, fisiología, evolución, etc. El impacto real de todo ello no se puede prever, pero no cabe duda que el PGH sienta las bases de un salto cualitativo y cuantitativo en nuestra visión del mundo vivo.

Desde el mismo inicio del PGH los propios científicos plantearon la conveniencia de emprender, en paralelo a la parte técnica del Proyecto, estudios y debates interdisciplinarios sobre los posibles impactos éticos, sociales y legales derivados de la avalancha de datos genéticos que suministrará esta magna empresa.

En 1989 se establece en los EEUU el subprograma "ELSI" (*Ethical, legal and social issues*), ligado al Ministerio de energía (DOE) y a los Institutos Nacionales de la Salud (NIH), como parte esencial del PGH, y con una generosa financiación, para asesorar sobre temas éticos, sociales y legales al Parlamento y al Gobierno, y para patrocinar actividades que promuevan la educación pública y el debate social sobre la secuenciación del genoma humano. Los otros proyectos nacionales, así como la coordinación internacional (HUGO, UNESCO) cuentan igualmente con secciones específicas del mismo tipo (en Europa contamos con el ESLA, *Ethical, social, and legal aspects*).



Esta ha sido una iniciativa sin precedentes por parte de la comunidad científica: por primera vez un gran proyecto tecnocientífico cuenta entre sus objetivos explícitos el analizar las cuestiones y dilemas sociales que una nueva tecnología puede suscitar, con amplia participación de filósofos, juristas, responsables sociales, líderes religiosos, entre otros.

En el fondo late la preocupación social sobre el uso/abuso de los datos genéticos. La historia de las ideas eugenésicas proyecta la sombra de la duda sobre si la información genética servirá para discriminar a individuos o poblaciones y para inculcar derechos fundamentales, sobre todo en una sociedad que se fuera impregnando de prejuicios sobre el determinismo genético de cualidades humanas (algo insostenible científicamente, pero que tiende demasiado a menudo a ser susceptible de instrumentalización política destinada a justificar posibles discriminaciones e injusticias).

REFERENCIAS

CAPÍTULO 1

- Griffiths, Anthony J.F.; Miller, Jeffrey H.; Suzuki, David T.; Lewontin, Richard C.; Gelbart, William M. Introduction to Genetic Analysis. 7th ed. New York: W. H. Freeman & Co.; (1999).
- Klug, W. y Cummings, Michael R. *Concepts of Genetics. 5th ed.* PRENTICE HALL, INC (1999)
- Mathew, C.; Van Holde, K. and Ahern, K. Biochemistry. 3rd ed. San Francisco; Addison Wesley Longman. (1999).
- Rawn, D. Biochemistry. Volumen (II), MCGRAW HILL - INTERAMERICANA DE ESPAÑA (1989).

CAPÍTULO 2

- Griffiths, Anthony J.F.; Miller, Jeffrey H.; Suzuki, David T.; Lewontin, Richard C.; Gelbart, William M. Introduction to Genetic Analysis. 7th ed. New York: W. H. Freeman & Co.; (1999).
- Hattori, M. *et al.* The DNA sequence of human chromosome 21. Nature, 405, 311-319 (2000).
- The human genome. Nature (no.6825), 409, 745-964 (2001).
- <http://www.nature.com/nature/focus/humangenome>
- <http://www.um.es/%7Emolecula/anucl.htm>
- http://www2.uah.es/biomodel/c_enlaces/temas_conten.htm
- http://www.ornl.gov/sci/techresources/Human_Genome/project
- http://www.ornl.gov/sci/techresources/Human_Genome/graphics/DNASeq
- http://www.ornl.gov/sci/techresources/Human_Genome/project/info.shtml
-
- http://www.ornl.gov/sci/techresources/Human_Genome/graphics/chr51619
- <http://www.ncbi.nlm.nih.gov/Genomes/index.html>



GLOSARIO

Ácido desoxirribonucleico (ADN): Es el componente químico dentro del núcleo de una célula que lleva las instrucciones para elaborar los organismos vivientes

ADN mitocondrial: El material genético de la mitocondria, los organelos que generan energía para la célula

Alelo: Una de las formas variantes de un gen en un locus o de un marcador particular en un cromosoma. Diferentes alelos de un gen producen variaciones en las características hereditarias tales como el color del cabello o el tipo de sangre.

Amplificación génica: Un aumento en el número de copias de un fragmento de ADN particular. Una célula tumoral amplifica o copia segmentos de ADN en forma aberrante, como resultado de las señales celulares y en ocasiones debido a daños causados por efectos ambientales.

Ácido ribonucleico (ARN): El componente químico que resulta de la transcripción del ADN. En el ARN, la letra U, que

corresponde al uracilo, substituye a la T del ADN. En las células encontramos tres tipos de ARN. El ARN mensajero, el ARN de transferencia y ARN ribosomal

Autosoma: Cualquier cromosoma diferente de los cromosomas sexuales. Los humanos tienen 22 pares de autosomas.

Bacteria: Las bacterias son organismos unicelulares y pueden ser benéficas o patógenas. La mayoría de las bacterias pueden vivir y replicarse fuera de un ambiente celular (extracelulares), mientras otras requieren de otras células para poder replicarse (intracelulares).

Biblioteca de ADNc: Una recolección de secuencias de ADN generadas a partir de secuencias del ARNm. Este tipo de biblioteca contiene sólo ADN codificador de proteínas (genes) y no incluye ADN no codificador

Cariotipo: Es el ordenamiento en base a número y morfología de la constitución cromosómica de un individuo. En el caso de los humanos es 46XX en el sexo femenino y 46XY en el sexo masculino.



Célula: Unidad básica de cualquier organismo viviente. Es un compartimento lleno de agua, componentes químicos y organelos que además contiene una copia completa del genoma de ese organismo.

Centimorgan: Es una medida de la distancia genética que nos dice que tan cerca o lejos dos genes o dos marcadores genéticos están el uno del otro. Un centimorgan es equivalente aproximadamente a 1 millón de pares de bases.

Centrómero: Es la región de constricción primaria en los cromosomas humanos y es el sitio en donde las cromátidas hermanas se unen durante la mitosis y meiosis.

Cigoto: Huevo fecundado originado por la unión de dos gametos con fusión de sus núcleos, hasta el momento de pasar a la forma de blastocisto y su implantación en el útero.

Clonación: El proceso de hacer copias de un fragmento específico de ADN, generalmente de un gen. Cuando los genetistas hablan de clonación, no se refieren al proceso de hacer copias idénticas de todo un organismo.

Clonación o clonaje posicional: Un proceso empleado por los genetistas para localizar los genes responsables de enfermedades cuando se conoce poca o ninguna información sobre las bases bioquímicas de ésta.

Codón: Tres bases en una secuencia de ADN o ARN, las cuales especifican un solo aminoácido.

Cromosoma: Un cromosoma es el resultado del empaquetamiento del ADN y las proteínas previo a la división celular para su segregación posterior en las células hijas. Los cromosomas se encuentran en el núcleo de las células y diferentes especies tienen diferente número y morfología de cromosomas. Los humanos tenemos 23 pares de cromosomas, 46 en total: 44 autosomas y 2 cromosomas sexuales. Cada uno de los progenitores aporta un cromosoma a cada par, de manera que los hijos reciben la mitad de los cromosomas de la madre y la mitad del padre.

Cromosoma artificial bacteriano (BAC): Es un vector de clonación generado en el laboratorio que permite insertar segmentos largos de ADN, de 100 000 a 200 000 bases, provenientes de otras especies en el genoma de bacterias. Una



vez que este ADN ha sido clonado dentro de la bacteria huésped puede replicarse junto con el genoma bacteriano y así se pueden obtener muchas copias de él.

Cromosoma artificial de levadura (YAC): Es un cromosoma artificial que se emplea como vector de clonación de grandes fragmentos de ADN de otras especies en el genoma de la levadura. Los YACs se propagan junto con los otros cromosomas de la célula de levadura.

Cromosoma sexual: Uno de los dos cromosomas que definen el sexo genético de un organismo. Los humanos tienen dos clases de cromosomas sexuales, uno se llama X y el otro Y. Las mujeres normales poseen dos cromosomas X y los hombres normales poseen un cromosoma X y uno Y.

Código genético (ACTG): Las instrucciones contenidas en un gen que le dicen a la célula cómo hacer una proteína específica. A, T, G, y C son las "letras" del código genético y representan las bases nitrogenadas adenina, timina, guanina y citosina, respectivamente. Estas bases junto con un azúcar y un enlace fosfato constituyen los nucleótidos que son la unidad fundamental del ADN. En cada gen se combinan las cuatro

bases en diversas formas, para crear palabras de 3 letras que especifican cuál aminoácido es necesario en cada paso de la elaboración de la proteína.

Cóntigo: La construcción física del mapa del ADN a menudo incluye el aislamiento de grandes fragmentos de ADN en clones, como los YAC y los BAC. Estos clones se analizan para determinar cuáles contienen ADN en común, o en otras palabras cuales están superpuestos. Estos clones contiguos constituyen un "contig" donde los clones adyacentes tienen en común parte de su secuencia. Los "contigs" son importantes porque brindan la posibilidad de estudiar un segmento del genoma completo especialmente cuando se buscan genes con ciertas características y cuando se desea establecer la secuencia de grandes fragmentos de un cromosoma.

Delección: Un tipo especial de mutación que consiste en la pérdida de un fragmento de ADN de un cromosoma. La delección de un gen o de parte de un gen puede ocasionar una enfermedad o una anomalía.



Diploide: El número de cromosomas en la mayoría de las células, excepto en los gametos o células germinales. En los humanos el número diploide es 46.

Doble hélice: Es la estructura espacial del ADN que la podríamos representar como una escalera enrollada en forma de hélice o espiral sobre un eje central. Las estructuras laterales de esta "escalera" están formadas por moléculas de azúcar y fosfato y los "peldaños" están compuestos por nucleótidos unidos por enlaces de hidrógeno.

Electroforesis: Es la técnica por la cual mezclas complejas de moléculas como proteínas, ADN o ARN se separan en un campo eléctrico de acuerdo al tamaño y a su carga eléctrica. La electricidad empuja las moléculas a través de los poros de un gel, que es una sustancia firme como la gelatina. El gel puede hacerse de manera que sus poros tengan distintas dimensiones para separar las moléculas según un rango específico de tamaños y formas. Las moléculas más pequeñas migran más rápido que las más grandes.

Enzimas: Son proteínas que catalizan reacciones químicas, generalmente acelerándolas.

Enzimas de restricción: Grupo de enzimas, cada una de las cuales se une al **ADN** en una secuencia de bases distinta y produce fragmentos de oligonucleotídicos de distintos tamaños longitud.

Fago: Virus cuyo huésped es una bacteria.

Familia génica: Conjunto de genes que tienen en común uno o varios fragmentos de ADN, por haberse originado a partir de un gen ancestral común.

Fenotipo: Conjunto de características observables de un organismo o grupo, fruto de la interacción entre su genotipo y el ambiente en que éste se expresa.

Gameto: Célula germinal madura, funcional que contiene el número haploide de cromosomas de la célula somática. Los gametos provenientes de sexos opuestos (óvulo y espermatozoide) se fusionan para formar el cigoto.

Gen: Unidad de herencia que ocupa una posición concreta en el genoma (locus) y está constituido por una secuencia de ADN que codifica un ARN funcional.



Gen candidato: Gen al que se hace responsable de una enfermedad, tanto por la posición que ocupa en el mapa genómico (candidato posicional) como por las propiedades de la proteína que codifica (candidato funcional).

Gen mutante: Gen que ha experimentado un cambio en su secuencia de bases como pérdida, ganancia o intercambio de material genético, lo que afecta a la transmisión normal y a la expresión del carácter para el que codifica. Estos genes pueden convertirse en inactivos o mostrar actividad reducida, aumentada o antagonista.

Gen recesivo: Gen que sólo se expresa si están presentes dos copias, una de cada progenitor.

Genoma: Complemento cromosómico básico que contiene toda la información genética del individuo.

Genotipo: Conjunto de los alelos de un individuo en uno, varios o todos sus loci.

Haploide: Célula u organismo con un solo complemento cromosómico, como sucede en los gametos tras la meiosis. El

número haploide se simboliza con la letra N (en humanos, el número haploide es N=23 cromosomas).

Herencia: Proceso por el cual determinados rasgos o características se transmiten de padres a hijos. Implica la separación y recombinación de genes durante la meiosis y las posibles influencias posteriores sobre el material genético durante la embriogénesis.

Heterocigoto: Célula o individuo diploide con alelos diferentes en uno o más loci de cromosomas homólogos.

Hibridación: Unión entre dos individuos con fenotipos o genotipos distintos, o bien procedentes de dos poblaciones o especies diferentes. En biología molecular, el emparejamiento específico entre cadenas complementarias de ADN o ARN.

Hibridación in situ fluorescente (FISH): Técnica usada para localizar una sonda marcada en una región cromosómica, haciéndola hibridar con el ADN de una preparación de células en interfase o en mitosis.

Homocigoto: Célula o individuo con alelos idénticos en uno o más loci de cromosomas homólogos.



Intrón: Secuencia de pares de bases en el ADN que genera aquellas partes del ARN precursor que se escinde durante la transcripción y que no entrará a formar parte del ARN mensajero por lo que no especificará la estructura primaria del producto de los genes.

Kilobase (Kb): Unidad de tamaño de los ácidos nucleicos, correspondiente a una longitud de 1000 nucleótidos. En ADN bicatenario se denomina kilopares de bases (Kbp).

Ligamiento: Tendencia de dos o más marcadores genéticos a heredarse juntos en una proporción mayor a la explicada por el principio de distribución independiente que aumenta con su proximidad al reducirse la probabilidad de ser separados durante una reparación del ADN o los procesos de replicación (fisión binaria en procariontas, mitosis o meiosis en eucariotas).

Locus (plural=Loci): Posición que ocupa un gen en el genoma.

Lugar de restricción: Secuencia de nucleótidos que es reconocida específicamente por una endonucleasa de restricción.

Mapa de restricción: Representación lineal de los lugares (o dianas) de restricción contenidos en una secuencia de ADN.

Mapa físico: Serie ordenada de genes y marcadores genéticos localizados en un cromosoma mostrando las distancias físicas relativas (expresadas en pares de bases). Se construye utilizando métodos físicos: secuenciación, mapas de restricción de clones solapantes, hibridación in situ, estudios de delecciones, y otros.

Mapa génico: Serie ordenada de loci genéticos en un cromosoma, deducida tanto por métodos genéticos (estudios de ligamiento) como físicos.

Mapa de ligamientos: Un mapa de las posiciones relativas de los loci en un cromosoma basándose en las frecuencias con que dichos loci se heredan juntos. Las distancias se miden en centimorgans.

Mapeo: Término que designa colectivamente los distintos procedimientos (tanto genéticos como físicos) empleados en la construcción de mapas génicos.



Marcador genético: Locus genético con alelos fácilmente detectables, bien porque producen un fenotipo característico o porque pueden estudiarse por métodos moleculares. Son utilizados en estudios de ligamiento y en la creación de mapas físicos.

Marco de lectura: Cada una de las posibles lecturas de una secuencia de ADN en forma de tripletes. Un marco de lectura abierto es el que da lugar a un ARN mensajero traducible a una proteína.

Meiosis: División celular que tiene lugar durante la formación de los gametos en especies de reproducción sexual, mediante la cual una célula germinal diploide da lugar a cuatro gametos haploides.

Nucleótido: Molécula constituida por una base nitrogenada, una pentosa y un grupo de ácido fosfórico. Es la unidad básica que compone los ácidos nucleicos.

Oligonucleótido: Secuencia lineal de nucleótidos unidos por enlaces fosfo-diéster, habitualmente no mayor de 50 nucleótidos.

Oncogén: **Gen** que induce una proliferación celular incontrolada.

p: En nomenclatura citogenética, brazo corto (del francés, "petit") de un cromosoma.

Palíndrome: Aplicado al ADN, se utiliza tanto para las secuencias nucleotídicas que se leen igual en ambos sentidos sobre la misma hebra (repeticiones invertidas en tándem) como para las secuencias que se leen igual en sentido 5' a 3' sobre hebras complementarias.

Par de bases: Dos nucleótidos complementarios en una molécula de ADN bicatenario.

Pirimidinas: Un tipo de bases nitrogenadas de las cuales la Timina se encuentra sólo en el ADN, el Uracilo sólo en el ARN y la Citosina en ambos.

Plásmido: Elemento genético extracromosómico presente en bacterias, consistente en una molécula circular de ADN bicatenario. En biología molecular se usan como vectores de clonación.



Poligénico: Rasgo fenotípico o enfermedad causado por la interacción de varios genes.

Polimorfismo: Locus genético que está presente en dos o más alelos distintos, de forma que el alelo más raro tiene una frecuencia mayor o igual a 1% (0,01) en la población general. Un polimorfismo puede ser transitorio (las frecuencias alélicas tienden a cambiar debido a una ventaja selectiva) o estable (las frecuencias alélicas permanecen constantes durante muchas generaciones).

Procariota: Perteneciente al super-reino Procariotas, que incluye los microorganismos que se multiplican por división binaria y carecen de núcleo delimitado por envoltura nuclear.

Promotor: Región de ADN que contiene diferentes dominios de unión a factores de transcripción y que determina el lugar donde la ARN polimerasa comienza la transcripción de un gen.

Proyecto Genoma Humano: Proyecto conjunto de investigación de escala internacional que pretende establecer el mapa y la secuencia de todos los genes humanos.

Purinas: Un tipo de bases nitrogenadas que contienen dos anillos (uno de seis y otro de cinco miembros). Las que forman parte del ADN y el ARN son la Adenina y la Guanina.

q: En nomenclatura citogenética, brazo largo de un cromosoma.

Radiación ionizante: Las ondas electromagnéticas de alta energía y los rayos constituidos por partículas en movimiento que en su trayectoria disocian a las sustancias en iones. Afectan directamente a los organismos vivos al interferir el desarrollo celular. Sus efectos genéticos comprenden la producción de mutaciones en los genes y diversas alteraciones cromosómicas.

Rasgo: Cualquier característica determinada genéticamente que se transmite asociada a un genotipo específico. La manifestación del rasgo en el fenotipo puede producirse de modo dominante o recesivo.

Recesivo: Rasgo fenotípico (y los alelos que lo determinan) que sólo se expresa en el estado homocigoto o hemicigoto. Los alelos recesivos se denominan con letras minúsculas para diferenciarlos de los dominantes.



Recombinación: Intercambio de material genético producido por sobrecruzamiento durante la meiosis y, en ocasiones, durante la mitosis.

Recombinante: Célula u organismo que resulta de la recombinación de genes en la molécula de ADN, independientemente de si se ha producido de forma natural o por medios artificiales.

Ribosa: Pentosa que entra en la composición de los nucleótidos.

RNA: Abreviatura inglesa que significa ARN.

Secuencia de consenso: Secuencia ideal que representa los nucleótidos o aminoácidos que se encuentran con mayor frecuencia en cada posición de un fragmento de ADN o de una proteína, respectivamente.

Segregación: Proceso de separación de los alelos de un locus durante la meiosis: al separarse los dos cromosomas homólogos de un par, cada alelo pasa a un gameto distinto. En sentido más amplio se aplica a la separación de alelos y su

distribución a células hijas diferentes, que se produce tanto en la **meiosis** como en la **mitosis**.

Separación (*Splicing*): Eliminación de los **intrones** y unión de los **exones** en **ARN** durante la **transcripción**.

Seudogén: **Gen** inactivo (no produce un producto proteico) cuya secuencia tiene un alto grado de homología con otro **gen** funcional que está en un **locus** distinto. Un pseudogén procesado es una copia de otro **gen** pero que carece de **intrones**, tiene una pequeña cadena de poliadeninas y está flanqueado por repeticiones cortas; se piensa que proviene de la integración en el **genoma** de **ARN** maduro retro-transcrito. Un pseudogén no procesado (o tradicional) surge por duplicación de un **gen** y posterior acumulación de **mutaciones** inactivantes, y suele estar cerca del **gen** funcional del que se originó.

Sonda: Fragmento marcado de **ADN** de cadena simple que al hibridarse con fragmentos de **ADN** que tengan una secuencia complementaria en un filtro de nitrocelulosa permite que éstos sean detectados y localizados.



STS: Acrónimo inglés de "Sequence-tagged site" (lugar con una secuencia marcada). Fragmento corto (200 a 500 pb) de **ADN genómico** cuya secuencia es conocida y su posición ha sido determinada en el mapa físico o genético. Esto permite que sean usados como **marcadores genéticos** en el desarrollo de **mapas físicos** del **genoma** humano.

Telocéntrico: **Cromosoma** que tiene su **centrómero** en un extremo.

Telómero: Extremo libre de los **cromosomas** lineales de **eucariotas**. En humanos, el **ADN** de los telómeros está compuesto por **repeticiones en tándem** de la secuencia T-T-A-G-G-G.

Transcripción: Proceso de síntesis de una molécula de **ARN mensajero** por acción de la **ARN-polimerasa**, tomando como molde la cadena antisentido del **ADN genómico**. Este es el primer paso de la **expresión** génica.

Traducción: Proceso por el que se sintetiza un **polipéptido** tomando un **ARN mensajero** como molde. Se lleva a cabo en los **ribosomas**.

Transcriptasa inversa: ADN-polimerasa ARN-dependiente, es decir, complejo enzimático que sintetiza una molécula de **ADN** tomando un **ARN** como molde.

Transformación: 1. Proceso de introducción de **ADN** exógeno en una bacteria por medios físicos o químicos. 2. Conversión de una célula animal **fenotípicamente** normal en una célula con crecimiento descontrolado, por acción de un virus **oncogénico**. Se manifiesta además en alteraciones de la forma celular y en la pérdida de inhibición por contacto.

Transgénico: Célula u organismo que contiene en su línea germinal un **ADN** exógeno introducido experimentalmente.

Triplete: Sinónimo de **codón**.

Uracilo: Base de pirimidina en el **ARN**.

Vector: En sentido amplio, sinónimo de vehículo. Se aplica a moléculas de **ADN** que se replican y sirven para transferir fragmentos de **ADN** entre células (**plásmidos**, **cósmidos**, BACs, PACs, **YACs**, etc); a sistemas de transferencia de **genes** utilizados en terapia génica (virus, liposomas, etc); o a



organismos capaces de transmitir una bacteria o un parásito (como el mosquito anopheles).

VNTR: Acrónimo inglés de "Variable number of tandem repeats". **Locus** cuyos **alelos** difieren por tener un número variable de **repeticiones en tándem**. Son muy **polimórficos**, por lo que se utilizan como **marcadores** en estudios de **ligamiento** y en la determinación de identidad en medicina legal.

X: En **nomenclatura citogenética** se refiere al **cromosoma X**.

Y: En **nomenclatura citogenética** se refiere al cromosoma sexual masculino.

YAC: Acrónimo inglés de "Yeast artificial chromosome" (Cromosoma artificial de levadura). **Vector** plasmídico de clonaje que contiene en su **ADN** las secuencias del **centrómero**, del **telómero** y la secuencia que controla la replicación autónoma del **cromosoma**, lo que posibilita el **clonaje** de fragmentos de hasta tres millones de bases de longitud.

Zigoto: Sinónimo de cigoto.