

# ***PARTE I***

## ***Herramientas Básicas***



# I - CAPÍTULO 1

## Establecimiento de cultivos de tejidos vegetales

Luis Mroginski, Pedro Sansberro y Eduardo Flaschland

### 1 Introducción

El cultivo de tejidos —en su acepción amplia— puede ser definido como un conjunto muy heterogéneo de técnicas que presentan en común el hecho de que un explante (una parte separada del vegetal que pueden ser protoplastos —células desprovistas de pared celular— células, tejidos u órganos) se cultiva asépticamente en un medio artificial de composición química definida y se incuba en condiciones ambientales controladas. La imprecisión de esta definición puede generar muchas polémicas, pero es actualmente aceptada. Generalmente es común dividir las técnicas del cultivo de tejidos en dos grandes grupos: a) cultivos en medios semisólidos y b) cultivos en medios líquidos, los que a su vez pueden ser agitados (mediante el empleo de agitadores de uso continuo) o estacionarios. También es frecuente dividir al cultivo de tejidos atendiendo a los niveles de complejidad en cultivo de órganos, cultivos celulares y cultivo de protoplastos. Para el establecimiento de los cultivos utilizando cualquiera de los sistemas es necesario tener en cuenta algunos aspectos generales comunes relacionados con el explante, la asepsia, el medio de cultivo y las condiciones de incubación. La discusión de estos aspectos constituye el objetivo de este capítulo. Adicionalmente se incluye el tema de la aclimatación de las plantas regeneradas *in vitro*, que es de gran importancia en la mayoría de las aplicaciones del cultivo de tejidos en la agricultura.

### 2 Explante

Varios factores deben ser tenidos en cuenta en la elección del explante apropiado para el establecimiento de los cultivos *in vitro*, entre ellos:

**1. Objetivo del cultivo:** si bien es difícil tratar de clasificar las aplicaciones que se persiguen con el cultivo de tejidos, se podría esquematizarlas en aplicaciones para:

- *Estudios básicos.* En este caso, los explantes cultivados pueden ser diversos. Si lo único que se quiere lograr es un sistema de callos para estudiar algún proceso fisiológico, se puede cultivar cualquier órgano, tejido o célula viva. En este caso —en que lo único que se busca es la inducción de callos— lo ideal es cultivar explantes jóvenes, derivados de semillas en germinación, donde se obtienen respuestas rápidas y, en general, hay menores problemas de contaminación con microorganismos. A veces se hace uso del cultivo de tejidos porque representa un sistema experimental que simplifica la complejidad generada por los fenómenos de correlación entre las distintas partes que normalmente están presentes en una planta entera. El explante que se usará estará condicionado por lo que se quiere estudiar. Un buen ejemplo lo constituye el cultivo de ovarios fecundados del tomate para estudiar los requerimientos nutricionales durante el crecimiento de los frutos. Asimismo, el cultivo de óvulos ha sido muy útil para estudiar aspectos relacionados con la formación de las fibras en algodón. El cultivo de discos de tallos brindó una valiosa ayuda para estudiar la rizogénesis *in vitro*.

- *Obtención de plantas con sanidad controlada.* Es muy común la utilización del cultivo de tejidos para la obtención de plantas libres de virus (ver VIII.-9). El explante ideal para ello es el meristema (dependiendo de la especie, de 0,2-0,5mm de longitud) consistente del domo y de un par de primordios foliares.

- *Micropropagación.* En este caso dependerá del sistema que se quiere utilizar. Si lo que se quiere explotar es la brotación de meristemas, los ápices terminales y los segmentos uninodales de ramas jóvenes constituyen excelentes explantes. En este caso el cultivador de tejidos debe conocer perfectamente la biología de la reproducción de la planta para aprovechar aquellos explantes que en forma natural son propágulos. En cambio, si se pretende micropropagar mediante el empleo de semillas sintéticas (ver Parte IV, capítulo 2) el explante original deberá posibilitar la inducción

de la embriogénesis somática. En este caso, la utilización de embriones zigóticos inmaduros u hojas, suelen ser frecuentes como explantes.

- *Obtención de híbridos interespecíficos.* Una de las primeras aplicaciones del cultivo de tejidos en la agricultura lo constituyó su empleo como ayuda para la obtención de híbridos derivados de cruzamientos interespecíficos, donde se produce el aborto temprano de embriones. En estos casos, el explante cultivado es el embrión cigótico en estadios tempranos de su desarrollo. Con la misma finalidad también se utilizan ovarios u óvulos fecundados.

- *Obtención de plantas de semillas con embriones rudimentarios.* Para esta finalidad los explantes pueden ser semillas (por ej. orquídeas) o bien, se aíslan los embriones en un estado temprano de desarrollo («corazón») como en el caso de yerba mate o de algunas variedades de duraznero.

- *Obtención de plantas haploides* (considerando como haploide a un esporofito que contiene el complemento gamético de cromosomas). En este caso, los explantes más utilizados son las anteras, aunque también pueden emplearse microsporas aisladas, óvulos y ovarios no fertilizados.

- *Inducción de variación somaclonal.* En este caso se pueden utilizar varios explantes, pero si la finalidad de su utilización es la aplicación en planes de mejoramiento genético de plantas, los explantes utilizados deben posibilitar la regeneración de plantas enteras.

- *Producción y/o conversión de sustancias útiles.* Se puede utilizar una gran variedad de explantes. Los de raíces suelen ser muy utilizados.

- *Obtención de híbridos somáticos.* Para esta finalidad se recurre a la fusión de protoplastos. En la mayoría de los casos se aíslan los protoplastos de mesófilos de hojas y de suspensiones celulares.

- *Conservación e intercambio de germoplasma.* Los meristemas son los explantes preferidos para el desarrollo de sistemas de conservación a mediano y largo plazo (crioconservación con nitrógeno líquido) y para el intercambio de material genético.

- *Establecimiento de suspensiones celulares.* En este caso se recomienda iniciar los cultivos a partir de explantes extraídos de semillas en germinación (hipocótilo, epicótilo, cotiledones, raíces).

**2. Posibilidad de contaminación con microorganismos.** De ser posible, se deben cultivar explantes de plantas donantes que crecen en condiciones de invernadero, con ello se reduce sustancialmente las tasas de contaminación. Por otra parte, es recomendable evitar el uso de «explantes sucios» (raíces, rizomas) que provienen de plantas crecidas en macetas o en el campo, dado que en la mayoría de los casos no es posible conseguir una buena desinfección de los mismos.

**3. Edad fisiológica.** Este es un aspecto de gran influencia en la morfogénesis. Como regla general se puede decir que cuando más joven e indiferenciado se encuentre el explante a cultivar, mejor será su respuesta *in vitro*. Es por ello que los meristemas apicales y axilares son ampliamente usados en numerosas especies. En el caso de la micropropagación de plantas leñosas, la edad del explante es un factor crítico. Si los tejidos son jóvenes, la micropropagación tiene mayores posibilidades de ser exitosa que con tejidos maduros. Este hecho genera la necesidad de realizar tratamientos de rejuvenecimiento de las plantas donantes de explantes.

**4. Tamaño.** En general, cuanto más grande sea el explante mayores serán las posibilidades de inducir la proliferación de callos o la regeneración directa de órganos. Sin embargo, a mayor tamaño de explante también son mayores las probabilidades de que los cultivos se contaminen con microorganismos. También es necesario tener en cuenta que existe un tamaño mínimo del explante, que depende de la especie y del material vegetal, por debajo del cual no es fácil lograr el establecimiento de los cultivos. Los explantes muy pequeños suelen requerir del empleo de medios más complejos o de los denominados medios acondicionados.

**5. Época del año.** Es un factor que suelen tener mucha importancia en la micropropagación y que generalmente está asociado al

grado de dormición que presentan ciertos explantes (yemas, por ejemplo) y también con la posibilidad de mayor o menor proliferación de microorganismos.

### 3 Asepsia

Uno de los principales problemas que se presentan cuando se tratan de establecer los cultivos es el de la contaminación de los mismos con diversos tipos de microorganismos (hongos, levaduras, bacterias, fitoplasmas, virus). El ambiente generado por explante-medio de cultivo-condiciones físicas de incubación es altamente propicio para la proliferación de muchos de estos microorganismos que pueden provocar la destrucción de los cultivos. Es difícil cuantificar el impacto de estas pérdidas, pero en promedio, en laboratorios dedicados a la micropropagación se lo puede estimar en alrededor del 10%. En el mejor de los casos, estos microorganismos no destruyen los cultivos pero compiten con el explante por los nutrientes del medio de cultivo o bien lo modifican. Es muy difícil conseguir cultivos estrictamente asépticos dado que en la mayoría de los casos es altamente probable que los mismos contengan virus y fitoplasmas, por lo que en la práctica, cuando se refiere a cultivos asépticos, en general se quiere significar que son cultivos donde no se produce la proliferación de hongos y bacterias.

Dos son las fuentes de contaminaciones: a) microorganismos presentes en el interior o en la superficie de los explantes y b) fallas en los procedimientos de cultivo en el laboratorio. La correcta detección de estas fuentes y del tipo de microorganismo son aspectos importantes para el éxito de los cultivos, pues por un lado ayuda a determinar la fuente de contaminación y por otro lado, ayuda a la planificación de los procedimientos para controlarlos. Varios géneros de bacterias (*Pseudomonas*, *Xanthomonas*, *Erwinia*, *Enterobacter*, *Agrobacterium*, *Bacillus*, *Micrococcus*, *Staphylococcus*, *Lactobacillus*, *Mycobacterium*, *Corynebacterium*) y de hongos filamentosos (*Aspergillus*, *Penicillium*, *Fusarium*, *Alternaria*, *Cladosporium* y *Neurospora*) están frecuentemente en los cultivos. Es conveniente inspeccionar visualmen-

te –con la ayuda de un microscopio estereoscópico– los cultivos en forma periódica (por lo menos semanalmente). También se pueden realizar pruebas con medios de cultivo diferenciales y «test» bioquímicos específicos.

Para evitar y/o minimizar las contaminaciones de los cultivos con microorganismos es necesario:

1) Conocer el material vegetal con que se trabaja y los posibles contaminantes específicos.

2) Realizar una adecuada preparación de la planta dadora de explantes, cultivándola preferentemente en invernaderos tratadas con productos químicos que eliminen patógenos y eventuales microorganismos endófitos.

3) Proceder a la desinfección superficial de los explantes mediante el uso de compuestos químicos con el objeto de eliminar los microorganismos con el menor daño posible para el explante. Si bien no es posible recomendar un procedimiento general, se puede señalar que el procedimiento más popularizado consiste en una doble desinfección mediante la inmersión de los explantes en etanol (70%v/v) durante 20-60 segundos seguido de hipoclorito de sodio 1 -3%, contenido en el agua de lavandina comercial, durante 3 a 30 minutos, dependiendo de la naturaleza del explante. Algunos procedimientos se basan en el empleo de únicamente etanol o de hipoclorito de sodio. Finalmente, es necesario eliminar los restos de estos productos mediante varios lavados con agua destilada estéril.

En este último punto hay que aconsejar que se utilice agua destilada estéril de reciente preparación, dado que está demostrado que el almacenaje prolongado del agua estéril puede ser la causa de contaminaciones con bacterias. Es aconsejable realizar estas operaciones de desinfección superficial en una cámara de transferencia con aire estéril. En lugar del hipoclorito de sodio se puede utilizar hipoclorito de calcio (6 -12 %) o el cloruro de mercurio (0,1%- 1,5%). Es preciso recomendar extrema cautela con el empleo de este último compuesto, dado que es altamente tóxico y además no es removido con facilidad del explante.

En los casos en que no se utilice etanol, la adición de agentes tensoactivos junto con el desinfectante es una práctica recomendada. Entre los más usados figuran Tween-20 (0,01 – 0.1%) o unas gotas de Triton. El lavado previo de los explantes con agua corriente y detergentes, ayuda a una mejor desinfección. La inmersión de los explantes en soluciones conteniendo sustancias antibióticas y /o antimicóticas (gentamicina, sulfato de estreptomina, ampicilina, tetraciclina, carbenicilina, sulfato de gentamicina, pentacloronitrobenzono, rifampicina, anfotericina B, benomil, carbendazim) puede ser de utilidad, pero deben ser utilizados en casos excepcionales. Estos productos tienen el inconveniente de que alteran la composición de los medios de cultivo y además pueden ser metabolizados por los explantes.

Últimamente han aparecido soluciones biocidas que matan bacterias y hongos, previenen la germinación de esporas y a altas concentraciones pueden eliminar contaminaciones de microorganismos endófitos. Uno de estos compuestos es el PPM (Plant Preservative Mixture). Otro ejemplo es el denominado G-1, un compuesto derivado de los furfuralos de la caña de azúcar. Este compuesto, químicamente: 1-(5-bromofur-2-il)-2-bromo-2-nitroeteno fue desarrollado en la Universidad Central de las Villas (Cuba) y tiene efecto bactericida y fungicida de amplio espectro.

En los casos en que se utilicen plántulas crecidas *in vitro* como plantas donantes de explantes, es necesario desinfectar las semillas para su cultivo y germinación y luego es aconsejable desinfectar también las plántulas resultantes.

En algunos materiales vegetales se utiliza la preincubación de los explantes mediante lo cual éstos son desinfectados suavemente y precultivados durante 24 horas en un medio conteniendo sacarosa, y finalmente son desinfectados nuevamente y cultivados.

4) Emplear medios e instrumentos de cultivo esterilizados, es decir, liberados completamente de cualquier microorganismo vivo o esporas. Para la esterilización, en la mayoría de los casos se hace uso del calor en sus diversas formas: llama directa, calor húmedo (en forma de vapor abierto o bajo presión), calor seco (aire caliente). Se pueden usar hornos a mi-

croondas. El agua caliente también puede ser usada. En el caso de sustancias termolábiles, la esterilización se puede hacer mediante filtración a través de filtros bacteriológicos.

No es posible recomendar ningún sistema de esterilización dado que la exitosa destrucción de los microorganismos depende de múltiples factores entre los que interesan el tamaño del recipiente, el tiempo de esterilización y la naturaleza de la sustancia a esterilizar. Sin embargo, se pueden dar algunas sugerencias:

- La esterilización en estufas mediante calor seco (aire caliente) es recomendable para esterilizar recipientes de vidrios secos (pipetas, cápsulas de Petri, tubos). En estos casos, 2-4 horas en estufas a 180-200 °C brindan excelentes resultados.

- La esterilización con calor húmedo con vapor bajo presión (en autoclave o en una «olla a presión»). Es el procedimiento más empleado para la esterilización de los medios de cultivo (salvo, como se indicó más arriba, para aquellos que posean componentes termolábiles). En este caso, lo más común es usar una presión de 1.46 kg.cm<sup>-2</sup> durante 20 minutos, con lo que prácticamente se destruyen todas las formas de vida. Es importante que en todos los puntos del autoclave se alcance dicha temperatura, para lo cual hay disponibles cintas detectoras colorimétricas.

5) Cultivar los explantes en una cámara de transferencia con aire estéril («gabinete de flujo laminar»), localizada en un ambiente limpio y no expuesta a corrientes de aire. De no disponer este equipamiento, se pueden sustituir con cuartos esterilizados previamente con luz ultravioleta (nunca exponerse a la luz UV en forma directa). La mesada de trabajo y las paredes del gabinete deben ser desinfectadas previamente con etanol al 70%. De la misma manera deben ser desinfectados externamente todos los recipientes (conteniendo medios de cultivo, agua, etc.) que ingresen en el área del aire estéril.

Los operarios constituyen frecuentemente una importante fuente primaria de contaminación, porque es recomendable que, antes de comenzar a trabajar laven sus manos y antebrazos con abundante agua y jabón y se des-

infecten con etanol al 70 %. La utilización de guardapolvos, guantes y máscaras protectoras de la boca y de la nariz, así como los gorros protectores de los cabellos, ayudan a reducir sensiblemente los niveles de contaminación.

Los instrumentos de trabajo (pinzas, pipetas, tijeras, agujas, cápsulas de Petri) deben ser esterilizados antes de su uso. Muchos de estos instrumentos pueden ser colocados en etanol al 95% y, antes de ser usados, se deben flamear cuidadosamente en la llama de un mechero. También es necesario flamear la boca de los recipientes que contienen los medios de cultivo antes y después de cultivar el explante.

6) Incubar los cultivos en cámaras o cuartos de cultivo, cerrados, libres de corrientes de aire y bien higienizados. En lo posible se debe restringir la circulación de personas, y los recipientes con cultivos contaminados deben ser rápidamente eliminados de este sector. Es conveniente que antes del lavado, estos cultivos sean esterilizados.

#### 4 Medios de cultivo

Un medio de cultivo puede ser definido como una formulación de sales inorgánicas y compuestos orgánicos requeridos para la nutrición y manipulación de los cultivos. Existen numerosas formulaciones, cada una de las cuales comprende entre 6 y 40 compuestos. Para dar una idea de la cantidad de formulaciones disponibles, George y col., luego de revisar más de 3.000 trabajos científicos describen en dos tomos (casi 1.000 páginas en total) más de 2.000 medios de cultivo. También dos empresas multinacionales ofrecen para la venta más de 60 medios cada una, listos para su utilización especialmente en la micropropagación comercial de plantas. En la Tabla 1 se presentan tres medios que son muy usados en la actualidad.

Básicamente, los medios de cultivo se componen de compuestos que suministran:

- Una fuente de carbono
- Nutrientes minerales
- Sustancias vitamínicas
- Sustancias reguladoras del crecimiento
- Agente gelificante (en el caso de medios semisólidos)

**Tabla 1:** Composición de tres medios básicos usados en el cultivo in vitro de tejidos.

Compuestos	Medio básico <sup>1</sup>		
	MS	N6	B5
NH <sub>4</sub> NO <sub>3</sub>	1.650	----	----
KNO <sub>3</sub>	1.900	2.830	2.500
KH <sub>2</sub> PO <sub>4</sub>	170	400	----
CaCl <sub>2</sub> .2H <sub>2</sub> O	440	166	150
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	----	463	134
MgSO <sub>4</sub> .7H <sub>2</sub> O	370	185	250
NaH <sub>2</sub> PO <sub>4</sub> .4H <sub>2</sub> O	----	----	150
KI	0,83	0,80	0,75
MnSO <sub>4</sub> .H <sub>2</sub> O	----	----	10,00
H <sub>3</sub> BO <sub>3</sub>	6,20	1,60	3,00
MnSO <sub>4</sub> .4H <sub>2</sub> O	22,30	4,40	----
ZnSO <sub>4</sub> .7H <sub>2</sub> O	8,60	1,50	2,00
Na <sub>2</sub> MoO <sub>4</sub> .2H <sub>2</sub> O	0,25	----	0,25
CuSO <sub>4</sub> .5H <sub>2</sub> O	0,025	----	0,025
FeSO <sub>4</sub> .7H <sub>2</sub> O	27,80	27,85	27,80
Na <sub>2</sub> EDTA	37,30	37,25	37,30
CoCl <sub>2</sub> .6H <sub>2</sub> O	0,025	----	0,025
Glicina	2,00	2,00	----
Tiamina -HCl	0,10	1,00	10,00
Piridoxina -HCl	0,50	0,50	1,00
Ácido Nicotínico	1,00	0,50	0,50
Mioinositol	100,00	----	100,00
Sacarosa	30.000,00	50.000,00	20.000,00
PH	5,7	5,8	5,5

<sup>1</sup>) Composición en mg·L<sup>-1</sup>.

MS = Medio de Murashige y Skoog (Physiol. Plant. 15: 473 - 97. 1962)

N6 = Medio de Chu, C.C., Wang, C.C., Sun, C.S., Hsu, C., Yin, K.C., y Chu, C. (Sci. Sinica 18: 659- 668. 1975)

B5 = Medio de Gamborg, O.L., Miller, R.A. y Ojima, K. (Exp. Cell Res. 50: 151 - 158. 1968)

#### • Otros compuestos.

**Fuente de carbono:** Prácticamente todos los cultivos son heterótrofos (comparativamente unos pocos son autótrofos) y por ende necesitan del suministro de una fuente de carbono. La sacarosa, en concentraciones de 2 al 5%, es el azúcar más utilizado. En algunos medios se la reemplaza por glucosa. En casos particulares se cita el empleo de maltosa o galactosa. También myo-inositol (100 mg/L) suele ser incorporado a los medios resultando un mejor crecimiento de los cultivos.

**Nutrientes minerales:** Los medios de cultivo deben suministrar los mismos macro y micronutrientes que son esenciales para el crecimiento de las plantas enteras. En general se destacan las concentraciones relativamente

altas de nitrógeno y potasio. El nitrógeno es suministrado en forma de amonio y/o nitrato. También se pueden utilizar urea, glutamina y caseína hidrolizada. Es fundamental que el hierro sea incorporado juntamente con un agente quelante (Na<sub>2</sub>EDTA), lo que lo hace disponible es un amplio rango de pH.

**Sustancias vitamínicas:** De todas las que comúnmente se incorporan a los medios, pareciera que la tiamina es la única realmente imprescindible para el buen crecimiento de los cultivos.

**Sustancias reguladoras del crecimiento:** En la mayoría de los casos los medios utilizados para el establecimiento de los cultivos contienen auxinas (ANA, 2,4-D, AIA, AIB, NOA, Dicamba, Picloram) y/o citocininas (BA, KIN, ZEA, 2iP, Thidiazurón). Las giberelinas (especialmente GA3) son requeridas en algunas ocasiones para el cultivo de meristemas o para la elongación de brotes. El ABA, es usado en algunos casos.

**Agente gelificante (en el caso de medios semisólidos):** El agar (entre 0,6 y 1%) es el compuesto más utilizado. También pueden emplearse «Agargel» (0,40-0,60%), «Transfergel» (2,0-2,60%), «Phytigel» (0,25- 0,40%), agarosa (0,80-0,90%) y «Gelrite» (0,10-0,20%).

**Otros compuestos:** Muchas sustancias, de variada composición química, suelen ser adicionadas a los medios básicos. Además de la glicina, otros aminoácidos se pueden agregar a los medios. Es el caso de L-tirosina, asparagina y cisteína, aunque hay que tener presente que en dosis altas pueden inhibir el crecimiento de los cultivos. El carbón activado (0,1 a 5%) suele ser incorporado al medio, dado que es probable que absorba metabolitos tóxicos para los cultivos.

En algunos medios se incorporan ácidos orgánicos como el málico, el cítrico, el pirúvico y el succínico y es frecuente el empleo de L-glutamina y de caseína hidrolizada. Aún hoy se siguen utilizando ciertos componentes de composición química no bien definida como el

agua de coco (5 a 15%), jugo de tomate y puré de banana. También en ocasiones es necesaria la incorporación de agentes antioxidantes (L-cisteína, ácido ascórbico, polivinilpirrolidona) para prevenir el ennegrecimiento tisular causado por la oxidación de polifenoles presentes en los explantes. Este ennegrecimiento puede causar la muerte de los mismos.

La preparación de los medios de cultivo puede llevarse a cabo de diferentes maneras, en «lecturas recomendadas» se citan manuales de laboratorio donde se describen detalladamente este punto.

## **5. Condiciones ambientales para la incubación.**

La incubación de los cultivos se debe llevar a cabo en condiciones controladas. Por lo menos en lo que se refiere a temperatura, calidad e intensidad de luz, fotoperíodo, humedad atmosférica e higiene. Estas condiciones se logran con el empleo de cámaras climatizadas o cuartos especialmente preparados con aire acondicionado (frío-calor) y una buena y uniforme circulación de aire en el interior y dotados de un buen sistema de alarma para cortar la iluminación en caso de no funcionar el aire acondicionado. En general, los cultivos son incubados a temperatura constante de 25-28 °C, con ciclo de luz/oscuridad de 16/8horas. La luz es generalmente provista por lámparas fluorescentes del tipo «luz día» con una irradiancia de entre 50 y 200 μmol m<sup>-2</sup>s<sup>-1</sup>. La humedad atmosférica debe ser elevada (80- 90%).

## **6 Aclimatación de las plantas regeneradas *in vitro*.**

Durante el período de incubación, los cultivos son expuestos a condiciones ambientales disímiles al ambiente externo. La atmósfera interna se caracteriza por presentar una considerable variación diurna en la concentración de CO<sub>2</sub>, humedad relativa elevada, temperatura constante e intensidad lumínica baja. A su vez, el medio de cultivo compuesto por concentraciones elevadas de azúcares (en aquellos sistemas heterótrofos y semiautótrofos de micropropagación), sales y reguladores del crecimiento,

sumado a la ausencia de microorganismos, generan anomalías morfológicas, anatómicas y fisiológicas que las plantas deberán corregir cuando son transferidas al ambiente externo. Este período de adaptación al nuevo hábitat es llamado *fase o etapa de aclimatación*. La estrategia a implementarse durante el mencionado ciclo deberá contemplar el control minucioso de los parámetros ambientales (humedad, temperatura y luz) de tal manera que permita disminuir la deshidratación y, al mismo tiempo, estimular la fotosíntesis con el objeto de generar un rápido crecimiento de los plantines.

El retraso en el desarrollo de la cutícula y la escasa funcionalidad del aparato estomático que presentan las hojas de la mayoría de las especies cultivadas *in vitro*, determinan una alta tasa de transpiración que puede ocasionar la muerte por deshidratación. El control de este proceso fisiológico es de vital importancia durante la aclimatación, teniendo en cuenta que la disminución de la transpiración será gradual y dependerá de la rehabilitación de los estomas, así como también del desarrollo de la cutícula. El equipamiento necesario estará sujeto a la especie, pudiendo utilizarse desde túneles de polietileno para plantas que posean un elevado control de la transpiración (por ej. *Malus pumila* o *Agave tequilana*) o bien, a través del empleo de cámaras climatizadas (Fig. 1) equipadas con sensores que permiten un descenso paulatino de la humedad relativa. En algunos casos puede resultar necesaria la aplicación exógena de ABA (hormona involucrada en el control del cierre de los estomas) o bien, el empleo de sustancias antitranspirantes que forman una capa semipermeable en la superficie de la hoja. En este último caso deberán tomarse algunas precauciones debido a que pueden observarse reacciones de fitotoxicidad.

Resulta imprescindible evitar la exposición a temperaturas extremas tanto en la fase aérea como en el sustrato. Mediante el empleo de extractores y/o acondicionadores de aire combinados con un sistema de niebla, es posible establecer la temperatura de la fase gaseosa entre los 25 y 30 °C durante la estación estival, mientras que en la época invernal es necesario, a veces, el empleo de mantas térmicas o serpentinas, sea de agua o aire caliente a nivel

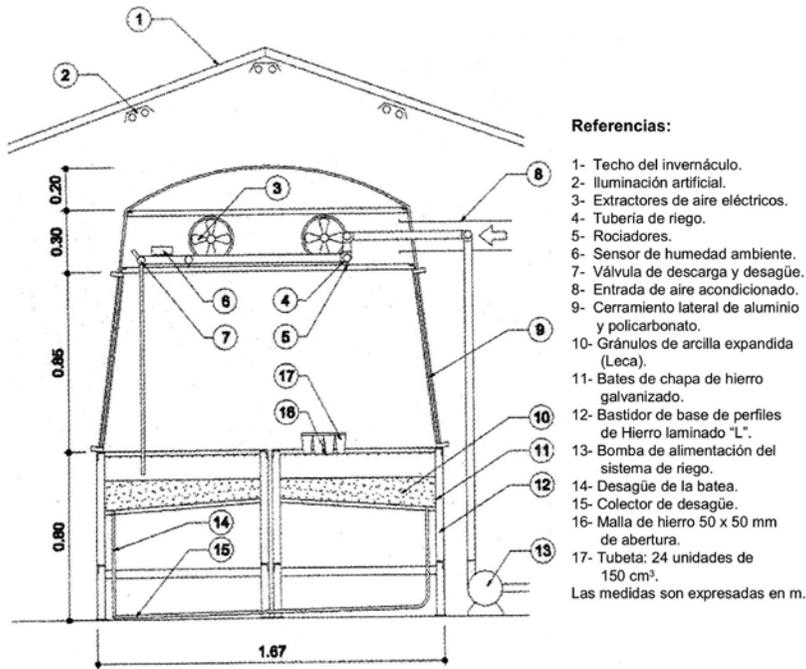
del sustrato, para mantener la temperatura por encima de los 18-20 °C.

Sin lugar a dudas, la opción más económica es el empleo de la luz natural, disminuyendo su irradiancia (20-50%) mediante el agregado de mallas de sombreado («saram»). No obstante, en aquellas latitudes donde el nivel medio de luz natural es bajo y los días son cortos durante una parte considerable del año, la luz artificial puede ser aplicada como complemento de la luz natural. Las lámparas tubulares fluorescentes del tipo «luz día» son empleadas en horticultura para prolongar el fotoperíodo. Asimismo, las lámparas tubulares de sodio alta presión presentan una distribución espectral de la energía adecuada para estimular fotosíntesis y se emplean para tal fin en una amplia variedad de cultivos.

Otro aspecto importante a tener en cuenta lo constituye la elección del sustrato, siendo el adecuado aquel que permita el normal crecimiento y desarrollo de las raíces. Se puede emplear: arena, perlita, turba, vermiculita, o mezclas de ellos, teniendo la precaución de realizar una esterilización previa. Es conveniente el agregado de fertilizantes, sea a través del sustrato (fertilizantes de liberación controlada) o bien mediante el sistema de riego (fertilizantes solubles); empleándose proporciones ricas en fósforo (N-P-K: 9-45-15) y potasio (N-P-K: 4-25-35) que favorecerán el desarrollo radicular y la rusticación de las plantas.

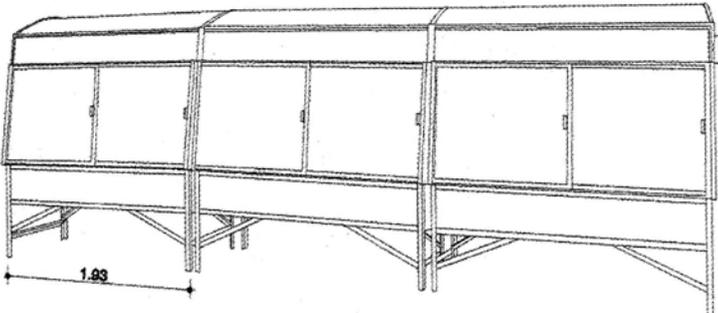
En todo momento deberá realizarse un riguroso control fitosanitario empleándose antibióticos, fungicidas e insecticidas de uso universal.

En la Figura 1, se detalla un modelo de cámara de aclimatación diseñada por los autores y empleada por el Laboratorio de Cultivo de Tejidos del Instituto de Botánica del Nordeste. Básicamente, está compuesta por una fase aérea construida en aluminio y policarbonato alveolar (9) y un recipiente o batea (11) que contiene gránulos de arcilla expandida a fin de facilitar el drenaje. Mediante el agregado de arena u otro sustrato adecuado, esta cámara puede ser utilizada tanto para el enraizamiento *ex vitro* de los brotes obtenidos en la etapa de multiplicación, así como también, para el enraizamiento convencional (a «raíz desnuda») de estacas de tallos u hojas.



- Referencias:**
- 1- Techo del invernáculo.
  - 2- Iluminación artificial.
  - 3- Extractores de aire eléctricos.
  - 4- Tubería de riego.
  - 5- Rociadores.
  - 6- Sensor de humedad ambiente.
  - 7- Válvula de descarga y desagüe.
  - 8- Entrada de aire acondicionado.
  - 9- Cerramiento lateral de aluminio y policarbonato.
  - 10- Gránulos de arcilla expandida (Leca).
  - 11- Bates de chapa de hierro galvanizado.
  - 12- Bastidor de base de perfiles de Hierro laminado "L".
  - 13- Bomba de alimentación del sistema de riego.
  - 14- Desagüe de la batea.
  - 15- Colector de desagüe.
  - 16- Malla de hierro 50 x 50 mm de abertura.
  - 17- Tubeta: 24 unidades de 150 cm<sup>3</sup>.
- Las medidas son expresadas en m.

A - Corte transversal.



B - Vista exterior.

La humedad relativa de la fase aérea es controlada por un sensor (6) ubicado por encima de la tubería de riego (4). El sistema humidificador está compuesto por picos tipo «fog» y es alimentado por una bomba de bajo caudal y alta presión (13). Con el propósito de evitar el goteo que pudiese ocasionar el agua remanente en las cañerías, el sistema consta de una válvula solenoide de descarga y desagüe (7). La fase gaseosa es renovada en forma intermitente mediante el empleo de extractores ubicados en ambos extremos de la cámara (3) los que, conjuntamente, introducen o extraen aire, generando un movimiento masal.

Debido a la latitud geográfica en que se encuentra, la cámara está equipada con un acondicionador de aire (3000 frigorías) para evitar situaciones de estrés térmico en época estival. A su vez, y dado que la mayoría de las especies estudiadas son originarias de climas tropicales y subtropicales, el sistema cuenta con mantas térmicas que son colocadas por debajo de los tubetes, manteniendo la temperatura del sustrato durante el invierno. En este último caso se agrega un material aislante entre la leca (10) y la malla de hierro (16) de tal manera de dirigir el calor hacia la fase aérea.

## 7. Agradecimientos.

Los autores agradecen al Ing. Pablo A. Luna por la planificación digital de la cámara de aclimatación. A la Secretaría General de Ciencia y Técnica (UNNE) y al Establecimiento La Cachuera por el aporte financiero recibido para construir la misma.

## 8. Lecturas Recomendadas

- GEORGE, E.F.; D.J.M. PUTTOCK y H.J. GEORGE. 1987. Plant Culture Media . Vol. 1 (Formulations and Uses). Exegetics Limited. England. 567 pág.
- GEORGE, E.F.; D.J.M. PUTTOCK y H.J. GEORGE. 1988. Plant Culture Media . Vol. 2 (Commentary and analysis). Exegetics Limited. England. 420 pág.
- HURTADO, D.V. y MERINO, M.E. 1991. Cultivo de Tejidos Vegetales. Ed.Trillas. México. 232 pág.
- PÉREZ PONCE, J.N. 1998. Propagación y Mejora Genética de Plantas por Biotecnología. Instituto de Biología de Plantas.Santa Clara. Cuba. 390 pág.
- POSPOSILOVA, J.; I. TICHA, P. KADLECEK, D. HASEL, S. PLZAKOVA. 1999. Acclimatization of micropropagated plants to ex vitro conditions. Biologia Plantarum 42: 481-497.
- ROCA, W.M. y L.A. MROGINSKI. 1993 (segunda edición). Cultivo de tejidos en la Agricultura: Fundamentos y aplicaciones. CIAT. Cali, Colombia, 970 pág.
- TORRES, A.C.; L.A. CALDAS y J.A. BUSO. 1998. Cultura de Tecidos e Transformacao Genética de Plantas. (Vol.1) Embrapa. Brasil.509 pág.
- TORRES, A.C.; L.A. CALDAS, y J.A. BUSO. 1999. Cultura de Tecidos e Transformacao Genética de Plantas. (Vol.2) Embrapa. Brasil.355 pág.

## I - CAPÍTULO 2

### Morfogénesis

Silvia Radice

#### 1.- Introducción

La *embriogénesis somática* y la *organogénesis* son dos procesos morfogénicos muy frecuentes en el cultivo *in vitro* de especies vegetales. La embriogénesis somática es el proceso por el cual se obtiene una estructura similar a un embrión cigótico sin que medie la fertilización de las gametas, mientras que por organogénesis pueden obtenerse tallos, raíces o flores. Estos órganos son inducidos a partir de una célula o de un grupo de células que, según las condiciones de cultivo, tienen la propiedad de mantenerse en activa división.

Esta totipotencialidad celular fue enunciada por Haberlandt en 1902, quien propuso la teoría de que todas las células vegetales tienen la capacidad de regenerar plantas completas. Haberlandt no llegó a demostrar su hipótesis debido a que no pudo lograr la división celular. Los medios de cultivo que empleaba no incluían reguladores del crecimiento debido a que esos compuestos eran aún desconocidos. Los avances en el cultivo de tejidos vegetales fueron muy lentos en sus inicios. En 1934, White pudo mantener, en forma ilimitada, el crecimiento de raíces en medios líquidos a partir de ápices de tomate. Al mismo tiempo se identificó el ácido indol acético (AIA), que permitió el mantenimiento indefinido de callos de zanahoria y tabaco *in vitro*. Posteriormente se descubrió el efecto de la leche de coco como estimulante de la formación de callo sobre el cultivo de embriones de *Datura stramonium*. En 1948 Skoog y Tsui, trabajando con cultivos de callo de tabaco, demostraron la existencia de una regulación química en la parte aérea y en la raíz. Trabajos posteriores en callos de la misma especie, con el agregado de cinetina, la primera citocinina descubierta, permitieron demostrar que la diferenciación de brotes, raíces o de ambos, era regulada por el balance de auxinas/citocininas.

#### 2.- Tipos de morfogénesis. Definiciones

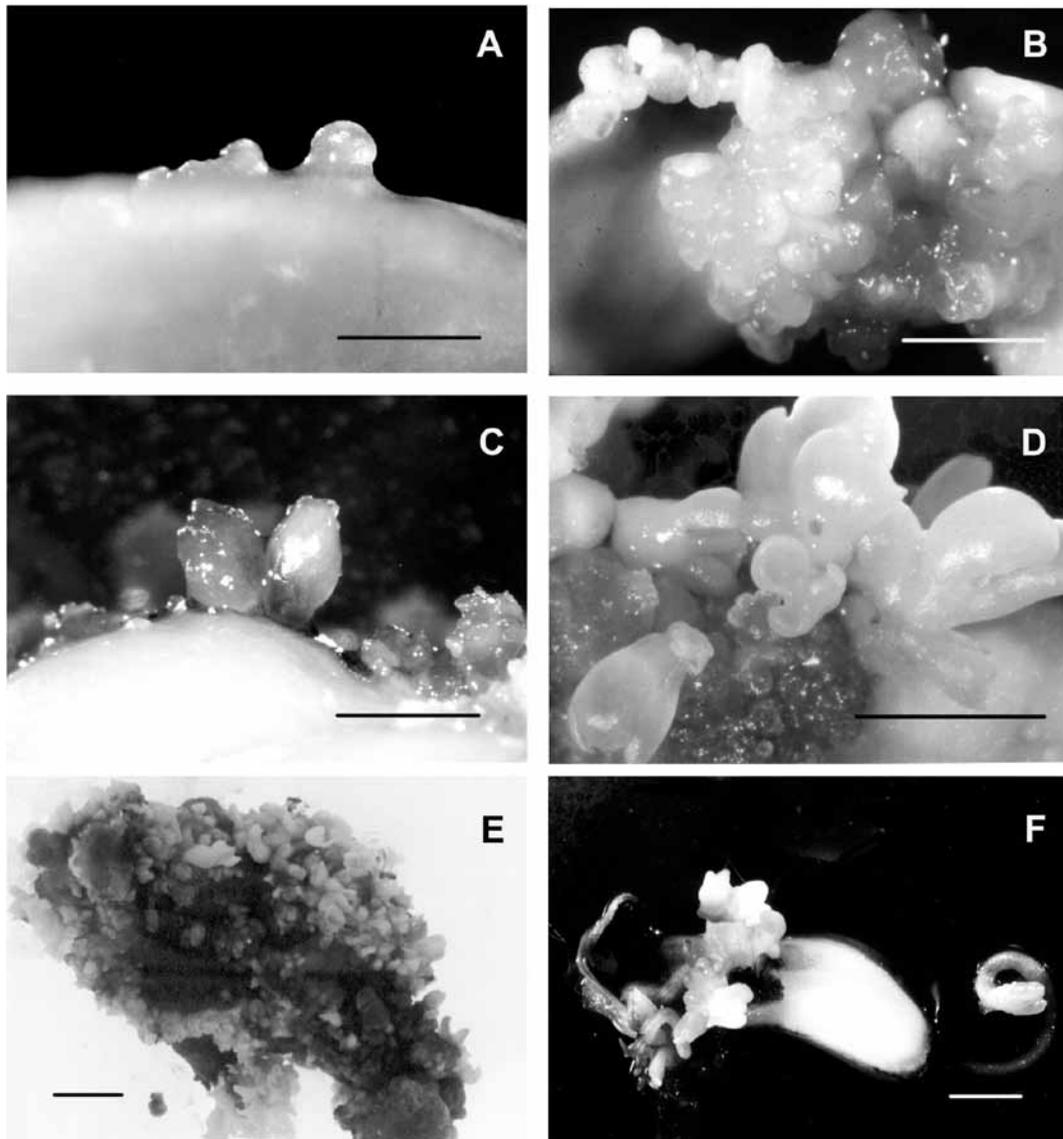
En condiciones de cultivo *in vitro*, las células somáticas pueden regenerar embriones (Fig 1) o brotes, raíces y/o flores (Fig 2) como respuesta a un determinado estímulo. La regeneración es un proceso que comprende diferentes fases que se suceden de manera similar para los tres tipos de morfogénesis citadas. De Klerk y colaboradores, en 1997, denominaron a estas diferentes fases como *adquisición de la competencia*; fase de *inducción* y fase de *realización*.

En la primera fase, las células no responden al estímulo organogénico pero adquieren esa competencia durante una fase de desdiferenciación. En la segunda fase o fase de inducción, las células son receptivas al estímulo morfogénico y hay una relación directa con el tipo, concentración y combinación de reguladores del crecimiento agregados al medio de cultivo y el órgano a desarrollar. En la fase de realización, la célula sufre las sucesivas divisiones para formar el órgano determinado.

A partir de la siembra *in vitro* de diferentes explantes relativamente grandes y en condiciones de cultivo adecuadas, puede inducirse la formación de nuevos órganos de manera directa, sin la formación de callo. Si la formación es de brotes, raíces o flores se denomina *organogénesis directa*. Si en cambio se induce la formación de embriones somáticos, este proceso se denominará *embriogénesis directa* (Figs. 1 y 3). Si por el contrario, a partir de la siembra de un explante *in vitro* se observa la proliferación de células en forma desordenada y sin ninguna función predeterminada, se iniciará la producción de *callos* o *suspensiones celulares* (Figs. 2 A y 3). La diferenciación de órganos a partir de callos, denominada *morfogénesis indirecta*, estará condicionada a la previa formación de los *meristemoides*. *Morfogénesis directa* o *indirecta* fueron términos propuestos por Hicks en 1980.

#### 3.- Histología de la morfogénesis

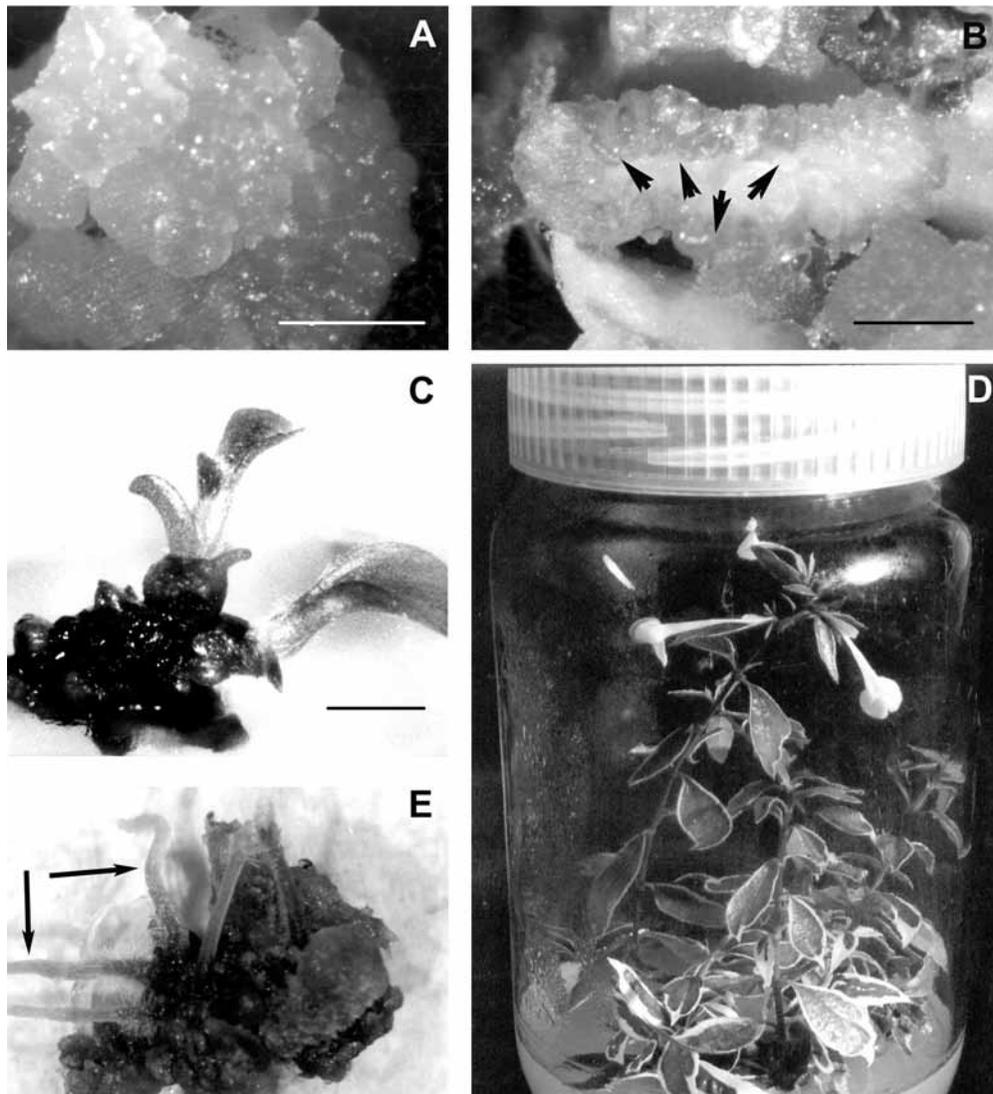
Después de 48 horas de realizada la siembra del explante en el medio de cultivo adecuado,



**Figura 1: A-F** embriogénesis somática obtenida por cultivo *in vitro*. **A-C-E**, embriogénesis somática en diferentes fases de crecimiento observada en *Codiaeum variegatum* (L) Blume cultivado en MS + 1 mg l<sup>-1</sup> de BAP. **B-D-F**, embriogénesis somática obtenida en diversos cultivares de *Prunus sp* a partir de embriones zigóticos inmaduros cultivados en MS + 0,1 mg l<sup>-1</sup> de ANA + 1 mg l<sup>-1</sup> de Kin con una previa inducción en MS + 2 mg l<sup>-1</sup> de 2,4-D. Las reglillas representan 5 mm.

a partir de una célula epidérmica o del mesófilo foliar, como ocurre en las coníferas, se suceden una serie de divisiones mitóticas. Así se pueden observar, a través del estudio histológico, pequeñas masas de células que contienen citoplasma denso y grandes nucleolos. Este conjunto de células, agrupadas a modo de esferas, fueron denominadas *meristemoides* por Torrey en 1966, y son responsables de la dife-

renciación de los nuevos órganos. Este tipo de meristema puede iniciarse en forma directa, a partir de células diferenciadas pertenecientes a un explante cultivado *in vitro*, o indirectamente, a partir de una célula o grupo de células diferenciadas que promueven la proliferación celular (Fig 4 A). Su evolución determinará el crecimiento de un órgano, por ejemplo una yema adventicia (Fig 4 B).



**Figura 2: A-E** Organogénesis somática obtenida por cultivo *in vitro*. **A**, callo organogénico obtenido en *Abelia sp.* (André) Rehder cultivada en MS + 1 mg l<sup>-1</sup> de picloran. **B**, Inicio de brotes (flechas) en hojas crecidas *in vitro* de “Crimson Gold” (*Prunus persica* L.) cultivadas en MS + 0,1 mg l<sup>-1</sup> de ANA + 1 mg l<sup>-1</sup> de Kin. **C**, Brotes de *Gerbera jamesonii* Bolus obtenidos a partir del cultivo de capítulos florales en MS 0,05 mg l<sup>-1</sup> de IBA + 0,75 mg l<sup>-1</sup> de BAP + 0,1 mg l<sup>-1</sup> de GA<sub>3</sub>. **D**, brotes florecidos *in vitro* de *Abelia sp.* (André) Rehder cultivados en MS + 1 mg l<sup>-1</sup> de BAP. **E**, raíces (flechas) crecidas de callo de *Abelia sp.* (André) Rehder cultivadas en inducidas en MS + 1 mg l<sup>-1</sup> de 2iP, previamente inducidas con 1 mg l<sup>-1</sup> de picloran. Las reglillas representan 5 mm

Las células embriogénicas son similares a las células meristemáticas debido a que no son demasiado voluminosas, tienen gran cantidad de ribosomas, citoplasma denso, un nucleolo agrandado y pequeños granos de almidón. Estas células, que en general se agrupan, pueden variar en tamaño y número. Dentro de un gru-

po de células embriogénicas, el desarrollo de las mismas no está sincronizado. Estas células son capaces de dividirse o de transformarse en embriones según las condiciones del medio de cultivo. Aquellas células que no desarrollan proembriones comienzan el proceso de diferenciación, es decir se vacuolizan, disminuye

el volumen de núcleo y nucleolo y desaparecen los gránulos de almidón.

Las experiencias demuestran que las células embriogénicas se desarrollan sólo cuando se modifican las condiciones de cultivo. En general, cuando se reducen las cantidades de auxina y citocinina o se elimina la auxina.

Las células embriogénicas desarrollan como embriones cuando las condiciones experimentales permiten que estas expresen su potencial. Pueden ocurrir dos condiciones ontogénicas diferentes, es decir que el origen de los embriones sea unicelular o pluricelular.

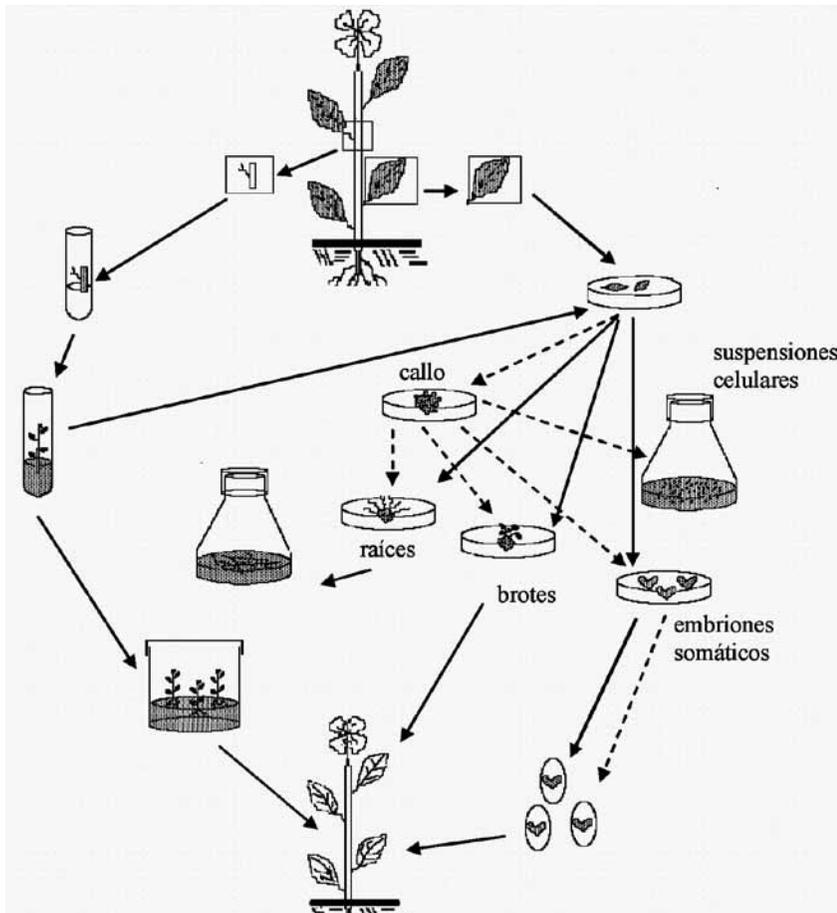
En el caso de iniciarse a partir de una célula, la misma se aísla de las demás por una importante modificación en su pared celular, en particular por la gelificación de la laminilla media. Esta célula, rodeada por un polisacárido mucilagi-

noso, sufre divisiones polarizadas formando un embrión globular. Previo a esta polarización se deposita almidón, luego se forma la epidermis y adquiere simetría bilateral. Sucesivamente se observa el crecimiento de uno o más cotiledones, el desarrollo de los tejidos provasculares, la iniciación de los ápices radicales y de tallo y una progresiva acumulación de almidón, lípidos y proteínas.

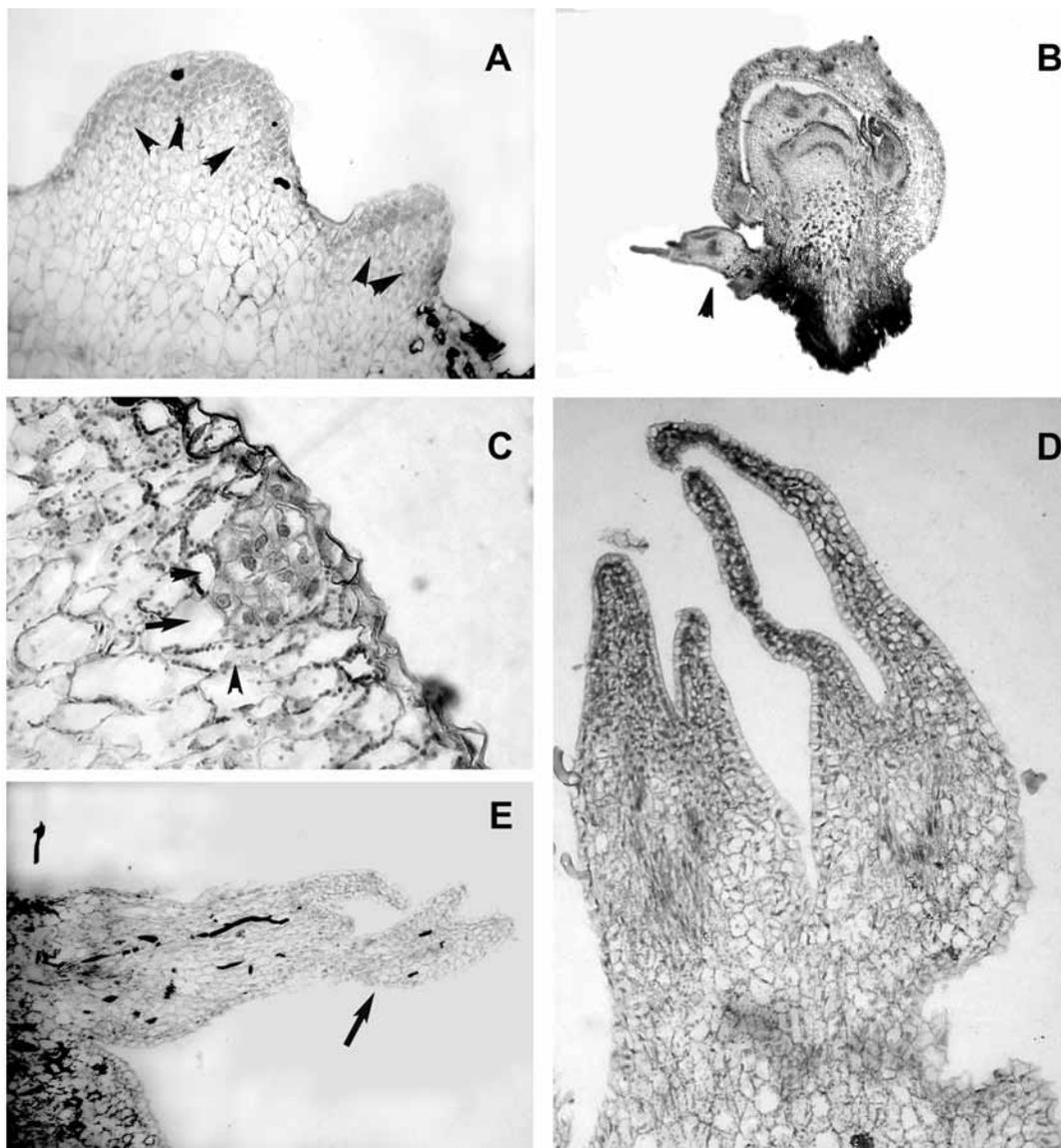
En el caso de un origen multicelular de los embriones somáticos pueden observarse diversos patrones. Los embriones somáticos se forman a partir de grupos de células epidérmicas y subepidérmicas (Fig 4 C). Este tipo de ontogenia de origen multicelular se llama comúnmente *budding* o *embriogénesis adventicia*. Estas células pequeñas tienen una alta relación núcleo/citoplasma, citoplasma denso y un nucleolo vo-

luminoso que se tiñe fuertemente. Se diferencian de las células embriogénicas por la falta de sustancias de reserva, por su delgada pared celular y su frecuente división celular. Estas estructuras se denominan proembriones globulares (Fig 1 A, B). En una etapa posterior desarrollan una epidermis, un tejido provascular, acumulan material de reserva y se polarizan. Estas estructuras globulares desarrollan cotiledones y se transforman en embriones somáticos que presentan ápices caulinar y radicular.

Para una misma especie puede ocurrir uno u otro tipo de inicio según las condiciones de cultivo. El desarrollo de un embrión somático de origen unicelular es comparable a la de un embrión cigótico. En dicotiledóneas, la etapa *globular* es seguida por una etapa *corazón* que se corresponde con la adquisición de la simetría bilateral: desarrollo de la epidermis,



**Fig. 3:** Esquema de los diferentes procesos morfogénicos obtenidos a partir de la siembra *in vitro* de diferentes explantes. Las líneas continuas expresan organogénesis o embriogénesis directa mientras que las líneas quebradas indican que la morfogénesis se obtiene por vía indirecta.



**Figura 4: A-E** Cortes histológicas de diferentes explantes cultivados *in vitro*. **A**, meristemoides (flechas) observados en cultivo de *Silybum marianum* L. en MS + 1 mg l<sup>-1</sup> de AIA + 5 mg l<sup>-1</sup> de BAP 10 x. **B**, yemas adventicias (flechas) en ápices de *Fragaria x ananassa* Duch. cultivados en Boxus + 0,1 mg l<sup>-1</sup> de IBA + 0.5 mg l<sup>-1</sup> de BAP + 0,1 mg l<sup>-1</sup> de GA<sub>3</sub> 4 x. **C**, grupo de células embriogénicas (flechas) en *Codiaeum variegatum* (L) Blume cultivados en 1 mg l<sup>-1</sup> de tidiazurón. 20 x. **D**, embriones somáticos en *Codiaeum variegatum* (L) Blume cultivados en 1 mg l<sup>-1</sup> de tidiazurón 10 x. **E**, embriogénesis secundaria en *Codiaeum variegatum* (L) Blume cultivados en 1 mg l<sup>-1</sup> de tidiazurón 10 x.

de los estratos procambiales y de manera sucesiva el desarrollo del ápice radical. El ápice de tallo se desarrolla más tarde, durante la etapa *cotiledonar*. En los embriones somáticos de algunas especies hay una etapa intermedia entre la globular y corazón llamada *oblonga*, que corresponde a la individualización del ápice de

la raíz y al procambium en respuesta a la elongación axial del embrión somático debido al transporte de auxinas (Fig 1 C).

Los embriones de origen multicelular obtenidos por *budding* muestran la misma ontogenia que los cigóticos, como así también la producción de sustancias de reserva.

Sin el seguimiento histológico detallado, la formación del embrión somático sólo puede ser confirmada por la germinación (Fig 1 F). Sin embargo, debido al bajo porcentaje de embriones que llegan a esta etapa, esta tarea es casi imprescindible para una correcta evaluación de las respuestas encontradas con los diversos medios de cultivo empleados.

Comparados con los embriones cigóticos de la misma especie, los embriones somáticos frecuentemente desarrollan de manera anormal o son inmaduros. La anomalía más frecuentemente encontrada es la formación doble o triple del sistema vascular, causada por la pobre polarización del transporte de auxinas. También se puede encontrar un contenido excesivamente alto, reducido o ausente de las reservas de almidón y/o proteicas, que el meristema apical o radical no estén desarrollados o que la embriogénesis sea repetitiva o secundaria (Fig 4 E).

La falta del meristema apical o una escasa deposición de sustancias lipoproteicas en los embriones somáticos de algunas especies no debe ser tomada como una anomalía. Esto puede ser un síntoma de inmadurez debido a la rápida formación de los mismos. En este caso una etapa intermedia que permita la maduración de las estructuras formadas permitirá incrementar el porcentaje de embriones somáticos capaces de germinar.

#### **4.- Factores que afectan los procesos morfogénicos**

Los cuatro principales factores que condicionan la obtención y el crecimiento de nuevos órganos *in vitro* son:

4.1.- el genotipo

4.2.- las condiciones químicas seleccionadas para realizar el cultivo.

4.3.- las condiciones físicas seleccionadas para realizar el cultivo.

4.4.- el explante.

4.1.- El genotipo es un factor determinante en todos los procesos morfogénicos, desde la capacidad del explante para su establecimiento en condiciones *in vitro* a la proliferación de callo, o la diferenciación y crecimiento de nuevos órga-

nos. Por esta causa, no es posible generalizar metodologías o protocolos de trabajo debido a que los medios de cultivo, así como las condiciones de cultivo seleccionados, deben ser específicos para cada situación en particular. Un ejemplo de ello es el protocolo empleado por Stamp and Meredith en diferentes cultivares de *Vitis vinifera*. En cuatro de ellos fue posible la obtención de embriones somáticos a partir de embriones cigóticos, pero estos resultados no se repitieron con el cultivar "Pinot Noir".

4.2.- Varios son los compuestos químicos que influyen en los patrones morfogénicos *in vitro* dentro de los cuales podemos considerar:

4.2.1.- *la composición salina del medio de cultivo*. La composición salina más empleada para inducir la formación de callo, la organogénesis directa o indirecta en la mayoría de las especies vegetales, es la de Murashige & Skoog (1962) (MS). Sin embargo, existen otras formulaciones diseñadas para inducir determinados patrones morfogénicos. Existe además una estrecha relación entre la composición hormonal del explante y la concentración de reguladores del crecimiento agregada al medio de cultivo.

4.2.2.- *reguladores del crecimiento*. Estos compuestos pueden promover la morfogénesis aún cuando la concentración salina no sea la adecuada. En condiciones óptimas de cultivo pueden aumentar significativamente la diferenciación de órganos. Para la inducción de embriones somáticos en muchas especies vegetales es necesario el agregado de auxinas, como ocurre en *Tilia spp*. Sin embargo para otras, como es el caso del crotón (*Codiaeum variegatum*), es suficiente con el agregado de thidiazurón. El genotipo y el tipo y concentración de reguladores del crecimiento empleados están estrechamente relacionados.

4.2.3.- *antibióticos*. En procesos de transformación con *Agrobacterium tumefaciens* es muy común el agregado de antibióticos del tipo cefotaxime; kanamicina y carbenicilina para la eliminación de la bacteria. En explantes de hoja de diferentes cultivares de *Malus domestica* se encontró que 250 mg L<sup>-1</sup> de cefotaxime aumentan significativamente la regeneración de brotes mientras que 500 mg L<sup>-1</sup> de carbenicilina promueven la formación de callo y la

kanamicina inhibe los procesos morfogénicos.

4.2.4.- *carbón activado*. En general se usa para adsorber compuestos tóxicos de la micro atmósfera gaseosa o el exceso de reguladores del crecimiento presentes en el medio de cultivo pero E.K. Pettersson y su equipo de colaboradores de la Swedish University of Agriculture, demostraron que puede promover la embriogénesis somática debido a la incorporación de ciertos compuestos al medio de cultivo por difusión.

4.2.5.- *agar*. Otro aspecto importante a tener en cuenta es la consistencia del medio de cultivo. Puede emplearse como semi-sólido o líquido. El agente gelificante más utilizado en el cultivo *in vitro* es el agar, extraído de diversas algas marinas. Las diferentes calidades existentes en el mercado modifican la expresión morfogénica debido a que pueden contener sustancias inhibitorias o promotoras del crecimiento. Tal es el caso del cultivo de hojas de *Actinidia chilensis* (en medio de MS con el agregado de 0,1 mg/L de AIA + 1 mg/L de BAP) cuya respuesta morfogénica varió según el tipo de agar utilizado. Cuando el gelificante empleado fue Chubut-agar, de producción nacional, se observó una gran diferenciación de yemas adventicias, mientras que con el agregado de agar SIGMA<sup>R</sup> sólo se indujo la proliferación de callo.

En caso de seleccionar medios de cultivo líquidos, la respuesta morfogénica observada varía de manera considerable. El cultivo líquido es imprescindible cuando se busca promover la morfogénesis indirecta a través de suspensiones celulares. Para ello es necesario cultivar los callos en medio líquido con agitación para permitir que las células se disgreguen y puedan diferenciar raíces, brotes o embriones somáticos en forma aislada.

La elección de un medio de cultivo líquido o sólido depende, además, de la especie con la que se trabaja. En *Cymbidium*, por ejemplo, se ha observado que el empleo de medio líquido promueve una mayor diferenciación de protocormos respecto del mismo gelificado. A su vez, este proceso puede mantenerse durante varios subcultivos, mientras que en medio de cultivo gelificado se observó que los protocormos tienden a formar tallos.

4.2.6.- *atmósfera gaseosa*. Es un factor determinante en los procesos morfogénicos y esta con-

dicionada por el tipo y tamaño del envase seleccionado para el cultivo, así como también por el sistema de cobertura del mismo. Las tapas usadas en cultivo *in vitro* varían desde el tapón de algodón; papel de aluminio; película de resinite transparente o tapas rígidas de polipropileno. El intercambio gaseoso es diferente para cada tipo de tapa, en consecuencia la atmósfera interna también sufrirá variaciones.

En condiciones *in vivo* la atmósfera contiene 78 % de nitrógeno; 21 % de oxígeno y 0,035 % de dióxido de carbono. En cultivos *in vitro* se han registrado además, etileno y otros compuestos hidrocarbonados.

El nivel de oxígeno disponible para el explante condiciona el crecimiento y los procesos morfogénicos. En general se necesita una buena aireación para obtener cualquier tipo de proceso morfogénico. En alfalfa se puede obtener un buen incremento de material a partir de suspensiones celulares embriogénicas cuando la solución se satura con 70% de oxígeno. Por el contrario, una tensión de oxígeno del 5 % estimula la producción de plantas a partir de anteras de tabaco.

La concentración de dióxido de carbono (CO<sub>2</sub>) en la atmósfera gaseosa *in vitro* varía según la respiración y la actividad fotosintética de las plantas. En condiciones de oscuridad, el CO<sub>2</sub> incrementa mientras que en condiciones de luz, disminuye. En trabajos realizados con *Ficus lyrata* se han encontrado concentraciones variables, entre 0,5 % y 8,5 %, según el tipo de sello o tapa empleados. Concentraciones superiores al 1 %, que son generalmente tóxicas en condiciones *in vivo*, probablemente no fueron limitantes para la actividad fotosintética.

La presencia de etileno en condiciones *in vitro* promueve diversas respuestas. Su acumulación tiene efectos inhibitorios sobre el cultivo de células, callos y anteras, La cantidad de etileno producida *in vitro* varía con la especie, el tipo de explante, la concentración de citocininas en el medio de cultivo, el tipo de agar utilizado y con la luz. Una forma de incorporar etileno al envase de cultivo es a través de la esterilización del material de disección realizada con el material embebido en alcohol y flameado con mechero Bunsen. En muchos casos el etileno actúa como promotor de la morfogénesis pero luego es inhibitorio del crecimiento de los órganos diferenciados

4.3.- Entre las condiciones físicas podemos mencionar los efectos de la temperatura, la humedad relativa y la luz.

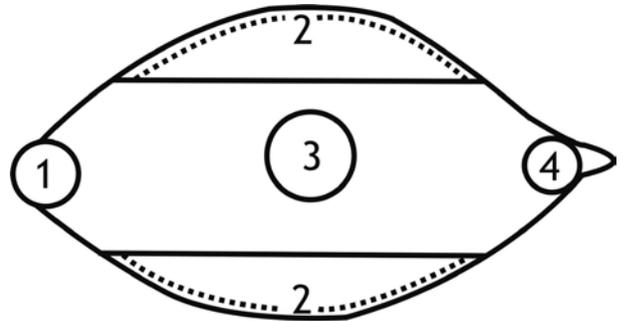
4.3.1.- La *temperatura* de incubación de los cultivos es un factor importante a tener en cuenta. Si bien en condiciones naturales las plantas experimentan diferencias térmicas durante el día y la noche, las temperaturas *in vitro* se mantienen casi estables. Es importante señalar que cuanto más se asemejen las condiciones *in vitro* a las óptimas de crecimiento de la especie estudiada, mayor será la respuesta obtenida. En cultivos de hojas de *Streptocarpus x hybridus* sometidos a diferentes temperaturas se observó que la regeneración mayor de brotes se producía a 12 °C mientras que por encima de los 30 °C, prácticamente no se observó diferenciación.

4.3.2.- La *humedad relativa* (HR), como medida de la cantidad de vapor de agua contenida en la atmósfera gaseosa, es otro de los parámetros físicos a tener en cuenta. La HR dependerá del sello o cobertura del envase empleado. Si este cierre es hermético, la humedad interior será del 100 %. Si en cambio existe la posibilidad de un intercambio gaseoso la humedad interna puede descender a niveles cercanos al 50 %. Este importante descenso de la HR puede promover una pérdida veloz de agua del medio de cultivo variando la concentración de sus compuestos hasta llegar a niveles tóxicos (Debergh, comunicación personal).

4.3.3.- La *luz* suministrada a los cultivos debe ser evaluada en cuanto a la calidad, intensidad y período de suministro. La respuesta morfogénica de un explante puede variar según si se le proporciona luz o no. Para estimular la formación de callo, es común que se prefiera la oscuridad. El suministro de luz favorece la diferenciación de órganos.

4.4.- Tal como ya se señaló anteriormente, los procesos morfogénicos dependen del genotipo, pero a esto debe sumarse el efecto del explante seleccionado. El tratamiento de la planta madre, las condiciones físicas y fisiológicas en las que ésta se encuentre y el sector del cual se tome el explante determinarán a su vez la respuesta morfogénica en condiciones *in vitro*. Un ejemplo muy interesante es la diferente respuesta morfogénica observada en hojas de *Camellia japonica*

cultivadas en un mismo medio de cultivo, según la porción de la lámina utilizada (Pedroso y Pais, 1993). Estos investigadores cultivaron estas hojas *in vitro* durante seis semanas previa inmersión del explante en una solución de IBA (1 mg l<sup>-1</sup>) durante 20 minutos (Fig 5).



**Figura. 5:** Esquema de las diferentes respuestas morfogénicas obtenidas después de seis semanas de cultivo *in vitro* de hojas de *Camellia japonica* L. según la porción de la lámina. 1, callo organogénico. 2 embriogénesis somática directa. 3, regeneración directa de raíces. 4, callo organogénico.

### Lecturas recomendadas

- Buddendorf-Joosten J.M.C. and Woltering E.J. 1994. Components of the gaseous environment and their effects on plant growth and development *in vitro*. *Plant Growth Regulation* 15: 1-16.
- De Klerk G.J., Arnholdt-Schmitt B., Lieberei R. and Neumann K.H. 1997. Regeneration of roots, shoots and embryos: physiological, biochemical and molecular aspects. *Biologia Plantarum* 39 (1): 53-66.
- George E. 1993. *Plant propagation by tissue culture Part 1 The technology*. Butler and Tanner Ltd (ed). Gran Bretaña. 1361pp.
- Hicks G.S. 1980. Patterns of organ development in plant tissue culture and the problem of organ determination. *Bot. Rev.* 46: 1-23.
- Williams E.G. and Maheswaran G. 1986. Somatic embryogenic factors influencing coordinated behavior of cells as an embryogenic group. *Ann. Bot.* 57: 443-597.

# I - CAPÍTULO 3

## La citogenética molecular e inmunocitogenética en el estudio de los genomas vegetales

Guillermo Seijo; Graciela I. Lavia;  
Germán Robledo; Aveliano Fernández;  
Viviana G. Solís Neffa

### 1 Introducción

La comprensión de la organización de los genomas lograda a través del análisis citogenético ha tenido un gran impacto en la evaluación y utilización de la biodiversidad, así como en el mejoramiento genético y la ingeniería genética, en particular, desde el desarrollo de las técnicas de citogenética molecular. El empleo sondas de ADN y de anticuerpos conjugados con fluorocromos en un número creciente de especies vegetales ha permitido, entre otros logros, la localización de secuencias específicas, la identificación de cromosomas particulares, el estudio del origen y la estructura de los genomas de híbridos y poliploides, la comparación de regiones genómicas homólogas, el estudio del comportamiento cromosómico y el análisis de la expresión génica. Por ello, las técnicas de hibridación *in situ* (ISH) e inmunocitogenética constituyen actualmente herramientas indispensables para la comprensión de la organización y la expresión del genoma de las plantas.

En este capítulo se describen, en primera instancia, algunos de los criterios empleados tradicionalmente para la caracterización cariotípica y, posteriormente, se presentan los principios y las aplicaciones de diferentes técnicas de hibridación *in situ* de ácidos nucleicos y de inmunodetección de modificaciones epigenéticas de nucleótidos e histonas.

### 2 Citogenética clásica

La caracterización cromosómica de especies o variedades de plantas se inicia con la descripción del **cariotipo** a través del análisis del **número**, el **tamaño** y la **forma** de los cromosomas que presentan (**complemento cro-**

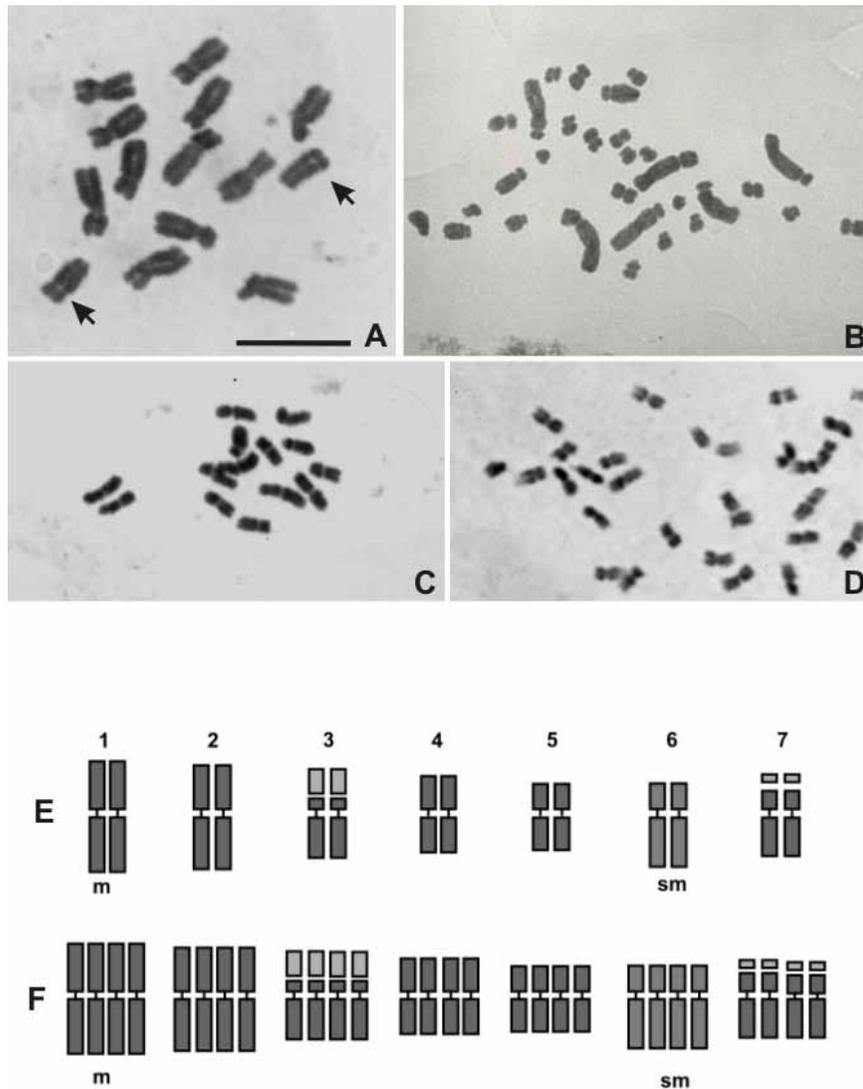
**mosómico**). Esta caracterización se realiza en **metafase mitótica** debido a que, en esta fase, los cromosomas han alcanzado su máximo grado de condensación y sus límites están perfectamente definidos. Para este análisis generalmente se emplean tejidos meristemáticos porque presentan un alto **índice mitótico** (i.e. la razón entre el número de células en división y el número total de células observadas). Los órganos (ápices radicales o caulinares, óvulos, etc.) son pre-tratados con diversas sustancias que inhiben la formación del huso mitótico (colchicina, 8-hidroxiquinoleína, paclosol, etc.) a los efectos de acumular metafases, dispersar los cromosomas en el citoplasma y producir una mayor condensación de los mismos. Los materiales se fijan y luego se tiñen con colorantes nucleares (fucsina básica, orceína, carmín, giemsa, etc.). Finalmente, se confeccionan los preparados por macerado y aplastado ("squash") del material sobre un portaobjetos. Con estos métodos de tinción, los cromosomas metafásicos observados con un microscopio óptico aparecen uniformemente teñidos, excepto en el **centrómero** (constricción primaria) y en algunas constricciones menores (secundarias).

Mediante la observación microscópica se determina el **número cromosómico somático (2n)** y, a través del análisis de microfotografías o dibujos de metafases con cromosomas bien definidos y separados (Fig. 1A), se establecen los cariotipos. Para ello, se analiza la **morfología** de los cromosomas determinando la **posición del centrómero** y la **longitud** de los **brazos corto (bc)** y **largo (bl)**, así como la **longitud total (lt)** de cada cromosoma. A partir de estas medidas, se calcula el **índice centromérico** [ $IC = (bc / lt) \times 100$ ], según el cual los cromosomas se clasifican en cuatro tipos morfológicos: **metacéntricos (m, IC = 50 - 37,5)**, **submetacéntricos (sm, IC = 37,5 - 25)**, **subtelocéntricos (st, IC = 25-12,5)**, **acrocéntricos (t, IC = 12,5 - 0)** y **telocéntricos (T, IC = 0)**. Otro aspecto a considerar para caracterizar morfológicamente a los cromosomas es la presencia y la posición de las constricciones secundarias, así como la presencia y el tipo de **satélites**. En general, las primeras corresponden a **regiones organizadoras de los nu-**

**cleolos (NORs)** y los últimos a las porciones cromosómicas distales delimitadas por la constricción secundaria y el telómero de los brazos que los portan. Los cromosomas que presentan satélites se denominan **cromosomas SAT** (Fig. 1A, E y F).

Para la confección del cariotipo los cromosomas del complemento se ordenan según su morfología (de metacéntricos a submetacéntricos, subteloacéntricos, acrocéntricos y telocén-

tricos) y, dentro de cada tipo, según su tamaño (de mayor a menor) ubicando los centrómeros alineados a una misma altura y los brazos cortos hacia arriba. En las plantas diploides (Fig. 1C y E) y alopoliploides, el cariotipo se compone de **pares** de cromosomas homólogos; mientras que en las autopoliploides (Fig. 1D y F) se compone de **grupos** de tres o más cromosomas de acuerdo con el nivel de ploidía. El orden que ocupa en el cariotipo cada par o



**FIGURA 1. A-D:** Metafases mitóticas teñidas con coloración de Feulgen. **A:** *Lathyrus macropus*,  $2n = 2x = 14$ , las flechas señalan los satélites de los cromosomas SAT. **B:** *Oziroë argentinensis* [citada en Fernández y Daviña (1991) como *Fortunatia biflora*],  $2n = 2x = 30$ , con un cariotipo asimétrico. **C y D:** *Turnera sidoides* subsp. *pinnatifida*, **C:** citotipo diploide,  $2n = 2x = 14$  y **D:** citotipo tetraploide,  $2n = 4x = 28$ . **E-F:** Idiogramas de los citotipos diploide y tetraploide de *Turnera sidoides* subsp. *pinnatifida* ilustrados en C y D, respectivamente. En gris se representan los satélites de los cromosomas SAT. La barra representa 10 micras en A y B, 5 micras en C y D y 2,5 micras en E y F.

grupo de cromosomas se indica mediante un número y, mediante una letra, el tipo cromosómico al que pertenece de acuerdo a su morfología. La representación gráfica del cariotipo se denomina **idiograma** (Fig. 1E y F). La composición de un complemento cromosómico también se puede expresar mediante una **fórmula cariotípica**, en la que se indica el número de cada tipo morfológico de cromosomas que presenta el material analizado (ej. 16 m + 10 sm + 2 st + 2 t). Entre los parámetros cuantitativos más utilizados para caracterizar los cariotipos se encuentra el **grado de asimetría** que presentan los mismos. Para ello, se emplean **categorías** o **índices de asimetría cromosómica** que consideran las diferencias de tamaño entre los cromosomas del cariotipo (**asimetría intercromosómica**), y las diferencias morfológicas de los cromosomas derivadas de la proporción entre sus brazos (**asimetría intracromosómica**). Un **cariotipo simétrico** es aquel cuyos cromosomas son aproximadamente del mismo tamaño y en su mayoría metacéntricos, mientras que uno **asimétrico** es aquel que presenta cromosomas de diferentes tamaños y predominantemente submetacéntricos, acrocéntricos o telocéntricos (Fig. 1B). La información obtenida a partir del análisis de los cariotipos permite identificar cromosomas, clasificar los cariotipos, realizar análisis comparativos entre grupos de especies más o menos relacionadas e inferir sus tendencias evolutivas. Asimismo, permite detectar las anomalías numéricas y estructurales en los complementos cromosómicos de células o individuos.

En numerosos grupos de plantas, la identificación de los diferentes pares de homólogos del cariotipo puede resultar engorrosa debido a que presentan cromosomas con morfología muy similar. En estos casos, la aplicación de determinados tratamientos en combinación con diferentes tipos de tinción, puede revelar regiones cromosómicas con condensación o composición cromatínica diferencial (**bandas**), generando marcadores cromosómicos adicionales. Así, el tratamiento de los cromosomas con un álcali y la posterior tinción con Giemsa revela regiones de **heterocromatina constitutiva**, mientras que la impregnación argéntica (bandeo Ag-NOR) revela las **NORs activas**

(Fig. 2A). Además, pueden utilizarse diversos fluorocromos que tienen la capacidad de intercalarse preferentemente en regiones del ADN con una composición de bases particular. Por ejemplo, las regiones ricas en adenina y timina son reveladas con DAPI (4',6-diamidino-2-fenilindol diclorhidrato) o quinacrina, mientras que las ricas en guanina y citosina lo son con CMA<sub>3</sub> (cromomicina A<sub>3</sub>). Las bandas pueden ser de tamaño e intensidad diferentes y presentar una distribución particular sobre los cromosomas de un complemento generando un **patrón de bandeo** característico (Fig. 2B). Estas técnicas de **bandeo cromosómico** han permitido, en muchos casos, la identificación inequívoca de cromosomas homólogos, la caracterización de genomas así como el seguimiento de cromosomas o bloques cromosómicos en híbridos y líneas de introgresión.

La diferenciación lineal de los cromosomas por técnicas de bandeo no siempre es evidente en todos los grupos de plantas y, aún en aquellas especies que presentan buenos patrones de bandas, a menudo no se cuenta con un número significativo de marcadores como para integrar los mapas de ligamiento con los mapas físicos. La posibilidad de posicionar secuencias particulares sobre los cromosomas y, de esta forma, generar un gran número de marcadores cromosómicos ha sido posible gracias al desarrollo de las técnicas de **hibridación in situ**, cuyo uso se ha generalizado a partir del desarrollo de sistemas de marcado y de detección de sondas no radiactivas.

### 3 Hibridación in situ fluorescente (FISH)

Esta técnica se basa en reacciones de reconocimiento y asociación de bases entre un segmento del ADN cromosómico (**secuencia blanco**) y una secuencia complementaria marcada (**sonda**). La misma se aplica a preparaciones de células somáticas o germinales, previamente incubadas con enzimas (celulasa, pectinasa y citohelicasa) que remueven las paredes celulares. Las preparaciones son tratadas con ARNasa A para evitar hibridaciones inespecíficas de la sonda con ARNs y, posteriormente, son permeabilizadas mediante la incubación con proteasas. Las sondas pueden

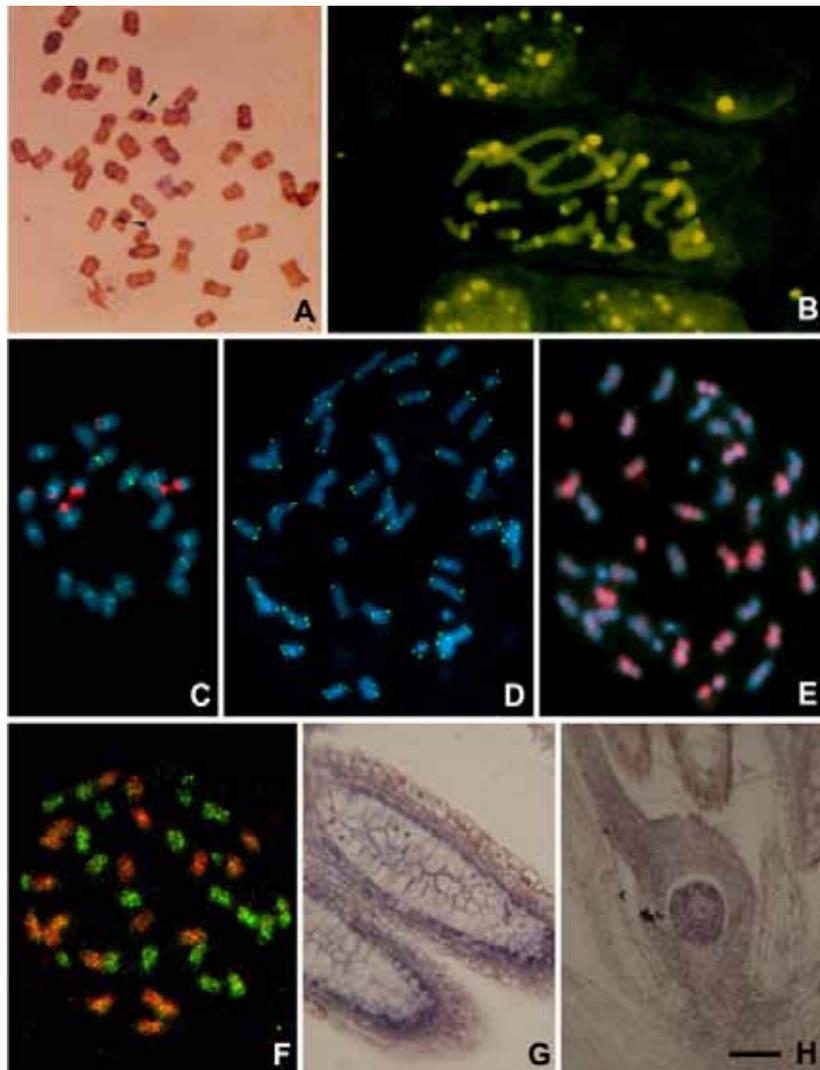
ser **marcadas** con nucleótidos unidos a haptenos (digoxigenina, biotina, etc.) o a fluorocromos (Cy3, Cy5, etc.) por la reacción en cadena de la polimerasa (PCR), por una transferasa terminal o por medio de la actividad exonucleotídica y de la subsiguiente polimerización realizadas por la ADN pol I a partir de escisiones simples causadas por una ADNasa (**nick translation**). La sonda marcada, componiendo una mezcla de hibridación que contiene formamida y ADN bloqueante (entre otros componentes) es colocada sobre las preparaciones cromosómicas, para luego realizar una **desnaturalización** conjunta del ADN de los cromosomas y de la sonda, a temperaturas que varían entre 70 °C y 90 °C durante unos 10 minutos. Posteriormente, los preparados son incubados a 37 °C durante al menos 1 hora (normalmente toda la noche) a fin de inducir la **hibridación**. En este paso, la sonda y la secuencia cromosómica complementaria al blanco compiten entre sí, pero debido a la alta concentración de la primera y a su menor tamaño, bajo condiciones óptimas, la sonda hibrida en todos los sitios complementarios a lo largo de los cromosomas. En el caso de las sondas marcadas con haptenos, la detección de los sitios de hibridación se realiza mediante una o dos rondas de anticuerpos conjugados con fluorocromos. Los preparados se analizan con microscopios de epifluorescencia utilizando filtros específicos para cada fluorocromo, registrando la **presencia**, el **tamaño** y el **número de señales fluorescentes** por complemento. Normalmente, se obtiene una microfotografía monocromática por cada sonda utilizada y una del complemento cromosómico teñido con DAPI o yoduro de propidio. Estas microfotografías son superpuestas e integradas en una sola imagen. El posicionamiento relativo de las señales con respecto al centrómero, los telómeros u otros marcadores cromosómicos permite mapear con precisión cada sonda sobre el complemento cromosómico y representarlas en un idiograma. A diferencia de la citogenética clásica, la identificación de cromosomas o de segmentos cromosómicos particulares usando sondas de FISH es independiente de la morfología cromosómica y, por lo tanto, es posible reconocerlos en diferentes estados del ciclo celular.

Desde el desarrollo inicial de la técnica de FISH, se han mejorado sustancialmente tanto la **sensibilidad** (i.e. el tamaño mínimo de una secuencia que puede ser detectada bajo el microscopio) como el **poder de resolución espacial** (i.e. la distancia física mínima a la que dos secuencias pueden ser identificadas como independientes). Bajo condiciones óptimas de hibridación y de detección, la sensibilidad de esta técnica depende de la **accesibilidad** de la sonda a la región complementaria sobre el ADN. La misma, a su vez, está determinada por el **grado de condensación** de la cromatina. El éxito de FISH se debe a las indudables ventajas que ofrece frente a las técnicas clásicas para identificar homologías y, sobre todo, a la posibilidad de contar con una amplia batería de sondas disponibles de acuerdo con la naturaleza de la secuencia blanco que se quiera detectar.

En general, se pueden utilizar sondas de ADN de **copia única**, **moderadamente repetitivo** o **altamente repetitivo**. Las secuencias repetitivas pueden ser **conservadas** o **variables** y estar dispuestas en **tándem** (i.e. una copia está seguida de otra en arreglos de decenas a miles de copias formando un locus definido) o **dispersas** (i.e. las repeticiones pueden estar dispuestas a lo largo de una región cromosómica, de algunos cromosomas o de todo el complemento). Entre las secuencias repetitivas, las sondas de los **genes ribosomales (ADNr) 45S** y **5S** son las empleadas con mayor frecuencia para la generación de marcadores cromosómicos por FISH. Esto es debido a que son secuencias altamente conservadas y a que presentan una disposición en tándem que facilita la detección de los *loci*. Los mismos pueden estar presentes en más de un sitio por complemento cromosómico, proporcionando marcadores útiles para la identificación de cromosomas y para realizar comparaciones cariotípicas (Fig. 2C). Otra de las sondas de secuencias conservadas y repetidas en tándem que se emplean habitualmente en experimentos de FISH son las **teloméricas** (Fig. 2D). Las mismas son utilizadas tanto para establecer con precisión los límites cromosómicos como para revelar inversiones que comprenden los segmentos cromosómicos distales y la ocurrencia de fusiones

cromosómicas. Las sondas de **secuencias no conservadas repetidas en tándem** son muy útiles para el desarrollo de marcadores específicos de cromosomas y de complementos cro-

mosómicos; mientras que las de **secuencias repetidas dispersas** lo son para definir regiones cromosómicas o genomas, en particular, para identificar los genomas parentales en los



**FIGURA 2.** **A:** Bando Ag-NOR en una metafase de *Arachis hypogaea* donde se identifican las regiones organizadoras del nucleolo en marrón oscuro (flechas). **B:** bando con quinacrina que distingue las regiones heterocromáticas ricas en AT (en amarillo intenso) de las regiones eucromáticas (amarillo pálido) en los cromosomas prometafásicos de *Oziroe argentensis*. **C:** FISH doble en una metafase de *A. batizocoi* en donde se localizan los genes ribosomales 5S (en verde) y 45S (en rojo), la contratinción realizada con DAPI diferencia bandas heterocromáticas centroméricas ricas en AT (en blanco) de las porciones cromosómicas eucromáticas (en azul). **D:** FISH que revela las regiones teloméricas (verde) de *A. hypogaea*. **E:** FISH que revela la distribución preferencial de un retroelemento en el genoma A de *A. hypogaea*. **F:** GISH doble sobre una metafase de *A. hypogaea* utilizando como sondas ADN genómico de *A. ipaensis* (rojo) y de *A. duranensis* (verde) que demuestra la constitución alopoliploide del cultígeno y las especies diploides progenitoras más probables. **G y H:** ARN-FISH en antera y ovario de *Paspalum notatum* utilizando como sonda un ARN marcado de expresión diferencial y tejido-específica en plantas apomícticas pero no sexuales. La intensidad del color violeta indica cantidades relativas del ARN analizado presente en los diferentes tejidos de plantas apomícticas. La barra representa 5 micras en A, C-F, 10 micras en B, 200 micras en G y 100 micras en H.

híbridos interespecíficos y en los aloploiploides (Fig. 2E). Ambos tipos de secuencias repetidas pueden mostrar diferencias notorias en los motivos que las componen, así como en su abundancia y distribución, aún entre especies estrechamente emparentadas. El análisis del conjunto de las secuencias repetitivas constituye una herramienta excelente, en sí misma o como complemento de otras técnicas, para la identificación de cromosomas y genomios, el estudio de la arquitectura nuclear y los análisis filogenéticos. También nos permiten realizar estudios detallados de aberraciones cromosómicas estructurales y numéricas, hacer inferencias sobre los mecanismos involucrados en la evolución cromosómica y realizar el seguimiento de cromosomas o regiones particulares en las sucesivas generaciones de los programas de mejoramiento. Las sondas de **secuencias únicas** hibridan con el ADN cromosómico correspondiente a un locus específico. A diferencia de las secuencias repetidas que pueden ser localizadas fácilmente por FISH, la detección de las secuencias únicas es más laboriosa debido a que, por lo general, las regiones codificantes de los genes presentan longitudes menores que las detectables por FISH convencional (10 kb). El empleo de grandes bloques de ADN genómico clonado en vectores, como los cromosomas artificiales de bacterias (**BACs**), en combinación con FISH (**BAC-FISH**), constituye un método efectivo para el mapeo físico de secuencias únicas de ADN. Sin embargo, el ADN repetitivo presente en las regiones no codificantes de los insertos muchas veces genera patrones de señales dispersas o inespecíficas (background). A pesar de esta dificultad, hasta el momento, BAC-FISH constituye una de las mejores técnicas para correlacionar los mapas genéticos con los físicos, ya que se pueden desarrollar sondas específicas para cada cromosoma o locus en particular.

#### 4 Hibridación in situ genómica (GISH)

Esta técnica se utiliza para identificar los diferentes genomios presentes en un organismo híbrido o aloploiploide. Aunque aplica todos los principios de FISH, difiere de ésta porque las sondas que utiliza están constituidas por **ADN**

**genómico total** y no por un fragmento de ADN de secuencia conocida. El ADN utilizado para construir estas sondas es extraído mediante protocolos estándares y luego es marcado por nick translation o por amplificaciones al azar usando cebadores cortos (random priming). El éxito de GISH depende del grado de diferenciación que presenten los genomios que se pretende analizar. Por lo general, las secuencias codificantes de los genomas de dos especies más o menos relacionadas no varían substancialmente entre sí, por lo tanto, la hibridación diferencial de las sondas en GISH está determinada principalmente por el grado de divergencia del componente repetitivo de los genomas en cuestión. En la medida en que los dos genomas presenten más secuencias en común, mayor será la dificultad para distinguirlos. Una estrategia para mejorar la capacidad discriminante de esta técnica consiste en realizar un **GISH simple** utilizando ADN de un genoma marcado (sonda) junto con una determinada proporción de ADN sin marcar del genoma que no se va a detectar (**sonda fría**). En cambio, cuando la diferencia entre las secuencias de los genomas considerados es alta se puede utilizar **GISH doble o triple**, colocando sobre el preparado mezclas equimolares de las sondas marcadas diferencialmente para cada uno de los genomas que se quiere distinguir en una metafase. Las regiones cromosómicas que hibridan con una sola sonda se visualizan con el color original del fluorocromo, mientras que las que presentan complementariedad con dos o más sondas presentan colores intermedios.

Entre las diferentes aplicaciones de GISH, se destaca el estudio de la **composición genómica** de los **alopoliploides** y de los **híbridos naturales o artificiales**. En estos casos, la técnica se basa en la hibridación del ADN genómico marcado de los probables ancestros o progenitores sobre las preparaciones cromosómicas del material en estudio. La hibridación diferencial de las sondas con sólo un subgrupo de cromosomas en una metafase constituye una clara evidencia de que el material estudiado está compuesto por diferentes genomios. **Mediante el empleo de GISH** se han determinado los progenitores probables de algunas especies aloploiploides cultivadas como el ta-

bacho, la avena, el trigo y el maíz (Fig. 2F). Otra de las ventajas de analizar los híbridos y los alopoliploides con GISH doble o múltiple, consiste en la capacidad que ofrece esta técnica para **detectar translocaciones intergenómicas, las que pueden pasar inadvertidas con los métodos convencionales de análisis genómico. Aplicado a células meióticas, GISH proporciona un método preciso para el análisis de los apareamientos cromosómicos que ocurren en los genomas híbridos. El estudio de la capacidad de los cromosomas de aparearse en meiosis es el procedimiento utilizado tradicionalmente para establecer las relaciones genómicas entre diferentes especies. Sin embargo, en muchas configuraciones metafásicas resulta difícil distinguir los apareamientos ocurridos entre cromosomas homólogos de aquellos ocurridos entre cromosomas homeólogos (i.e. parcialmente homólogos). Mediante GISH es posible “pintar” los cromosomas y, de este modo, determinar el origen genómico de los cromosomas apareados y no apareados, tanto en profase como en metafase de la meiosis I.** La técnica de GISH también constituye una herramienta irremplazable para el desarrollo de programas de mejoramiento vegetal que involucran hibridaciones interespecíficas o intergenéricas ya que permite: 1) confirmar el origen híbrido de la progenie, 2) determinar el origen de los cromosomas en las distintas generaciones, en particular, cuando los híbridos  $F_1$  muestran inestabilidad cromosómica tal como la eliminación cromosómica uniparental, 3) identificar bivalentes interparentales en la meiosis de los híbridos  $F_1$  sexuales y 4) comprobar si los cromosomas recombinantes son transmitidos a las siguientes generaciones. Esta técnica también ha sido utilizada para detectar regiones cromatínicas introgresantes, especialmente aquellas que portan caracteres agronómicos de interés. Un ejemplo de ello es el análisis de una línea élite de cebada introgresada con *Hordeum bulbosum* resistente a roya. Mediante una combinación de GISH y FISH se pudo determinar la localización de los fragmentos cromosómicos de *H. bulbosum* introgresados en el complemento cromosómico de cebada que le conferirían dicha resistencia.

## **5 Mapeo por amplificación directa de secuencias de interés por PCR in situ (PRINS)**

La técnica de PRINS (“primed *in situ*”) es considerada como una alternativa a ISH para la detección de ciertos tipos de secuencias de ácidos nucleicos. Esta técnica involucra la extensión de cebadores diseñados para las secuencias blanco por medio de la *Taq* polimerasa. Durante la extensión, se incorporan nucleótidos marcados con fluorocromos o haptenos a fin de permitir la localización física de las secuencias en forma directa o indirecta mediante protocolos adicionales de detección. Comparada con FISH convencional, PRINS ofrece algunas ventajas tales como: 1) el empleo de cebadores diseñados a partir de un mínimo de información de la secuencia blanco, 2) dichos cebadores pueden hibridar con secuencias de ADN más fácil y rápidamente que las sondas (de mayor tamaño) usadas en FISH, aún en cromatina compacta, 3) no requiere el marcado de sondas y 4) los cebadores pueden ser extendidos aún dentro de las secuencias adyacentes (a diferencia de FISH que depende de una secuencia específica), reforzando la intensidad de la señal para una secuencia corta. Mediante PRINS se han localizado secuencias repetitivas en tándem (secuencias teloméricas, elementos móviles y ADN<sub>r</sub>) en cromosomas de leguminosas (*Vicia faba* y *Pisum sativum*) y de gramíneas (*Hordeum vulgare*, *Avena sativa* y *Lolium multiflorum*). Todos los *loci* encontrados fueron coincidentes con los hallados por FISH. Los resultados demuestran que la técnica de PRINS es apropiada para la localización rápida de repeticiones en tándem sobre cromosomas vegetales cuando la distancia entre los cebadores es pequeña (1 kb) y los blancos son largos (100 – 500 kb). Sin embargo, no se obtienen buenas señales cuando las secuencias repetidas están muy separadas entre sí, como ocurre con el gen de la vicilina. Este gen está repetido muchas veces en el genoma de *Vicia faba*, pero cada copia está separada por una distancia mayor de 12 kb. En estos casos, el empleo de FISH resulta más adecuado dado que se obtienen mejores señales. Por otra parte, PRINS no ha presentado ventajas con respecto a FISH convencional para la detección

de secuencias simples en la mayoría de los casos, aún utilizando PRINS cíclico que involucra de 5 a 10 ciclos de amplificación. Una de las mayores desventajas de PRINS es el alto grado de aparición de señales inespecíficas, probablemente debido a la incorporación de nucleótidos marcados en los extremos 3'OH libres que resultan de roturas espontáneas de los cromosomas durante el procesamiento del material. Estas incorporaciones indeseables pueden disminuirse mediante el tratamiento con ligasa o incorporando 2', 3' - didesoxinucleótidos (ddNTP) antes de PRINS.

En síntesis, esta técnica es útil para la detección rápida de repeticiones en tándem grandes, pero no constituye una alternativa a FISH para la detección de genes de copia simple o para el pintado de los cromosomas en plantas mediante el uso de un cóctel de cebadores cromosoma-específicos.

## 6 FISH de alta resolución

Bajo este concepto se agrupan una serie de técnicas que, utilizando los mismos principios de FISH convencional, permiten hacer hibridaciones sobre cromosomas profásicos o aún sobre núcleos interfásicos. Estas estrategias permiten aumentar tanto la resolución como la sensibilidad de la detección.

- **FISH sobre cromosomas poco condensados:** consiste en el análisis de cromosomas meióticos en paquitene o de cromosomas mitóticos extendidos.

La técnica de **cromosomas paquiténicos** se aplica a células madres del polen fijadas en etanol : ácido acético (3:1) y tratadas con una mezcla de enzimas (celulasa, pectinasa y citohelicasa) que digieren la calosa de las paredes celulares. En general, los cromosomas paquiténicos vegetales son de 7 a 40 veces más largos que aquellos de metafases mitóticas, están compuestos por cuatro hebras de ADN (iguales en los homocigotos), son más accesibles a las sondas (debido a que tienen una estructura cromatínica menos condensada) y presentan marcadores cromosómicos particulares (bloques heterocromáticos y knobs) que permiten individualizarlos con facilidad. Estas

características hacen que los cromosomas paquiténicos constituyan el material de elección para los estudios de FISH de alta resolución, en particular, para identificar secuencias blanco de alrededor de 10 kb y resolver dos *loci* separados por menos de 100 kb.

Por otra parte, la técnica de FISH sobre **cromosomas mitóticos extendidos** utiliza cromosomas metafásicos seleccionados por **citometría de flujo** a partir de meristemas radiculares sincronizados. El estiramiento de los cromosomas se realiza mediante una digestión moderada con proteinasa K, antes de la fijación con etanol-ácido acético (3:1). Mediante este procedimiento, se pueden obtener cromosomas con una longitud de 10 a 100 veces mayor que la original. Esto permite aumentar la resolución de los arreglos de secuencias génicas y repetidas sobre los cromosomas, sin perder la información sobre la orientación relativa con respecto al centrómero y los telómeros. Esta técnica posee una resolución similar a la que utiliza cromosomas paquiténicos y es muy útil en aquellas especies en las que la obtención de buenos preparados meióticos es difícil o en aquellas que poseen cromosomas paquiténicos difíciles de resolver.

- **FISH en núcleos interfásicos:** esta técnica se aplica a núcleos en cortes de tejidos así como a núcleos aislados, y permite incrementar tanto la sensibilidad como la resolución de FISH.

Para el análisis de los **núcleos en tejidos**, los órganos de la planta se fijan con formaldehído o glutaraldehído (este último en baja proporción ya que puede inducir autofluorescencia), luego se cortan con micrótomo y los cortes se montan directamente sobre un portaobjetos. Debido a que la sonda necesita penetrar en el tejido para alcanzar el interior del núcleo, se realizan diversos pre-tratamientos tendientes a incrementar la permeabilidad de los tejidos y, además, se emplean como sondas fragmentos marcados cuyo tamaño no supera los 100-500 pb. El análisis de las hibridaciones se realiza con un microscopio confocal, debido a que este tipo de microscopio permite visualizar las estructuras mediante la reconstrucción tridimensional de las secciones ópticas de la muestra.

Para el análisis de **núcleos aislados**, los tejidos se disgregan en un buffer de estabilización apropiado y la fracción de interés se obtiene por centrifugación y tamizado en mallas con diferentes diámetros de poro. Los núcleos aislados se colocan directamente sobre un portaobjetos y se secan al aire y deshidratan en serie alcohólica. Este procedimiento, que también puede ser aplicado a suspensiones celulares, no presenta mayores problemas de permeabilidad para las sondas. Las regiones de hibridación de las sondas sobre las secuencias de copia simple se observan como señales individuales dentro del núcleo (dos señales en los núcleos diploides), cada una de ellas correspondiente a cada uno de los cromosomas homólogos.

Mediante estas técnicas también se ha estudiado la organización y la disposición de los centrómeros, de los telómeros y de los cromosomas en núcleos interfásicos de diferentes órganos y tejidos de trigo, arroz, tabaco y *Arabidopsis thaliana*. En particular, las mismas resultan muy útiles para el análisis de los cambios que se producen durante el ciclo celular en las diferentes estructuras cromosómicas dependientes de secuencias. También se emplean para el mapeo de secuencias separadas por distancias menores a 500-1000 kb, ya que en los cromosomas metafásicos las señales de las sondas separadas por estas distancias se superponen.

- **FISH en fibras de ADN extendidas (fiber-FISH):** se basa en la hibridación de las sondas sobre fibras extendidas de ADN. Esta técnica involucra los siguientes pasos: 1) se aíslan los núcleos por medio de mallas de nylon de diferentes diámetros de poro, 2) se deposita un pequeño volumen de la suspensión de núcleos en uno de los extremos de un portaobjetos tratado con poli-L-lisina (este compuesto reduce al mínimo la aparición de señales inespecíficas y favorece la extensión de las fibras de ADN), 3) se agrega un buffer de lisis y se realiza la extensión usando un cubreobjetos de manera similar a como se hace un frotis de sangre, y 4) sobre estos extendidos se realizan las hibridaciones siguiendo el protocolo convencional de FISH. Normalmente, se realizan dos o tres

rondas de amplificación con anticuerpos para obtener una buena intensidad de las señales. La principal ventaja de fiber-FISH frente a FISH en núcleos interfásicos es que se logra una mayor sensibilidad (200 pb) y una resolución a nivel genómico de 1 kb. Esta técnica es de gran utilidad para analizar clones solapantes, detectar rearrreglos cromosómicos pequeños, determinar distancias físicas entre genes y su orientación, medir tamaños de *loci* y acelerar el clonado posicional. Sin embargo, su utilidad para el mapeo de secuencias es limitada debido a que sólo permite establecer posiciones relativas, perdiendo la orientación real de las sondas con respecto al centrómero y a los telómeros. Una variante de esta técnica es el empleo de **cromatina estirada**, un procedimiento que involucra la extensión de las fibras cromatínicas sin remover las proteínas, la que resulta muy útil para examinar las interacciones ADN-proteínas.

## 7 Mapas cromosómicos

Estos mapas constituyen una herramienta fundamental para integrar los mapas físicos (basados en el ordenamiento de secuencias y el armado de contiguos) y los mapas genéticos (derivados de los análisis de ligamiento). Los mapas cromosómicos permiten posicionar genes y/o marcadores ligados a éstos, con respecto a marcadores estructurales propios de los cromosomas (por ejemplo, centrómeros, telómeros, bandas heterocromáticas y constricciones secundarias) o a otros marcadores cromosómicos desarrollados por FISH. Esta integración facilita el mapeo y el aislamiento de genes particulares por métodos convencionales y son indispensables para el aislamiento de secuencias por microdissección y microclonado.

Para el clonado posicional de secuencias es importante tener una estimación precisa de la razón kb/cM existente en la región del locus de interés. Esto posibilita la conversión de las distancias genéticas establecidas entre los marcadores y el gen en estudio en distancias físicas. En general, se ha observado que la distancia entre loci en los mapas genéticos (y no el orden) puede diferir mucho de la distancia establecida en los mapas físicos. Los estudios

de los nódulos de recombinación en complejos sinaptonémicos han revelado que la discrepancia entre los mapas físicos y genéticos es el resultado de la distribución no al azar de los eventos de crossing-over a lo largo de los cromosomas. En este marco, el desarrollo de los mapas cromosómicos ha cobrado importancia debido a que, dependiendo de las regiones cromosómicas que se analicen y a la frecuencia de recombinación propia de cada una, los mapas genéticos pueden ser mejores o peores estimadores de las distancias físicas reales.

Por tal motivo, en varias especies modelo se ha puesto especial énfasis en el desarrollo de marcadores cromosómicos por las técnicas de ISH (principalmente BAC-FISH, y Fiber FISH) para poder integrar los grupos de ligamiento genético con cromosomas, y para relacionar las estimaciones de distancia por recombinación con la distancia física real (Ver capítulo de Genómica).

## **8 Análisis de la expresión génica en tejidos por ARN-ISH**

Los protocolos de ISH de ARN resultan apropiados para describir los patrones temporales y espaciales de expresión de un determinado gen a nivel celular o subcelular. En los organismos multicelulares, ISH constituye un complemento de los análisis de northern blotting, RT-PCR y microarrays, en los que la extracción de ARN de los tejidos resulta invariablemente en la pérdida de información espacial. Asimismo, si bien los microarrays constituyen una de las técnicas más poderosas para analizar la expresión de muchos genes simultáneamente, con frecuencia los resultados que se obtienen deben ser verificados por métodos independientes tales como ISH. El análisis de la expresión génica por ISH generalmente se efectúa sobre cortes de tejido, aunque también puede realizarse sobre órganos enteros ("whole-mount ISH", ISH-WISH). La técnica se basa en la utilización de una sonda (generalmente cDNA) marcada con haptenos que se hibridará directamente sobre los cortes de tejido. Las detecciones se realizan con anticuerpos conjugados con fosfatasa alcalina o con fluorocromos.

Para aquellos casos en los que el ARNm a

detectar está poco representado en las muestras, se pueden utilizar protocolos que amplifican la señal mediante el precipitado enzimático de tiramida sobre los sitios blancos (TSA, Tyramide Signal Amplification). La aplicación de TSA en los protocolos estándares de ISH, consiste en una detección inicial de la sonda marcada con anticuerpos o estreptavidina conjugados con una peroxidasa (por ejemplo la de rábano picante, horseradish peroxidase o HRP). Posteriormente, se agrega al preparado tiramida conjugada con algún hapteno o fluorocromo, la cual es precipitada masivamente por la HRP y unida en forma covalente al sitio de detección inicial. El sistema resultante es detectado mediante anticuerpos conjugados con fosfatasa alcalina (para detecciones cromogénicas estándares) o directamente con microscopía de epifluorescencia. Debido a que la tiramida adicionada sólo se deposita en las regiones próximas a la enzima, se logra aumentar significativamente la sensibilidad sin perder resolución. La principal limitante de ARN-ISH radica en que, generalmente, se puede analizar una sola sonda por cada hibridación. No obstante, esta técnica se ha utilizado para comparar los niveles y sitios de expresión de secuencias en numerosas especies de plantas. Uno ejemplo de ello, es el análisis espacial de transcritos expresados diferencialmente en plantas con reproducción sexual y apomíticas de *Paspalum notatum* (Fig 2 G y H).

## **9 FISH cuantitativo para estimar la expresión génica (QUISH)**

A diferencia de la técnica de ARN-FISH, que utiliza anticuerpos sobre cortes de tejidos para la detección de la expresión génica, QUISH se basa en la hibridación de oligonucleótidos gen-específicos marcados con fluorocromos directamente sobre porciones de tejidos u órganos enteros, combinada con el análisis de secciones ópticas realizadas con un microscopio láser confocal de barrido. De este modo, se logra mejorar notablemente la calidad de los resultados debido a que se eliminan los pasos que generan artefactos e impiden realizar una cuantificación precisa. La característica clave de QUISH es la cuantificación de la expresión

génica utilizando como estándar a los genes de mantenimiento celular (genes “housekeeping”, GAPDH y 18S ARNr). El empleo de este estándar permite corregir la variación de la razón citoplasma/vacuola en los diferentes tipos celulares, los efectos ópticos de los tejidos y las señales no específicas. La naturaleza cuantitativa de esta técnica hace posible el análisis de la expresión génica en órganos, a nivel tisular o celular, independientemente de las diferencias debidas a la edad o a los tipos de tejidos. Este método es más rápido que los métodos tradicionales y puede utilizarse para estudiar la localización celular de la expresión de uno o varios genes en un período relativamente corto de tiempo. Asimismo, con QISH es posible investigar los cambios en los niveles de los transcritos durante la ontogenia o, en respuesta a cambios de las condiciones ambientales en diferentes líneas de plantas modelo.

### **10 Identificación de modificaciones cromatínicas por inmunocitogenética**

La identidad celular está determinada por un programa nuclear establecido por un patrón complejo de interacciones entre el ADN y las proteínas. La organización del ADN en la cromatina regula la expresión y el mantenimiento de la información genética (replicación, reparación, recombinación y segregación) en forma dinámica. Las colas N-terminales de las histonas que conforman el núcleo de los nucleosomas están sujetas a diferentes modificaciones pos-traduccionales (acetilación, metilación, fosforilación, ubiquitinación, etc.) las que, conjuntamente con la metilación del ADN, controlan el empaquetamiento de los nucleosomas en arreglos de orden superior y proveen señales para diversos procesos celulares. Aunque las histonas y sus modificaciones son altamente conservadas, la distribución de las mismas en los cromosomas puede variar a lo largo de los procesos de diferenciación y del ciclo celular, así como dentro y entre grupos de eucariotas. Recientemente, se han producido grandes avances en el esclarecimiento del llamado código de las histonas y del intrincado mecanismo de regulación epigenética. La citogenética se encuentra en una posición privile-

giada para realizar estudios de este tipo ya que permite unir estructuras nucleares con características bioquímicas, tales como la actividad transcripcional, la replicación del ADN, las modificaciones de las histonas y la metilación del ADN mediante la inmunodetección de distintos blancos celulares con anticuerpos conjugados con fosfatasa alcalina o con fluorocromos. La preparación del material, la fijación, la digestión de las paredes celulósicas, la permeabilización y la incubación con anticuerpos se realizan de igual modo que en FISH, aunque se omiten los pasos de desnaturalización e hibridación. Las técnicas de inmunocitogenética se emplean para analizar la estructura cromatínica y la disposición de las diferentes moléculas que intervienen en las divisiones celulares. En particular, estas técnicas son útiles para estudiar la distribución de las modificaciones histónicas y la metilación del ADN en relación con la estructura cromatínica en cromosomas y núcleos durante el ciclo celular. Merecen especial mención aquellos análisis relacionados con las modificaciones que indican estructuras cromatínicas abiertas, con potencialidad de transcripción, y las que delimitan estructuras cromatínicas cerradas, con actividad transcripcional reprimida. El advenimiento de las técnicas de citogenética molecular y de inmunocitogenética ha permitido disectar progresivamente el núcleo a nivel de microscopía óptica. La detección individual de complejos proteicos y de secuencias de ARNs y ADN ha revelado la existencia de muchos compartimientos nucleares. Las variaciones en la arquitectura nuclear reflejan la relación existente entre la organización nuclear y la regulación génica. El análisis de dicha arquitectura incluye varios parámetros cuantitativos y cualitativos, tales como el tamaño y forma nuclear, el contenido relativo y la distribución de heterocromatina, los patrones generales de modificación cromatínica, la presencia de compartimientos nucleares (nucleolos, cromocentros, territorios cromosómicos y cuerpos nucleares) y la organización de los cromosomas en el núcleo (ie. la posición y la orientación de los cromosomas en el núcleo) y las diferencias en la condensación cromatínica a nivel local.

## 11 Análisis de la posición de la inserción de los transgenes

Diferentes ensayos han demostrado que la expresión de los transgenes puede variar significativamente entre especies diploides relacionadas, entre líneas transformadas independientes, y aún entre generaciones de las líneas transformadas. Los niveles de expresión de estas secuencias pueden ser afectados por diversos factores, entre ellos, la presencia de numerosas copias de la construcción en uno o varios *loci* y el contexto genómico en que se encuentran las copias. La técnica de FISH permite mapear los transgenes en cromosomas metafásicos, en núcleos interfásicos o en fibras extendidas de ADN y, a su vez, determinar el número de *loci* en que se encuentran los mismos. Para el mapeo se utiliza una sonda marcada correspondiente al transgen y otras sondas que permitan identificar los pares cromosómicos, un genoma o un subgenoma particular (en el caso de híbridos o alopoliploides). Uno de los ejemplos clásicos, en el cual se combinaron diferentes técnicas moleculares y citogenéticas para analizar los efectos del contexto genómico en la expresión de los transgenes fue realizado en tabaco. En el mismo se empleó FISH para mapear los transgenes sobre los cromosomas y GISH para identificar los genomas del alopoliploide. A partir de estos estudios, se ha demostrado que, en general, los transgenes con expresión estable están localizados en las proximidades de las regiones teloméricas y ricas en genes, mientras que los de expresión inestable se localizan en las regiones intercalares o paracentroméricas que son más pobres en genes. Por otra parte, la combinación de FISH con técnicas de inmunocitogenética permite analizar el contexto epigenético de la cromatina circundante a los transgenes y su relación con la expresión de los mismos.

### Lecturas / sitios recomendados

Bennett M.D. 1995. The development and use of genomic *in situ* hybridization (GISH) as a new tool in plant biosystematics. In: Brandham P.E. and Bennett M.D. (eds.). Kew Chromosome Conference IV. Royal Botanic Gardens, Kew, 1995. pp 167-183.

Fernández A. and Daviña J.R. 1991. Heterochromatin and genome size in *Fortunatia* and *Camassia* (Hyacinthaceae). Kew Bulletin, **46**, 307-316.

Fransz P.F., Alonso-Blanco C., Liharska T.B., Peeters A.J.M., Zabel P. and Jong J.H. 1996. High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibers. The Plant Journal, **9**, 421-430.

Küper H., Ort Seib L., Sivaguru M., Hoekenga O.A. and Kochian L.V. 2007. A method for cellular localization of gene expression via quantitative *in situ* hybridization in plants. The Plant Journal, **50**, 159-175.

Laspina N.V., Vega T., Seijo J.G., González A.M., Martelotto L.G., Stein J., Podio M., Ortiz J.P.A., Echenique V.C., Quarín C.L. and Pessino S.C. 2008. Gene expression analysis at the onset of aposporous apomixis in *Paspalum notatum*. Plant Molecular Biology, **67**, 615-628.

Lavia G.I., Fernández A. and Seijo J.G. 2008. Cytogenetic and molecular evidences on the evolutionary relationships among *Arachis* species. In: Sharma A.K. and Sharma A. (eds.). Plant Genome: Biodiversity and Evolution. Volume 1E: Phanerogam-Angiosperm. Science Publishers, Enfield, NH, USA, 2008. pp 101-134.

Moscone E.A., Matzke M.A. and Matzke A.J.M. 1996. The use of combined FISH/GISH in conjunction with DAPI counterstaining to identify chromosomes containing transgene inserts in amphidiploid tobacco. Chromosoma, **105**, 231-236.

Robledo G., Lavia G.I. and Seijo G. 2009. Species relations among wild *Arachis* species with the A genome as revealed by FISH mapping of rDNA loci and heterochromatin detection. Theoretical and Applied Genetics, **118**, 1295-1307.

Schwarzacher T. and Heslop-Harrison P. 2000. Practical *in situ* hybridization. BIOS Scientific Publishers, Oxford, 2000. pp. 203.

Seijo J.G., Lavia G.I., Fernández A., Krapovickas A., Ducasse D. and Moscone E.A. 2004. Physical mapping of 5S and 18S-25S rRNA genes evidences that *Arachis duranensis* and *A. ipaensis* are the wild diploid species involved in the origin of *A. hypogaea* (Leguminosae). American Journal of Botany, **91**, 1294-1303.

Seijo J.G., Lavia G.I., Fernández A., Krapovickas A., Ducasse D., Bertoli D.J. and Moscone E.A. 2007. Genomic relationships between the cultivated peanut (*Arachis hypogaea* - Leguminosae) and its close relatives revealed by double GISH. American Journal of Botany, **94**, 1963-1971.

Solís Neffa V.G. and Fernández A. Karyotypic stu-

- dies in *Turnera sidoides* complex (Turneraceae, Leiocarpaceae). 2002. *American Journal of Botany*, 89, 551-558.
- Stein N., Ponelies N., Musket T., McMullen M. and Weber G. 1998. Chromosome micro-dissection and region-specific libraries from pachytene chromosomes of maize (*Zea mays* L.). *The Plant Journal*, 13, 281–289.
- Terkelsen C., Koch J., Kølvrå S., Hindkjær J., Pedersen S. and Bolund L. 1993. Repeated primed in situ labeling: formation and labeling of specific DNA sequences in chromosomes and nuclei. *Cytogenetics and Genome Research*, 63, 235-237.

# I.-CAPÍTULO 4

## Herramientas Básicas de Ingeniería Genética

Ingrid Garbus, Marisa Gómez y Viviana Echenique

### Introducción

Hasta principios de los años 70, el ADN era la molécula de la célula que planteaba más dificultades para su análisis bioquímico. Excesivamente larga y químicamente monótona, la secuencia de nucleótidos del DNA tan solo podía ser estudiada por caminos indirectos tales como la determinación de la secuencia de proteínas, o del RNA. Actualmente, es posible separar regiones determinadas del ADN, obtener cantidades ilimitadas de esos fragmentos y determinar incluso la secuencia de esos nucleótidos. Estos adelantos técnicos forman parte de la tecnología del ADN recombinante, de cuyo desarrollo han sido pilares fundamentales el conocimiento de las enzimas de restricción, el esclarecimiento de los procesos de

replicación y reparación de ADN así como de la replicación de virus y plásmidos y la síntesis química de secuencias de nucleótidos. De esta manera, puede decirse que la tecnología del ADN recombinante esta integrada por diversas estrategias metodológicas, muchas de las cuales son adaptaciones de procesos naturales de la genética microbiana que han sido estudiados y se conocen en profundidad.

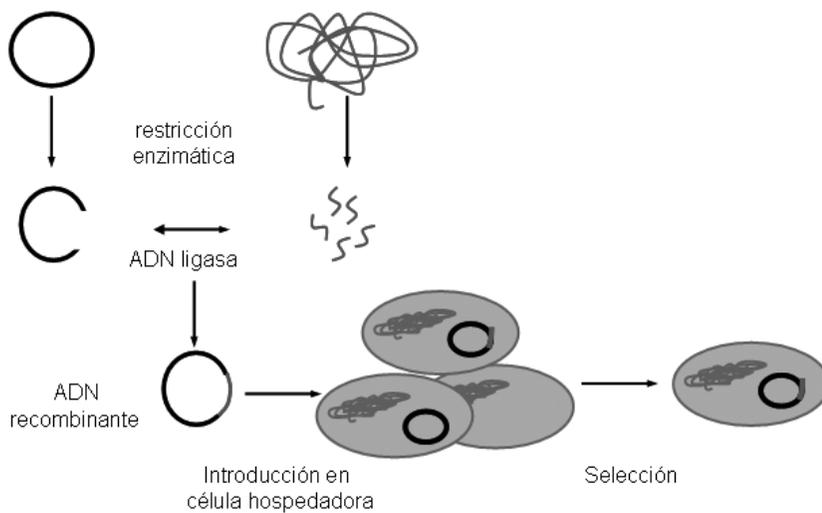
### LA CLONACIÓN MOLECULAR O CLONACIÓN DE GENES

La finalidad de la clonación molecular es la obtención de un determinado fragmento de ADN, que puede corresponder a un gen, inserto en un vector capaz de replicarse, pudiéndose posteriormente amplificar a varios millones de copias, por ejemplo para análisis de secuencia o para obtener grandes cantidades de un producto génico.

El desarrollo de la tecnología del ADN recombinante y la clonación molecular han aportado sofisticados procedimientos que permiten el aislamiento, purificación y replicación de fragmentos específicos de ADN. La clonación de segmentos de ADN, también llamada creación del ADN recombinante requiere esencialmente de las etapas que se enumeran a continuación (Fig. 1)

#### 1. Obtención del ADN a clonar

Según la estrategia experimental involucrada, el ADN a clonar puede tener diferentes procedencias. Puede tratarse de ADN de origen genómico, de un producto de amplificación por PCR, provenir de la síntesis *in vitro*, a partir de una retrotranscripción (ADNc) tomando como molde el ARNm o de la síntesis *in vitro* de secuencias previamente definidas. Cuando el material de



**Figura 1.** Etapas básicas en la clonación de un fragmento de ADN. En primer lugar el ADN a clonar y el vector (plásmido) se cortan con la misma enzima de restricción. Luego se ponen en contacto en presencia de la enzima ligasa para obtener una molécula de ADN recombinante que se introduce en bacterias que multiplicarán el plásmido junto con el fragmento de interés.

partida es el ADN genómico, los fragmentos a ser clonados deben ser escindidos mediante el empleo de enzimas de restricción. En algunas situaciones, particularmente cuando se trata de la clonación de productos de PCR, los fragmentos deben ser aislados en función de su tamaño. Este aislamiento se lleva a cabo a través de electroforesis en geles de poliacrilamida o agarosa.

## 2. Elección del vector de clonación

Un vector de clonación es una molécula de ADN que vehiculizará al fragmento de interés y permitirá su amplificación, resultando de esta manera indispensable para la creación del ADN recombinante. Uno de los principales requisitos para que una molécula de ADN pueda actuar como vector, es la capacidad de autorreplicación que irá acompañada de la replicación del fragmento inserto. También es necesario que los vectores tengan un tamaño que permita su manipulación fuera de la célula. Los más pequeños se manipulan con mayor facilidad que las moléculas de ADN más grandes, que tienden a ser más frágiles. Los vectores de clonación más utilizados derivan del genoma viral o de plásmidos bacterianos. Los componentes esenciales de un vector de clonación son el origen de replicación, un gen marcador responsable de alguna característica fenotípica para la selección y al menos un sitio único de restricción.

## 3. Generación del ADN recombinante

El término ADN recombinante se emplea para hacer referencia al material genético a clonar unido al vector de clonación. Para su obtención es requisito contar con los dos fragmentos de ADN en forma pura, y ligarlos enzimáticamente por acción de la enzima ADN ligasa, en presencia de ATP.

## 4. Introducción del ADN recombinante en el organismo hospedador

Este proceso es realmente el verdadero evento de clonación, en el cual el ADN recombinante es "clonado" para producir varias

copias idénticas. Los organismos hospedadores utilizados en biología molecular deben contener toda la maquinaria genética para la amplificación del ADN recombinante. Puede tratarse de organismos procariotas (bacterias) o eucariotas (levaduras, células de mamíferos cultivadas). En el caso de la utilización de sistemas bacterianos, la introducción se realiza valiéndose de los procesos naturales de la genética microbiana o modificaciones específicas de los mismos. De estos métodos modificados, el más ampliamente utilizado es el de transformación en célula competente (ver vectores de clonación). Otra forma de introducir el ADN en la célula hospedadora, de elección para moléculas de ADN de gran tamaño, es el proceso de electroporación, que consiste en la creación de poros en la membrana celular inducida por descargas eléctricas que permiten la introducción del ADN. Los plásmidos híbridos, introducidos a través de alguno de estos procesos, continuarán replicándose en la célula hospedadora mientras ésta crezca y se divida, siempre que sea mantenida la presión de selección.

## 5. Identificación y selección de las células que contienen al ADN recombinante

En el esquema de transferencia del ADN recombinante al hospedador descrito anteriormente, se asume que durante la transformación, no todas las bacterias reciben una copia del vector híbrido. Es necesario seleccionar aquellas células que han incorporado el vector, a través de la utilización de marcadores génicos presentes en los vectores de clonación. Los más comúnmente empleados son los genes de resistencia a antibióticos y el fragmento funcional de *lacZ*, el gen de *E. coli* que codifica para la enzima  $\beta$ -galactosidasa ( $\beta$ -gal). Estos genes marcadores tienen insertado un sitio múltiple de restricción (polylinker), de manera que cuando los fragmentos de ADN son clonados en el polylinker del gen *lacZ*, anulan la actividad de la  $\beta$ -galactosidasa. Esta enzima hidroliza un compuesto químico denominado X-gal, y se libera un colorante azul relativamente insoluble. El X-gal se incorpora al medio de cultivo en placa de Petri donde desarrollarán las células hospedadoras del ADN recombinante. Si el ADN clo-

nado se insertó en el polylinker y desactivó la  $\beta$ -galactosidasa, no se producirá pigmento azul y las células formarán colonias que se verán blancas en la placa de Petri. Caso contrario las colonias de las células hospedadoras se verán de color azul. Cuando el marcador es un gen de resistencia a los antibióticos, las células se inoculan en medio agarificado que contiene el antibiótico cuya resistencia confiere el vector, de manera que sólo sobrevivirán aquellas células que lo contienen. Luego deberá buscarse específicamente aquellas células que poseen el fragmento de interés. Estos dos marcadores suelen ser utilizados en forma conjunta de manera que solo las colonias portadoras del vector crecerán en medio con antibiótico, y, paralelamente, las colonias que poseen el plásmido recombinante serán de color blanco.

Una vez obtenidas las colonias recombinantes los pasos a seguir incluyen la amplificación de la colonia en medio líquido con antibiótico, purificación del plásmido a partir de la suspensión de bacterias y comprobación de la presencia del inserto esperado. Para la detección del inserto, pueden utilizarse diversas estrategias. La reacción en cadena de la polimerasa (PCR) se puede utilizar cuando la secuencia del fragmento clonado es conocida o para identificar la secuencia de los fragmentos clonados por amplificación con sondas complementarias a la secuencia terminal del vector. La hibridación molecular con una sonda específica marcada puede ser utilizada cuando se conoce la secuencia de interés y se dispone de dicha sonda. En otros casos, se realiza la selección en función del tamaño del inserto, para lo que el plásmido recombinante es digerido mediante enzimas de restricción y posteriormente se realiza una corrida electroforética en gel de agarosa.

## **HERRAMIENTAS Y PROCEDIMIENTOS IMPLICADOS.**

A pesar de que la obtención de ADN recombinante pudo ser explicada a través de cinco pasos sencillos, es muy amplio el conjunto de herramientas y metodologías experimentales que están involucradas en su desarrollo. El objetivo de esta parte del capítulo es introducir

al lector en las herramientas metodológicas de biología molecular más ampliamente utilizadas.

### **1. VECTORES DE CLONACIÓN**

Como hemos mencionado, un vector es una molécula de ADN a la que se unirá el fragmento de interés resultando en el ADN recombinante y, debido a su capacidad de autorreplicación, permitirá la amplificación del fragmento insertado. Los vectores de clonación de última generación son muy prácticos y sencillos de manejar. Incluyen un sitio múltiple de clonación o sitio de restricción múltiple ("polylinker") que es un segmento corto de ADN con sitios de restricción para varias enzimas diferentes, cada uno de ellos único para el vector (Fig. 2). Este sitio suele formar parte del marco abierto de lectura ("ORF") de un gen responsable de alguna característica fenotípica, por lo que resulta sencillo confirmar si tras la restricción se ha obtenido el plásmido híbrido, ya que la inclusión del inserto produce la pérdida de funcionalidad del gen mediante inactivación por inserción.

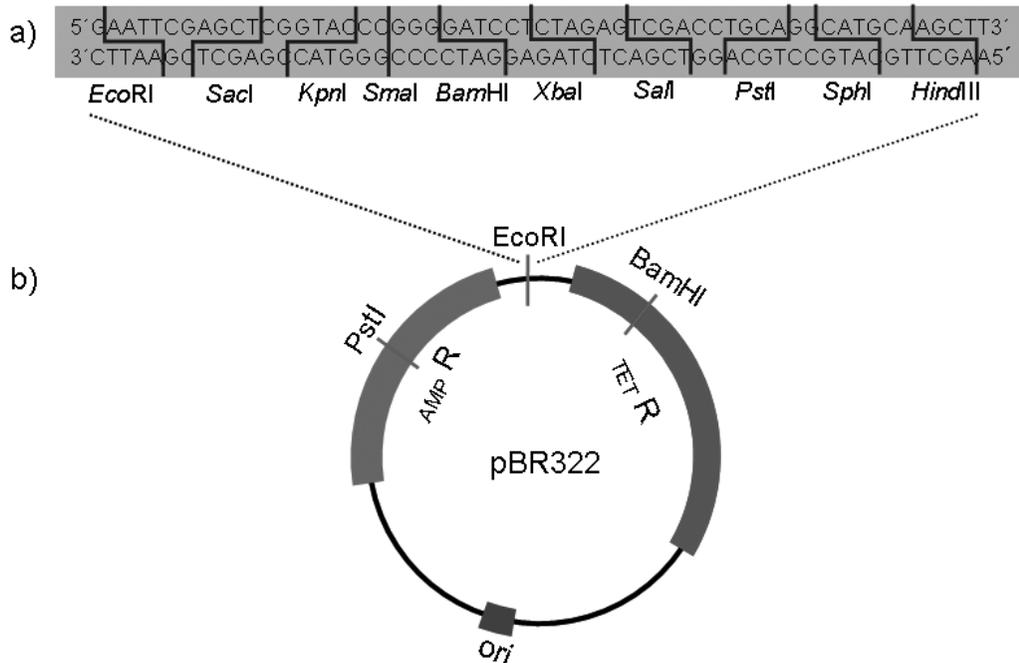
Existen diferentes tipos de vectores de clonación, entre los que se incluyen:

#### **A) Plásmidos:**

Son moléculas de ADN circular, de doble cadena, extracromosómicas, presentes en las bacterias, que poseen propiedades muy útiles como vectores de clonación, a saber:

1. Pequeño tamaño que hace sean fáciles de aislar y manipular.
2. Son circulares, lo que hace que el ADN sea más estable durante su aislamiento químico.
3. Su replicación transcurre independientemente del ADN nuclear en la célula bacteriana.
4. Existen múltiples copias en la célula, dependiendo del plásmido y de la especie hospedadora puede haber de varias a numerosas copias (por ej. 1000 a 3000 copias de pBR322 por célula).
5. La presencia de genes de resistencia a los antibióticos que actúan como marcadores seleccionables facilita la detección y selección de los clones que los contienen.

El tamaño de los insertos que se pueden clonar puede ser considerable, sin embargo si son



**Figura 2.** Elementos genéticos que constituyen un vector de clonación. a) Sitio de restricción múltiple que consiste en un corto segmento de ADN con varios sitios de restricción, cada uno de ellos único para el vector. b) representación esquemática del plásmido pBR322 con el origen de replicación (Ori) y los genes marcadores de resistencia a tetraciclina ( $_{TET}R$ ) y ampicilina ( $_{AMP}R$ ) que contienen los sitios de restricción de *Bam*HI y *Pst*I, respectivamente;

mayores de 10 kb, el plásmido se hace generalmente inestable.

Aunque en el ambiente natural los plásmidos conjugativos generalmente se transfieren por contacto célula a célula, los plásmidos vectores de clonación generalmente han sido modificados a fin de evitar su transferencia por conjugación y así lograr su contención biológica. Sin embargo, en el laboratorio es posible realizar la transferencia a la célula hospedadora por transformación, utilizando choque térmico en presencia de cloruro de calcio o mediante electroporación.

Aunque los primeros plásmidos utilizados existían en forma natural, los vectores de clonación actuales constituyen una generación posterior de vectores construidos *in vitro*, como el plásmido **pBR322** (Fig. 2). En este plásmido, el sitio de restricción de *Bam*HI está dentro del gen de resistencia a la tetraciclina, y el sitio para *Pst*I está dentro del gen de resistencia a la ampicilina. Si se inserta un fragmento de un ADN en uno de estos sitios, la resistencia al antibiótico conferida por el gen que contiene este sitio se pierde (inactivación por inserción). Por

tanto, cuando pBR322 es digerido con *Bam*HI y se liga a un ADN, y luego se aíslan los clones transformados, aquellos transformantes que posean resistencia a la tetraciclina y ampicilina no portan ningún ADN clonado. Aquellas células que continúan siendo resistentes a la ampicilina pero sensibles a la tetraciclina, contienen el plásmido con el fragmento del ADN clonado. Como la resistencia a la ampicilina y a la tetraciclina puede determinarse independientemente en placas con agar, resulta fácil aislar bacterias que contengan los clones deseados y eliminar las células que no los contengan.

## B) Bacteriófagos o fagos:

**Fago  $\lambda$ :** se caracteriza por ser un fago especializado en la transducción (transducción especializada) por lo tanto este fago puede utilizarse también como vector de clonación para la recombinación *in vitro*. Es un vector de clonación particularmente útil porque:

1. Se conoce bien su estructura, genética molecular y funcionamiento

Puede insertar mayor cantidad de ADN que la mayoría de los plásmidos (entre 10 y 20 kb)

El ADN puede ser eficientemente empaquetado *in vitro* dentro de las partículas del fago

Las partículas del fago son mucho más eficientes en la infección de las células hospedadoras (**transfección**) que la transformación.

El fago  $\lambda$  tiene un mapa genético complejo. El tercio central del cromosoma, que contiene genes requeridos para la lisogenia pero no para el ciclo lítico, puede ser escindido por restricción y substituido por ADN foráneo (Fig 3a). El ADN recombinante resultante puede ser empaquetado en las cabezas del fago *in vitro*, pudiéndolas el fago inyectar en *E. coli*, que se replica produciendo colonias bacterianas que contienen los fragmentos a clonar.

Dado que el fago  $\lambda$  silvestre tiene excesiva cantidad de sitios de restricción, no es indicado como vector de clonación por lo que se han construido fagos  $\lambda$  modificados (Charon), en los cuales los sitios de restricción no deseados han sido modificados

**Fago M13** es un fago filamentosos que contiene ADN monocatenario y se replica sin matar a su hospedador. Sin embargo, para poder usar M13 para la clonación es necesario disponer de una forma bicatenaria ya que las enzimas de restricción sólo trabajan sobre ADN bicatenario. El ADN bicatenario de M13 puede obtenerse de células infectadas, donde se encuentra en forma replicativa bicatenaria. La mayor parte del genoma del tipo silvestre contiene información genética esencial para la replicación. Sin embargo existe una pequeña región, llamada secuencia intergénica, que puede ser utilizada como sitio de clonación. Es posible clonar ADN de longitudes variables hasta 5kb sin afectar la viabilidad del fago. El ADN monocatenario de M13 y de sus vectores derivados, ha sido extremadamente útil para secuenciar ADN.

Tanto en el fago  $\lambda$  como el M13 se ha insertado como marcador el gen **lacZ**.

### C) Cósmidos:

Son híbridos entre plásmidos y el fago  $\lambda$  y combinan las ventajas de ambos como vectores. Poseen la habilidad del plásmido para replicarse autónomamente en las células de *E.*

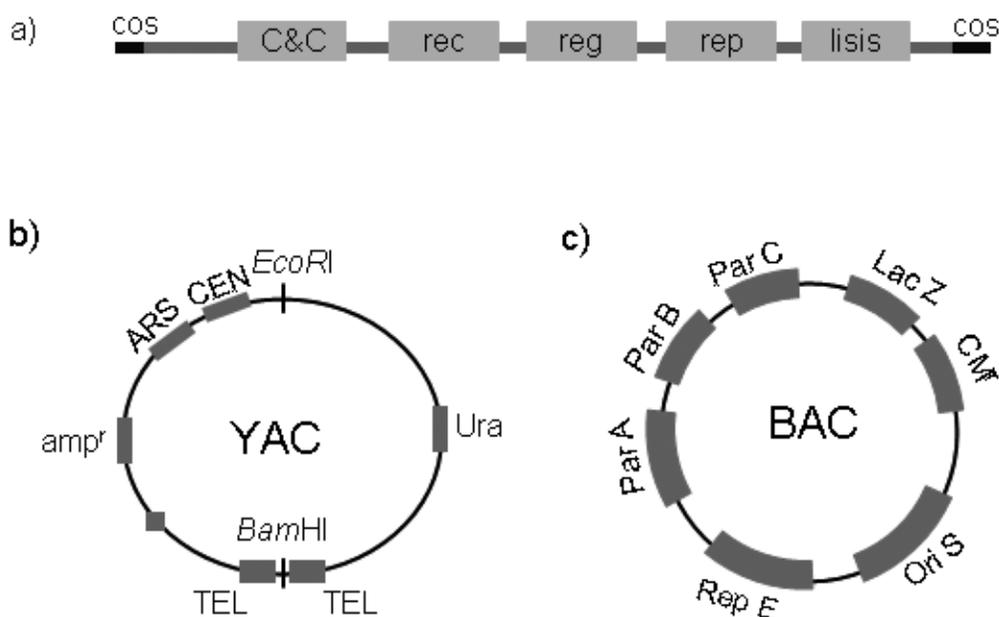
*coli* y la capacidad de empaquetamiento *in vitro* del cromosoma de  $\lambda$  y contienen el origen de replicación y los genes de resistencia a los antibióticos de su plásmido parental. *Cos* proviene de sitio cohesivo (“cohesive site”) en referencia a la terminal de simple cadena de 12 bases complementaria en el cromosoma de  $\lambda$  maduro. El sitio *cos* es reconocido por el sistema de empaquetamiento de ADN de  $\lambda$ , haciendo cortes escalonados que originan los extremos cohesivos complementarios en el cromosoma maduro del fago. La principal ventaja de los cósmidos como vectores es su habilidad para insertar fragmentos de ADN de 35 a 45 kb.

### D) Fásmidos (fagémidos):

Son vectores híbridos entre el fago M 13 y el plásmido pBR323 que contienen los orígenes de replicación de ambos. Normalmente la replicación depende del plásmido, pero cuando una célula que contiene un fásmido se infecta con un fago de tipo salvaje, el origen del fago es responsable de la replicación y se generan copias monocatenarias. Este ADN monocatenario es empaquetado en viriones y puede aislarse fácilmente y ser utilizado para secuenciar. Habitualmente los fásmidos pueden transportar de forma estable un fragmento de ADN clonado mayor que un vector típico derivado de M13.

### E) Cromosomas artificiales:

A partir de la creación de proyectos de secuenciación de genomas completos se gesta la necesidad de obtener vectores que alberguen porciones de ADN de mayor tamaño para efectuar el análisis genómico y obtener mapas físicos de cromosomas. El *primer* sistema de clonado de grandes fragmentos de ADN fue informado en 1987 por Burke y colaboradores, los cromosomas artificiales de levaduras (**YACs**), minicromosomas con la capacidad de contener insertos de ADN de 200-500 kbp (Fig. 3b). Los YACs son vectores de DNA lineares que contienen los elementos esenciales de los cromosomas de la levadura, como por ejemplo el origen de replicación ARS (Autonomously Replicating Sequence). Introducidos en la célula de levadura, los YACs se replican y segregan junto con el complemento cromosómico



**Figura 3.** Vectores de clonación. Se muestran los elementos genéticos que constituyen los vectores de clonación. a) Fago lambda. El cromosoma se organiza de acuerdo con las proteínas codificadas por los genes: C&C: cabeza y cola, genes de proteínas de la cápside, de unión cabeza-cola y de estabilización y empaquetamiento de ADN, rec: genes de recombinación, proteínas para la integración y escisión del ADN del fago en el cromosoma de la bacteria hospedadora, reg: genes regulatorios, para el cambio entre los distintos ciclos del fago, rep: región de replicación, con el origen de replicación y los genes O y P, lisis: genes que posibilitan la lisis celular, b) Cromosoma artificial de levadura (YAC). ARS: origen de replicación, CEN: centrómero, TEL: telómeros, URA: gen marcador seleccionable para el desarrollo en medio con uracilo. Amp: gen marcador seleccionable en medio con ampicilina. El ADN es clonado en el sitio *Bam*HI. c) Cromosoma artificial de bacterias (BAC). Deriva del factor bacteriano F. Su esqueleto contiene regiones esenciales que le otorgan estabilidad y bajo número de copias en cada bacteria, *parA*, *parB* y *parC*. *oriS*: origen de replicación unidireccional. CM<sup>r</sup>: gen marcador seleccionable en medio con cloranfenicol, repE: gen de proteína autoregulatoria esencial para la replicación desde *oriS*. pBeloBAC11, el vector BAC mas utilizado, tiene tres sitios únicos de clonado, *Hind*III, *Bam*HI y *Sph*I dentro del gen marcador *lacZ*.

habitual. El marcador seleccionable suele ser un gen de tipo silvestre que le confiere la capacidad prototrófica a la célula hospedadora. El mas comúnmente utilizado es URA3+, que confiere a las auxótrofos URA3- la capacidad de desarrollarse en un medio de cultivo sin uracilo. Durante la meiosis los YACs se aparean con cualquier YAC homólogo presente, resultando en una alta frecuencia de recombinación entre los fragmentos clonados. Esta alta frecuencia de quimerismo lleva a predicciones inexactas en los mapas físicos moleculares y,

junto con la falta de familiaridad en el manejo de levaduras entre biólogos moleculares, limitaron el uso de los YACs. Para superar estas dificultades fueron desarrollados los cromosomas artificiales de bacterias (BACs) constituyen un sistema de clonación que mantiene un bajo número de copias en cada bacteria, garantizando una recombinación mínima entre diferentes fragmentos de ADN y son de sencilla manipulación. Los **BACs** se basan en el plásmido F de 7kb de *E. coli* y pueden llevar insertos de 300 kb aunque el promedio es de 100

kb (Fig. 3c). Contienen como elementos esenciales una copia del origen de replicación de *E. coli*, un marcador de selección que suele ser un gen de resistencia a antibióticos y un sistema de recombinación sitio-específico (*cre-lox*), cuyo rol consiste en forzar la circularización de grandes insertos. Los cromosomas artificiales derivados del fago P1 (**PACs**) permiten la clonación de fragmentos de 80-100 Kpb y, al igual que los BACs, presentan bajo grado de quimerismo. El fago P1 se replica primero como plásmido de bajo número de copias y luego como un largo segmento de ADN compuesto por múltiples copias de la secuencia repetida (concatemero), que es empaquetado en la cabeza del fago por un mecanismo similar al del fago lambda. Las propiedades de los clones PAC permiten la iniciación de empaquetamiento en las etapas de clonación, escisión de las secuencias de alto número de copias después de la infección bacteriana y la inducción de la replicación inmediatamente antes la amplificación del ADN. Utilizan al igual que los BACs el sistema *cre-lox*.

### **Introducción del vector en la célula hospedadora**

En el caso de la utilización de sistemas bacterianos, la introducción del vector en la célula hospedadora se realiza valiéndose de los mecanismos naturales de la genética microbiana dado que en los procariontes existen mecanismos de intercambio genético, que permiten tanto la transferencia de genes como la recombinación del inserto, a través de uno de los siguientes procesos genéticos:

**Transformación:** es un proceso por el cual el ADN libre del medio ambiente se incorpora en una célula receptora, trayendo aparejado un cambio genético. Una cadena de este ADN libre, llamado donador, es captada y puede transformar genéticamente la célula receptora por recombinación con una región homóloga del cromosoma. El movimiento de las moléculas de ADN a través de la membrana y dentro del citoplasma de la célula receptora es un proceso activo, que requiere energía. Las células que son capaces de tomar una molécula

de ADN y ser transformadas se denominan competentes. Estas células secretan el factor de competencia, que es una pequeña proteína que induce la síntesis de otras 8 a 10 nuevas proteínas requeridas para la transformación. La transformación es un proceso que ocurre sólo en algunas especies bacterianas.

**Transducción:** se refiere a la transferencia del ADN de una célula a otra por medio de un fago, denominación utilizada para los virus cuando se trata de hospedadores bacterianos. Esta transferencia de genes puede ocurrir de dos maneras:

a) **transducción generalizada:** una fracción del ADN celular, que puede proceder del cromosoma o del plásmido del hospedador, durante el ensamblaje del virus pasa a formar parte del ADN de la partícula vírica madura, reemplazando el genoma del fago. Una vez que son liberados, estos fagos anormales pueden pegarse a otras células y donar ADN, pero no dan lugar a lisogenia ni a lisis dado que no pueden replicarse independientemente por carecer de gran parte del ADN fágico. Si los genes del donador no sufren recombinación homóloga con el cromosoma de la bacteria de la célula receptora se perderán.

b) **transducción especializada (restringida):** ocurre sólo en algunos virus atemperados, es decir, aquellos que se integran en el genoma como parte de su ciclo biológico. En este mecanismo el ADN de una región específica del cromosoma del hospedador se integra directamente en el genoma del fago, reemplazando algunos de sus genes. El genoma fágico recombinante puede empaquetarse seguidamente en una cabeza de fago y el resto del fago puede ensamblarse normalmente. Este fago puede carecer de la capacidad de replicar, no obstante puede inyectar su ADN, que contiene además algunos genes de la célula receptora a una nueva célula receptora. Puede también ocurrir recombinación homóloga, sin embargo, como el ADN donador bacteriano está formando parte del genoma de un fago atemperado se presentan dos posibilidades: que el ADN se integre en el cromosoma del hospedador durante la lisogenización, o que el ADN se replique en el receptor como parte de

una infección lítica. En ambos casos, la partícula vírica transductora es normalmente defectiva como virus, ya que los genes víricos han sido reemplazados. Si bien no todos los fagos tienen la capacidad de transducir, ni todas las bacterias son transducibles, el fenómeno está lo suficientemente extendido como para asumir que desempeña un importante papel en las transferencias genéticas que se producen en la naturaleza.

**Conjugación:** mecanismo de transferencia genética mediada por un plásmido que requiere contacto de célula a célula. Ciertos plásmidos y transposones conjugativos confieren a sus células hospedadoras la capacidad de transferir ADN a otras células por conjugación. La conjugación requiere una célula donadora, que contiene un tipo particular de plásmido conjugativo, y una célula receptora que carece de él. El proceso de conjugación, presenta características diferenciales según se trate de bacterias Gram positivas o Gram negativas. En el primer caso, las células receptoras secretan un péptido de pequeño tamaño que provoca que la célula donadora sintetice un componente adhesivo en la superficie celular. El ADN plasmídico es posteriormente transferido desde el donador a la célula receptora adherida. Estos cruzamientos pueden tener lugar en medios líquidos. En el segundo caso, el plásmido codifica, entre otras proteínas, las subunidades de la proteína del pelo sexual, mediante el cual se realiza el contacto. El pelo constituye un puente de conjugación a través del cual pasa el ADN permitiendo el apareamiento específico. Durante la conjugación, pueden resultar movilizados otros elementos genéticos, como grandes bloques del cromosoma del hospedador u otros plásmidos.

**Transposición conjugativa:** es una conjugación independiente del plásmido, un tipo de conjugación facilitada por un transposón conjugativo, que pueden encontrarse tanto en el cromosoma como en los plásmidos. El contacto entre las bacterias donadora y receptora induce la escisión del transposón en el donador: en cada extremo del transposón (integrado), se producen un par de cortes alternos de modo que al escindirse una de las cadenas en cada extremo del transposón porta con ella

(generalmente) seis nucleótidos de la molécula hospedadora. (Estas secuencias terminales de simple cadena se denominan secuencias acoplantes). El transposón escindido se circulariza entonces como resultado del apareamiento de entre las secuencias acoplantes. Parece probable que sólo se transfiera una simple cadena al receptor, y que la cadena complementaria se sintetice, en el receptor, antes de la inserción.

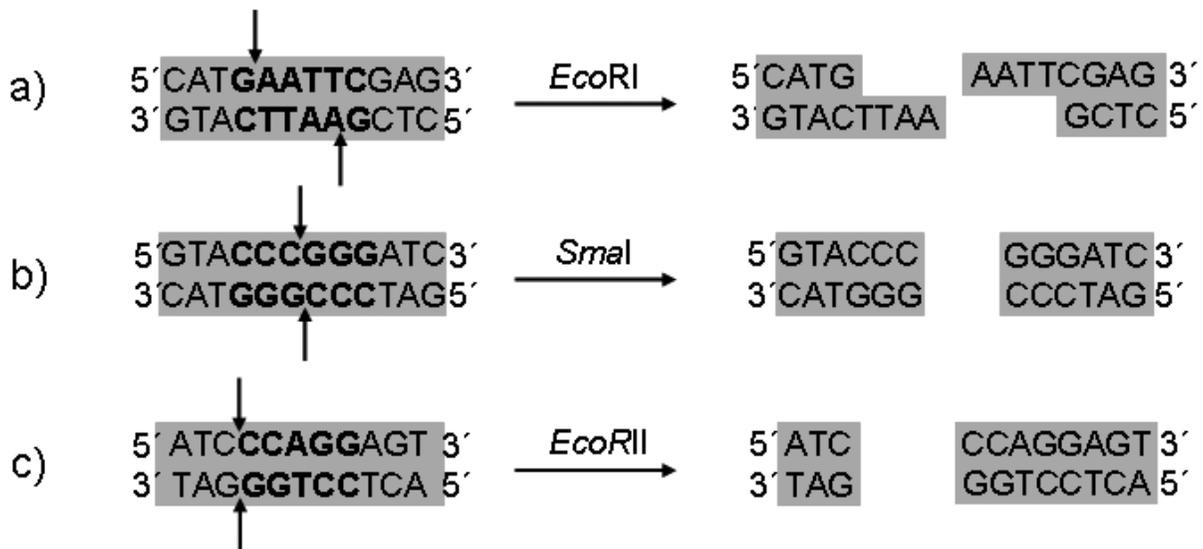
*Las estrategias de introducción de vectores en células hospedadoras utilizadas en las técnicas in vitro del ADN recombinante, han sido desarrolladas en base a modificaciones de estos procesos naturales de transferencia de material genético en procariontes.*

## 2. ENZIMAS UTILIZADAS EN LA GENERACIÓN DE ADN RECOMBINANTE

El requisito primordial para la clonación de un gen o segmento de ADN de interés es poder contar con dicho fragmento en forma individual, aislado desde su entorno génico. Para ello, la tecnología del ADN recombinante cuenta con una herramienta indispensable, las endonucleasas de restricción. Estas enzimas reconocen secuencias específicas de ADN de 4 a 8 bases de extensión y escinden los enlaces fosfodiéster del material genético en dichos sitios, denominados sitios de restricción. Para actuar como sustrato de estas enzimas, el ADN debe hallarse como doble hebra.

Son extraídas de organismos procarióticos, donde actúan como un mecanismo de defensa para degradar material genético extraño que entra en la célula, proveniente principalmente de fagos. Para diferenciar el ADN propio del extraño, las bacterias tienen la capacidad de metilar su propio ADN inmediatamente después de la replicación, haciéndolo de este modo irreconocible por las endonucleasas.

Las endonucleasas se nombran en relación a las bacterias de las que fueron aisladas por primera vez, considerando que 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 el orden en que fue identificada la enzima en esa cepa. Según este criterio, la enzima *EcoRI*, representa a la primer endonucleasa



**Figura 4.** Corte del ADN por enzimas de restricción. Reconocimiento y escisión de secuencias palindrómicas originando: a) extremos cohesivos o b) extremos romos. En este caso puede utilizarse la enzima transferasa terminal para la incorporación de extremos cohesivos. c) reconocimiento y corte de secuencias no palindrómicas. Las flechas indican los sitios de corte mientras que las secuencias reconocidas se señalan en negrita.

aislada en el género *Escherichia*, especie *coli*, cepa RV 13. Se utiliza el término isoquímicos para las enzimas que reconocen la misma secuencia de ADN pero han sido obtenidas de diferentes especies de bacterias.

Pueden diferenciarse tres tipos de enzimas de restricción. Las de tipo I y III escinden en sitios alejados de la secuencia de reconocimiento, requiriendo de ATP para realizar el traslado del sitio de reconocimiento al de acción y tienen actividad de restricción y metilación simultáneamente. En cambio, las de tipo II cortan en el sitio de reconocimiento, tienen actividad de restricción exclusivamente y reconocen secuencias palindrómicas (secuencias que se leen igual en ambas direcciones). Estas últimas son las utilizadas para el clonado de ADN, aunque tienen otras aplicaciones como la construcción de mapas de restricción de un plásmido o bacteriófago, fragmentación del ADN genómico para su separación por electroforesis con aplicación en técnicas como "Southern Blot" o fingerprinting o la generación de fragmentos para ser usados como sondas.

Las escisiones realizadas por las endonucleasas de tipo II, pueden resultar en extremos romos o cohesivos. En el primer caso, se producen cortes escalonados que crean colas cortas de cuatro bases de cadena simple en cada extremo del fragmento. Estas colas tienden a asociarse con una cadena complementaria por apareamiento de bases (Fig. 4a), a secuencias complementarias de otro fragmento con colas generadas por la misma enzima. Cuando se generan extremos romos, el ADN es clivado en el centro de la secuencia de reconocimiento, en el mismo punto en ambas cadenas, no quedando bases desapareadas en los extremos por lo que no presentan tendencia a unirse con otros fragmentos por complementariedad (Fig. 4b).

Aunque por su especificidad y versatilidad las endonucleasas de restricción son las enzimas las más renombradas, la tecnología del ADN recombinante no sería factible sin la participación en el proceso de otras enzimas de roles diversos (Tabla 1).

Enzima	Acción	Aplicación
<b>Endonucleasas de tipo II</b>	<b>Clivaje sitioespecifico del ADN</b>	<b>Generación de fragmentos de ADN</b>
<b>Ligasa de ADN</b>	Unión de dos fragmentos de ADN con extremos romos, en presencia de ATP.	Forma enlaces covalentes entre el extremo 5' de una cadena polinucleotídica y el extremo 3' de otra cadena
<b>Transcriptasa reversa</b>	Síntesis de una hebra de ADNc utilizando como molde ARN.	Construcción de bibliotecas de cDNA. Amplificación por PCR
<b>Polinucleotido kinasa</b>	Incorporación de un grupo fosfato al extremo 5' de un polinucleótido.	Permite ligado con otro polinucleótido o incorporación de fosfato marcado.
<b>Transferasa Terminal</b>	Adición de colas homopoliméricas al extremo 3'OH de ADN doble hebra lineal	Genera extremos cohesivos oligo dT u oligo dA
<b>Exonucleasa III</b>	Remoción de nucleótidos del extremo 3' de un ADN doble hebra.	Expone los extremos 5'.
<b>Exonucleasa del fago I</b>	Remoción de nucleótidos del extremo 5' de ADN doble hebra.	Expone los extremos 3'.
<b>Fosfatasa alcalina</b>	Remoción de fosfatos terminales de extremos 5'	Impide autoligado de vectores. Permite la incorporación de fosfato 5' marcado.
<b>Fragmento klenow</b>	Fragmento de una enzima ADN polimerasa, con actividad DNA polimerasa y exonucleasa en dirección 3'→5'.	Genera extremos romos desde el extremo 3', por síntesis complementaria o clivaje de nucleótidos desapareados.

**Tabla 1.** Algunas enzimas utilizadas en la tecnología del ADN recombinante

### 3. SEPARACION DE FRAGMENTOS DE ADN: ELECTROFORESIS EN GEL

La electroforesis es una técnica que permite la separación de moléculas en función de su diferente movilidad en un campo eléctrico, siendo el parámetro esencial su diferencia en carga eléctrica. Sin embargo, determinados soportes como los geles de agarosa o poliacrilamida, ofrecen una resistencia notable al avance de las moléculas. Aunque el parámetro que rige el avance es la relación carga/masa, como en los ácidos nucleicos la carga eléctrica es proporcional a la longitud en nucleótidos, la relación carga/masa es la misma para todas las moléculas. Como resultado, el efecto de retardo del gel depende del tamaño molecular por lo que la electroforesis separa los fragmentos de ácido nucleico según su tamaño o longitud. La agarosa es un polisacárido, cuyas disolu-

ciones (0,5 - 2%) forman un gel semisólido al enfriarse, constituido por una matriz o trama tridimensional de fibras poliméricas, que retarda el paso de las moléculas de ácido nucleico a su través, en función de su tamaño. Los geles de poliacrilamida se forman por la polimerización de acrilamida y bis-acrilamida, catalizada por el agente oxidante persulfato de amonio y la amina terciaria TEMED como agente reductor. La variación de la concentración de acrilamida posibilita la modificación controlada el tamaño del poro. Estos geles presentan menor rango de separación que los geles de agarosa, pero mayor poder de resolución, pudiendo discriminar fragmentos de ADN con diferencia de tamaño de sólo una base. Se usan además para la separación y caracterización de proteínas.

Para establecer el tamaño de fragmentos separados se utilizan marcadores moleculares, que consisten en una mezcla de fragmentos de

ADN de tamaño conocido, a partir de los que se puede inferir el tamaño de la muestra analizada. Se establece una curva de calibrado en la que se consigue la linealidad graficando el tamaño en función de la inversa de la movilidad o el logaritmo del tamaño respecto de la movilidad.

Dentro de las aplicaciones más comunes de la electroforesis en gel en biología molecular, podemos mencionar chequeo de la amplificación de fragmentos de PCR, identificación de polimorfismos y la identificación de genes mediante el establecimiento de sus patrones de restricción.

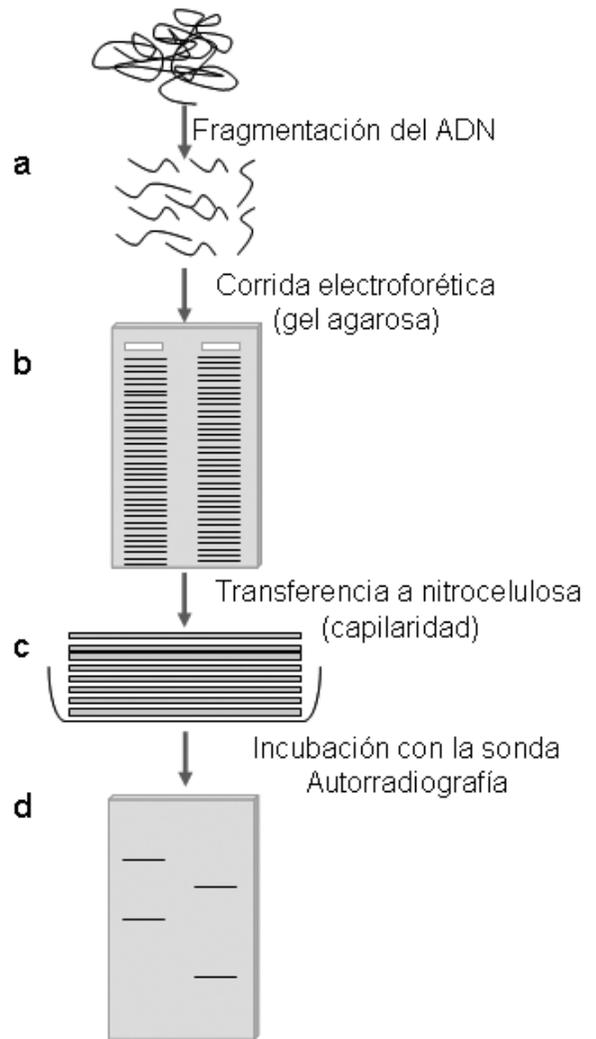
#### 4. HIBRIDACIÓN. ANÁLISIS MOLECULAR DE ADN, ARN Y PROTEÍNAS

La hibridación es la construcción artificial de un ácido nucleico bicatenario por apareamiento (emparejamiento) de bases complementarias de dos ácidos nucleicos monocatenarios. Para que exista la formación de híbridos estables debe haber un alto grado de complementariedad. Se define formalmente como sonda ("probe") a un fragmento determinado de ARN o ADN marcado química o radioactivamente, utilizado para localizar determinadas secuencias de ácidos nucleicos mediante hibridación.

##### A) Análisis de ADN por hibridaciones Southern Blot

Los procedimientos utilizados, para separar ácidos nucleicos y proteínas se rigen por el mismo principio, pero involucran algunas diferencias de técnicas debidas a las características particulares de cada clase de moléculas.

En 1975, E. M. Southern, publicó un nuevo procedimiento que permitió a los investigadores ubicar los genes y otras secuencias de ADN sobre fragmentos de restricción separados por electroforesis en gel. La principal característica de esta técnica es la transferencia de las moléculas de ADN que han sido digeridas por endonucleasas de restricción y separadas por electroforesis en gel a una membrana de nylon o nitrocelulosa. Esta transferencia se denomina Southern Blot, en honor al científico que desarrolló la técnica (Fig. 5). Para realizarla, el ADN es desnaturalizado (abierto en



**Figura 5.** Método de hibridación de Southern. a) Es-cisión enzimática o mecánica del ADN; b) inclusión del ADN en el gel de agarosa y corrida electroforética; c) transferencia de las moléculas de ADN desde el gel de agarosa hacia un filtro de nitrocelulosa por capilaridad, previa desnaturalización del ADN; d) hibridación con la sonda monohebra, seguida de lavados y detección por autorradiografía.

sus dos cadenas), ya sea antes o durante la transferencia, mediante tratamiento alcalino. Una vez completada la transferencia, el ADN es inmovilizado sobre la membrana por exposición a elevada temperatura o a radiación UV. Una sonda de ADN conteniendo la secuencia de interés es entonces incubada con el ADN inmovilizado. La sonda sólo va a hibridar con moléculas de ADN que contienen la secuencia

de nucleótidos complementaria a su secuencia. El excedente de sonda no hibridada se elimina mediante repetidos lavados de la membrana. Las sondas pueden marcarse por diferentes métodos: con elementos radioactivos, o compuestos que producen color o luminiscencia, etc. Luego de la hibridación, la membrana se somete a diferentes procedimientos para poner en evidencia la sonda, de acuerdo al método con el que haya sido marcada.

### **B) Análisis de ARN por transferencia e hibridación: Northern Blot**

De manera similar al tratamiento realizado con las moléculas de ADN, las moléculas de ARN también pueden ser separadas por electroforesis en geles de agarosa, transferidas a membranas y analizadas. La transferencia de ARN se denomina Northern blot, en reconocimiento al hecho de que el procedimiento es la imagen de espejo de la técnica Southern blot.

Ambos procedimientos son esencialmente idénticos. Sin embargo, las moléculas de ARN son muy sensibles a la degradación por enzimas ARNasas. Por lo tanto, se debe tener mucha precaución para evitar la contaminación de los materiales con estas enzimas. Además, la mayoría de las moléculas de ARN contienen estructuras secundarias (producidas por apareamiento complementario intracadenas), por lo tanto deben mantenerse desnaturalizadas durante la electroforesis para poder separarlas en base a su tamaño. La desnaturalización se provoca incorporando formaldehído u otro químico desnaturalizante a la solución tampón ("buffer") utilizada en la electroforesis. Después de la transferencia a la membrana apropiada, las moléculas de ARN pueden hibridar con sondas de ADN o ARN.

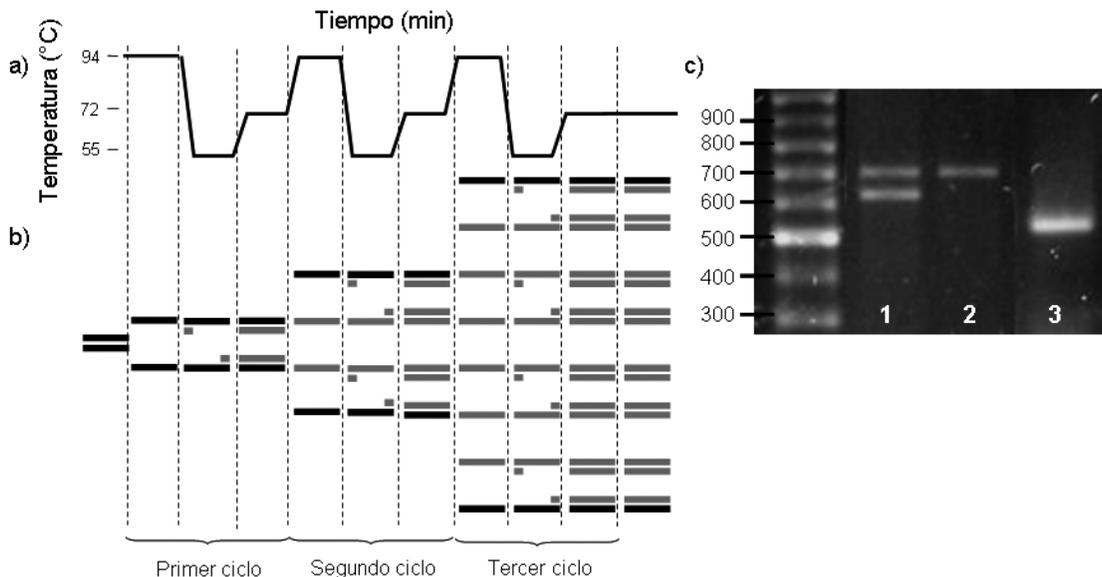
### **C) Análisis de proteínas por transferencia e inmunodetección o Western Blot**

La electroforesis en gel de poliacrilamida es una herramienta muy importante para la separación y caracterización de proteínas. Debido a que muchas de las proteínas están compuestas por dos o más subunidades, las cadenas polipeptídicas individuales son separadas por electroforesis en presencia del detergente dodecil sulfato de sodio (SDS), que actúa como

agente desnaturalizante de las proteínas. Los polipéptidos separados por electroforesis también pueden ser transferidos del gel a una membrana de nitrocelulosa y pueden ser detectados utilizando anticuerpos específicos. Esta transferencia de proteínas se denomina Western blotting y se realiza utilizando una corriente eléctrica para trasladar las proteínas desde el gel a la superficie de la membrana. Después de la transferencia, la proteína de interés es identificada mediante la unión a un anticuerpo específico que está conjugado (marcado) ya sea con isótopos radioactivos, que permiten la detección por autoradiografía, o con enzimas que producen un producto visible cuando se adiciona el sustrato correspondiente.

## **5. AMPLIFICACION DEL ADN: REACCIÓN EN CADENA DE LA POLIMERASA (PCR)**

La PCR constituye una tecnología poderosa que involucra la síntesis enzimática *in vitro* de millones de copias de un segmento específico de ADN. La reacción se basa en la hibridación de un par de oligonucleótidos, diseñados en base a la secuencia de interés y sintetizados artificialmente, a la secuencia de ADN (Fig 6). Estos oligonucleótidos funcionan como iniciadores o cebadores ("*primers*") que delimitan la secuencia a ser amplificada. La primera etapa de la reacción involucra la desnaturalización del ADN de doble cadena, por elevación de la temperatura a 92-95°C. Posteriormente comienzan los ciclos de PCR, que comprenden: desnaturalización de la doble cadena, unión por complementariedad del cebador al ADN de simple hebra y replicación del ADN a partir del cebador, catalizada por la enzima *Taq* polimerasa. La temperatura de unión de los cebadores al ADN de cadena simple depende de la secuencia y tamaño del oligonucleótido, así como de la estrategia específica de trabajo, comprendiéndose en términos generales entre los 35-60°C. La fase de síntesis de ADN complementario se lleva a cabo a la temperatura óptima para que la *Taq* polimerasa realice la replicación a partir de cada extremo 3' de los cebadores. Se trata de una ADN polimerasa termo-resistente aislada de *Thermus aquaticus*, bacteria que vive en fuentes termales.



**Figura 6.** Reacción en cadena de la polimerasa (PCR). a) Representación de una reacción de PCR de tres ciclos. En cada ciclo se suceden tres etapas: i- desnaturalización por calor del ADN (94°C) para separar las hebras, ii- unión por complementariedad de los cebadores a las monohebras (en este ejemplo 55°C), iii- reacción de polimerización de por la enzima TAQ polimerasa, dando la hebra complementaria al molde de ADN (72°C); b) Comportamiento de una fragmento de ADN blanco de los cebadores, a través de los mencionados ciclos de PCR. En cada ciclo, se incrementa la cantidad de ADN blanco, siguiendo la relación  $2^n$ , donde n es el número de ciclos. En el ejemplo, con tres ciclos de PCR, se obtienen  $2^3$  copias. c) Verificación de un producto de PCR sobre un gel de agarosa. En la calle de la izquierda se muestra el marcador de peso molecular. En la primera calle, se observan dos fragmentos de 700 y 650 pares de bases (bp), respectivamente, sugiriendo que el cebador ha encontrado complementariedad en más de un sitio en el genoma. Las calles 2 y 3 presentan banda única de 700 y 550 bp aproximadamente en cada caso.

Este ciclo es repetido por algunas decenas de veces y, como el producto de cada polimerización sirve como molde para el siguiente, en cada ciclo se duplica la cantidad de producto sintetizado. Es necesario tener en cuenta que, además de la enzima, los cebadores y el ADN de interés, para que se produzca la amplificación es necesaria la presencia de un medio tamponado y desoxinucleótidos (dNTPs) y cloruro de magnesio que actúa como cofactor de la enzima. El resultado de la reacción es un fragmento de ADN de doble cadena cuyos extremos corresponden a los extremos 5' de los cebadores y su tamaño a la distancia entre los mismos. A pesar de que se forman moléculas más largas a partir del molde original en cada ciclo, se acumulan solo a una tasa lineal y no contribuyen significativamente a la masa final de secuencia blanco.

Después de apenas 20 ciclos se logra más de un millón de veces la cantidad inicial de la secuencia de interés. Esta escala de amplificación permite, por lo tanto, iniciar el proceso con cantidades mínimas de ADN (del orden de pico o nanogramos) y terminar la reacción con grandes cantidades de una secuencia de interés. La versatilidad de esta reacción es enorme y la combinación de la PCR y la secuenciación constituye una poderosa herramienta para el análisis de genes.

Como hemos mencionado, la técnica de PCR es de suma utilidad para amplificar ADN a partir de fragmentos clonados en distintos vectores, donde los cebadores son complementarios a los sitios del vector que flanquean el fragmento inserto. La versatilidad de la PCR incluye la amplificación a partir de ADN de material embebido en parafina por varios años,

de material momificado, de restos fósiles, etc. Dentro de los usos mas frecuentes, podemos mencionar la detección de enfermedades genéticas, determinación del sexo en embriones humanos, clonación de genes, mutagénesis *in vitro* y mapeo y secuenciación de genomas.

#### **A) RT-PCR**

Se trata de una reacción de PCR, secundaria a la transcripción reversa del ARNm, catalizada por la enzima transcriptasa reversa (RT). Esta enzima es capaz de sintetizar ADN utilizando ARN como molde y forma parte de la batería enzimática de los retrovirus, virus de ARN que requieren la síntesis de ADN para ser integrados en el genoma del hospedador. El esquema general de la RT-PCR consiste en un primer paso de síntesis de ADN complementario (ADNc) de cadena simple tomando como molde el ARN. En esta etapa, llamada síntesis de la primera hebra (first strand synthesis) se utilizan cebadores que pueden ser específicos para el gen de interés, hexámeros de secuencia al azar u oligodT, según se busque sintetizar sólo el ADNc específico o los ADNc correspondientes a todos los ARNs del tejido en cuestión. Por lo tanto, mientras que el uso de primers específicos aumenta la especificidad, los otros dos tipos de cebadores maximizan el número de moléculas de ARNm que pueden ser analizadas en una muestra pequeña de ARN. El uso de hexámeros al azar, puede ocasionar distorsiones por exceso en el número de copias de un determinado ARNm, comparado con los cebadores específicos. La segunda etapa consiste en subsecuentes ciclos de PCR con los cebadores específicos para la secuencia de interés, lo que permite obtener cantidades apropiadas del ADN para diversas manipulaciones genéticas. Esta estrategia conjunta de retrotranscripción y PCR puede ser utilizada para el estudio de ARNm a nivel de una célula individual. Los principales alcances de esta técnica incluyen la determinación de la presencia o ausencia de un transcripto, estimación del nivel de expresión y para el clonado de ADNc sin la construcción de una genoteca.

*La técnica de PCR por transcriptasa reversa es denominada como "RT-PCR", la cual desafortunadamente, puede ser confundida con la PCR de Tiempo Real, ya que ambas tienen la misma*

*sigla (RT-PCR). De lo expuesto es obvio, la necesidad de ser específicos en las denominaciones, sobre todo cuando se combinan ambas técnicas: "Real Time de RT – PCR"*

#### **B) PCR cuantitativa (qPCR) o PCR en tiempo real (RT-PCR).**

Una variante de la PCR convencional que se basa en la detección y cuantificación simultánea de la fluorescencia emitida por los productos de PCR que se acumulan durante el proceso de amplificación. La PCR cuantitativa ofrece la posibilidad de detectar, en tiempo real, el nivel de amplificación de una secuencia de interés, con el objeto de estimar la cantidad de esa secuencia presente en la muestra original. En la PCR a tiempo real, los procesos de amplificación y detección se producen en el mismo vial cerrado, sin necesidad de ninguna acción posterior. Además, mediante detección por fluorescencia se puede medir durante la amplificación la cantidad de ADN sintetizado en cada momento, ya que la emisión de fluorescencia producida en la reacción es proporcional a la cantidad de ADN formado. Esto permite conocer y registrar en todo momento la cinética de la reacción de amplificación.

Los termocicladores para llevar a cabo la PCR a tiempo real incorporan un lector de fluorescencia y están diseñados para poder medir, en cualquier momento, la fluorescencia emitida en cada uno de los viales donde se realice la amplificación. Los sistemas de detección por fluorescencia empleados en la PCR a tiempo real pueden ser de dos tipos: agentes intercalantes y sondas específicas marcadas. Los agentes intercalantes son fluorocromos cuya emisión de fluorescencia aumenta notablemente cuando se unen a ADN de doble hélice. El más empleado en PCR a tiempo real es el SYBR Green I. El incremento de ADN en cada ciclo se refleja en un aumento proporcional de la fluorescencia emitida. Este sistema de detección es de fácil implementación y mas económico que el uso de sondas específicas. Sin embargo, presenta baja especificidad. El riesgo de amplificaciones inespecíficas puede disminuirse iniciando la reacción de PCR a temperaturas elevadas (hot-start PCR). Pueden utilizarse polimerasas recombinantes que funcionan después de ser activadas por temperatura o polimerasas unidas a anticuerpos que mantie-

nen bloqueado el centro activo de la enzima hasta que son desnaturalizados por calor. Además, la mayoría de los equipos de RT-PCR tienen la posibilidad de determinar el  $T_m$  de los fragmentos amplificados, que depende de su longitud y de la composición de sus bases, resultando característico de cada fragmento. Las sondas de secuencia específica, están marcadas con dos tipos de fluorocromos, un donador y un aceptor. Las más utilizadas son las sondas de hidrólisis, denominadas también sondas TaqMan que contienen un fluorocromo de extinción de fluorescencia (Q, quencher) unido en el extremo 3' y otro reportero (R) en el extremo 5'. Cuando se produce la extensión de la cadena de la sonda por acción de la Taq polimerasa, como resultado del clivaje del extintor por su actividad de exonucleasa es emitida la fluorescencia. La fluorescencia detectada durante cada ciclo de la PCR será proporcional a la cantidad de ADN que se está amplificando.

La confiabilidad de los resultados obtenidos mediante esta técnica requiere de la realización en paralelo una curva patrón en las mismas condiciones para conocer la cantidad total de ADN que se está amplificando.

Entre las principales aplicaciones de PCR en tiempo real, podemos mencionar:

**Amplificación y detección de ADN o ARN específico en una muestra.** Los blancos de detección pueden estar asociados a la presencia de agentes patógenos, resultando una herramienta de extrema utilidad para fines de diagnóstico clínico. En el caso del ARNm, debe ser previamente retrotranscrito a cDNA, mediante una transcriptasa reversa que posee actividad endo H, es decir, remueve el ARNm permitiendo que se forme la segunda hebra de ADN. También puede aplicarse al estudio de las variantes de "splicing" (corte y empalme de intrones) del ARNm.

**Cuantificación del ADN o ARN diana en la muestra.** Se puede cuantificar la concentración inicial de ADN (o ARN) diana en la muestra de manera muy sencilla, a través de la construcción de una curva de calibrado. El control de la amplificación como medida de la concentración en la muestra se obtiene en las fases iniciales de la reacción, en las que la concentración de los reactivos no es limitante y el efecto de la variabi-

lidad en la eficiencia de amplificación es menos importante.

### **C) Variantes de la qPCR**

**PCR múltiple:** es una variante de la PCR cuantitativa en la que se amplifican simultáneamente dos o más secuencias blanco en una única reacción. Este análisis en simultáneo de varias secuencias optimiza el aprovechamiento de la muestra cuando se trata de materiales de disponibilidad limitada y ofrece la posibilidad de realizar controles internos, por ejemplo, la expresión de un gen constitutivo en ensayos de expresión. La PCR múltiple suele ser más complicada que la simple adición de varios cebadores a la PCR cuantitativa simple. Se requiere extremar los cuidados en cuanto al diseño específico de cebadores, sugiriéndose que tengan igual temperatura de fusión (meeting point,  $T_m$ ), un contenido de GC del 45-60% y que carezcan de homología entre sí. La alta eficiencia de los cebadores en la PCR individual, no asegura el éxito en la PCR múltiple, pero aumenta considerablemente las posibilidades de obtenerlo. La combinación de fluoróforos a utilizar en los diferentes cebadores también debe elegirse de manera cuidadosa, para que no exista superposición entre los espectros absorción y emisión entre ellos.

**Análisis de curvas de disociación.** Se basa en la aplicación de un gradiente de temperaturas creciente después de la PCR para monitorizar la cinética de disociación de los fragmentos amplificados, pudiéndose establecer su  $T_m$  para comprobar su especificidad. También permite el análisis de mutaciones puntuales, usando una sonda complementaria con el ADN de tipo salvaje, que abarque la posición del polimorfismo. El híbrido formado entre sonda y ADN de tipo salvaje tendrá una estabilidad mayor que el formado entre la sonda y el ADN mutado, puesto que en este caso la complementariedad no es absoluta, reflejándose en una  $T_m$  superior. Adicionalmente, permite establecer además, de forma relativa, la cantidad de ADN de tipo salvaje y la de tipo mutado cuando ambos están presentes en la misma muestra.

*En la Sección IX, Capítulo 4, se detalla más este punto y se aplica esta técnica para la detección de organismos genéticamente modificados (OGMs).*

## 6. GENOTECAS

Una genoteca (“gene library”) es una colección de fragmentos de ADN clonados que representan en su conjunto el ADN total de un organismo de interés o bien el ADNc, que representa el conjunto de genes que se están expresando en un órgano o tejido determinado o bajo una situación particular o momento de crecimiento o desarrollo. En el primer caso hablamos de una genoteca genómica, donde se encuentran todas las secuencias que se expresan y no se expresan en el organismo mientras que en el segundo se trata de una genoteca de ADNc, donde se encuentran solo las secuencias expresadas.

Estas genotecas consisten en una colección bacterias, conteniendo cada una un vector con el ADN inserto, dispuestas en pocillos. Las genotecas carecen de un catálogo a través del cual se pueda saber cuál es el clon que contiene una secuencia de interés, por lo que es necesario realizar un relevamiento de las colonias bacterianas. Para ello se utiliza una sonda de ADN complementaria a la secuencia o gen buscados, que puede ser conocida o deducirse a partir de la secuencia de aminoácidos de la proteína purificada.

Una de las principales dificultades en el clonado de ADN genómico es que algunas secuencias están representadas una sola vez en el genoma y es difícil hallarlas. Para superar esta limitación de la metodología, puede clonarse directamente el ADNc, ya que el número de copias del ARNm en el tejido correspondiente a un gen que se está expresando es más elevado. El problema se plantea cuando no se sabe en qué circunstancias o en qué tejido se expresa un gen en particular. Actualmente es posible clonar un ADNc a partir de mensajeros poco abundantes en la célula, con concentraciones de 1 a 2 moléculas por célula.

### A) Genotecas de ADNc

El primer paso en la construcción de una genoteca de ADNc consiste en el aislamiento del ARN a partir del tejido o condición experimental planteada. Los cambios en la expresión génica asociados a la exposición a estímulos diversos, hacen necesaria la preservación del material

nitrogeno líquido para impedir cambios cuali o cuantitativos de la expresión. La susceptibilidad del ARN al clivaje por RNAsas, requiere de cuidados especiales en el material del laboratorio a ser empleado en el aislamiento. A partir del ARN total es posible separar el ARNm, tomando ventaja de la característica cola de poliadeninas que posee en el extremo 3', mediante cromatografía de afinidad, utilizando columnas de celulosa que tienen ligados oligonucleótidos compuestos solo por desoxitimina-oligo(dT). Al pasar el ARN por la columna, el ARNm queda unido por complementariedad al oligo(dT), a través de la cola de poliA. Una vez separado del ARN total, el ARNm es eluido de la columna utilizando una solución tampón adecuada.

Estas colas de poliA también son utilizadas en el próximo paso de clonado, donde funcionan como sitio de unión de los cebadores oligo(dT), para la reacción catalizada por la enzima transcriptasa reversa. El producto de la reacción es un híbrido ARN-ADN. La principal limitación de esta técnica radica en que en los casos en los que el ADNc es muy largo, al comenzar en el extremo 3', muchas veces no llega la retrotranscripción al extremo 5'. Para subsanar esta dificultad, se utilizan primers al azar de 6 a 10 nucleótidos de longitud y están confeccionados de manera de representar muchas secuencias diferentes, por lo que la reacción comienza a partir de muchos sitios diferentes, no solo del extremo 3'. Cualquiera de estas estrategias conduce a la obtención de un híbrido ARN-ADN, a partir del cual mediante PCR, se obtiene ADN de doble cadena que puede clonarse en el vector apropiado.

Para obtener la segunda cadena, se desarrolló inicialmente un método que tomaba ventaja de una “vuelta” o “giro” que da la hebra recién sintetizada, como un efecto de cambio de rumbo de la transcriptasa reversa al llegar al final de la cadena de ARN. Este artefacto provee un cebador adecuado para la síntesis de la segunda cadena de ADN y puede eliminarse, una vez obtenida la segunda cadena, con una nucleasa S1, perdiéndose parte de la secuencia correspondiente al extremo 5' del mensajero.

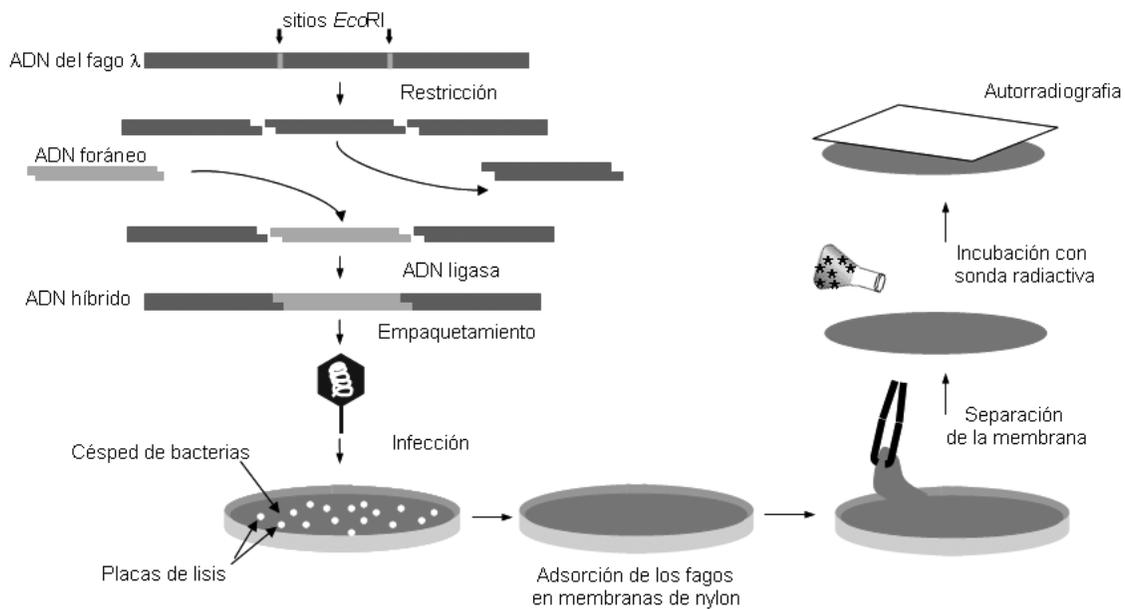
Existe una segunda técnica, que presenta dos ventajas sobre la anterior. Una es que ge-

nera moléculas de ADNc más largas y la segunda es que contiene prácticamente toda la secuencia correspondiente al extremo 5'. Se basa en la utilización de la enzima RNasaH, que reconoce moléculas híbridas ARN-ADN y digiere la cadena de ARN dejando trozos pequeños que permanecen unidos a la primera cadena del ADNc y sirven como *primers* para la ADN polimerasa I, que usa el ADNc original como molde para sintetizar la hebra complementaria de ADN. Solo queda un pequeño segmento de ARN en el extremo 5'. La segunda cadena de ADN tiene algunos sectores no unidos que son sellados por una ligasa de ADN.

Estas moléculas de ADNc de doble cadena están listas ahora para ser insertadas en un vector, que puede ser un plásmido o un deriva-

do del fago. Para ello se utiliza la transferasa terminal, que adiciona colas de poliA o poliT, o se agregan adaptadores, que son sitios artificiales de reconocimiento de alguna enzima de restricción que se unen a los extremos de la secuencia. Se trata de oligonucleótidos artificiales (8-12 bp) que se unen al fragmento utilizando una ligasa de ADN y son cortados con la enzima de restricción apropiada (Fig. 7). Ambos procedimientos sirven para generar extremos cohesivos a fin de unir el fragmento al vector, que posee extremos cohesivos cortados por la misma enzima.

Si se utilizan como vectores los plásmidos, se introducen en las células huésped (bacterias) por transformación, Si se eligen los fagos, son empaquetados *in vitro* para formar



**Figura 7.** Clonado de ADNc de doble cadena en un fago. El ADN del fago contiene dos sitios de restricción *EcoRI*. El ADN de la región central no esencial del fago se reemplaza por ADN foráneo, uniendo los fragmentos del fago con el ADN a clonar a través de la enzima ADN ligasa. Es necesario previamente adicionar al ADNc adaptadores que llevan sitios *EcoRI*, para generar extremos cohesivos para acoplarse con el ADN del fago. Se genera así una serie de moléculas recombinantes dispuestas en tandem, flanqueadas por sitios *cos*. Este ADN recombinante se empaqueta formando partículas infecciosas del fago, que se utilizarán para infectar un césped de bacterias. Esto se visualizará como placas o calvas. Cada placa surge de una molécula recombinante individual, que se propaga ahora como un fago. Para la localización del gen de interés en la genoteca, es necesario hacer una copia de la placa en membrana de nylon, que luego será hibridada con una sonda para el gen marcada apropiadamente. Si la marcación es radiactiva, por ejemplo, será revelada por autorradiografía y el clon se identifica por comparación entre la marca de la membrana de nylon y la placa de bacterias.

partículas víricas, que introducirán su ADN en el hospedador apropiado por infección de un césped de bacterias crecido sobre la superficie de una placa de agar (Fig. 7). Si bien los vectores plasmídicos son más fáciles de manipular, las genotecas construídas en los fagos poseen fragmentos de mayor tamaño.

Como resultado del cultivo en placa (plaqueo) de la genoteca, se obtienen cientos de miles a un millón colonias bacterianas o de placas de fagos (zonas claras o de lisis resultado de la producción de partículas víricas), dependiendo del vector utilizado, cada una conteniendo un fragmento clonado, distribuidas sobre la placa de agar. A fin de realizar la búsqueda de un gen particular ("screening"), la genoteca es replicada en membranas de nylon o nitrocelulosa, de manera tal que el patrón de placas en la caja original se vea exactamente reproducido en la membrana de nylon o filtro (Fig. 7). La búsqueda se lleva a cabo por hibridación con una sonda de ácido nucleico, lo cual requiere un conocimiento previo de la secuencia de interés.

**Elección de la sonda:** En algunos casos, donde parte del gen ha sido clonado, este se utiliza para buscar las partes faltantes. Si se usa una sonda donde la homología es completa el "screening" puede realizarse en condiciones de alta rigurosidad (es decir, con lavados más fuertes, que tienden a despegar lo que se ha unido inespecíficamente). Si la sonda no es completamente homóloga el "screening" deberá realizarse a menores temperaturas y utilizando concentraciones salinas más elevadas en los lavados. Si no se tiene conocimiento previo de la secuencia puede construirse una sonda a partir de la secuencia de aminoácidos de la proteína. Cuando se utiliza esta estrategia, el tamaño mínimo de la sonda debe ser de 15 a 16 nucleótidos (que permite encontrar secuencias únicas en genotecas de ADNc eucarióticos. Generalmente se utilizan sondas de 17 a 20 nucleótidos (correspondientes a 6 aminoácidos contiguos). Otra consideración a tener en cuenta cuando se construye la sonda es que existe más de un codón o triplete para definir la mayoría de los aminoácidos (es lo que se conoce como la "degeneración" del código genético) por lo que un aminoácido puede estar especificado por hasta 6 codones diferentes.

Por ello, estas sondas oligonucleotídicas están constituidas por una mezcla de todas las combinaciones posibles. Una de ellas corresponderá a la secuencia correcta del gen buscado. El problema de esta estrategia es el elevado número de falsos positivos que pueden dar las secuencias adicionales. Una variante consiste en usar un menor número de oligonucleótidos o uno solo pero de mayor longitud (35-75 nucleótidos), eligiendo una región de la proteína que contenga el menor número posible de codones degenerados, utilizando el conocimiento de los codones más probables para la especie en estudio.

Existen otras técnicas para localizar genes en las genotecas de ADNc, como hibridación diferencial, si se trata de genes expresados en diferentes tejidos o la utilización de sondas de regiones conservadas en familias de proteínas para encontrar genes relacionados o bien la utilización de vectores de expresión.

### **B) Genotecas Genómicas**

Un genoteca genómica puede obtenerse de cualquier tejido ya que todas las células de un organismo tienen la misma constitución genética. Se encuentran en ella, no solo las secuencias expresadas sino también las correspondientes secuencias regulatorias. Pueden utilizarse fagos como vectores, pero sucede que muchos de los genomas eucariotas son muy grandes, el de mamíferos por ejemplo contiene  $3 \times 10^9$  pb de ADN. Si el promedio típico de inserto es de 15.000 pb, se necesitarán unos 200.000 fagos para contener el genoma completo. Para asegurar que todas las secuencias estén representadas al menos una vez, los cálculos estadísticos dicen que deberán analizarse aproximadamente de 1 a 2 millones de fagos.

Por este motivo, especialmente cuando se trata de genomas muy grandes, como es el caso del genoma humano o de varias especies vegetales y animales, se utilizan las genotecas de YACs o BACs.

### **Genotecas genómicas en cromosomas artificiales**

La capacidad de clonar fragmentos de más de 100 kb es crucial para la **genómica estruc-**

**tural, funcional y comparativa** de organismos complejos. Las primeras genotecas genómicas fueron construidas utilizando como vectores los YACs, aunque por sus limitaciones se prefiere actualmente el uso BACs y PACs como vectores. Cuando se trata de genomas de plantas, pueden utilizarse como vectores los cromosomas artificiales de bacterias binarios (BIBACs), que presentan la ventaja adicional de pueden usarse directamente para transformar plantas utilizando *Agrobacterium tumefaciens*.

Para preparar una genoteca genómica el ADN es escindido con enzimas de restricción o mecánicamente al azar, si se desea evitar la presencia de fragmentos demasiado largos o cortos. Un paso crítico en el proceso es el aislamiento de moléculas intactas de ADN de elevado peso molecular, esencialmente ADN cromosómico. Para ello las células se embeben en bloques de agarosa de baja temperatura de gelificación y se tratan con enzimas y otros reactivos para separar el ADN de las proteínas celulares y del RNAs. La función del bloque de agarosa es proteger a las grandes moléculas, que, embebidas en esta matriz, se someten a la digestión con enzimas de restricción que realicen cortes poco frecuentes ("rare cutters"). La preparación de ADN de alta calidad en plantas es más complicada que la correspondiente para el caso de animales debido a la presencia de la pared celular, que dificulta el embebido en bloques de agarosa. Para superar esta dificultad se trabaja directamente con los núcleos aislados. Si se desea separar estos fragmentos por electroforesis, los bloques de agarosa conteniendo el DNA digerido pueden ser directamente embebidos en la matriz del gel que se correrá.

La cantidad de clones BACs que constituyen una genoteca, se relaciona con el tamaño del genoma, el tamaño promedio de los insertos y la cobertura genómica esperada. Considerando el tamaño de inserto promedio de los clones BACs, para cubrir cinco veces el genoma, una genoteca de BACs de la especie modelo *Arabidopsis thaliana*, cuyo genoma es pequeño, requerirá la obtención de 7.000 clones BACs, mientras que una de trigo candeal, cuyo genoma es dos ordenes de magnitud mayor, tendrá que estar compuesta por 500.000 clones.

**Separación de grandes trozos de ADN: electroforesis de campo pulsátil.** Las técnicas estándar de separación de ADN en geles de agarosa no son adecuadas para moléculas mayores de 10 kb. Esta limitación ha sido subsanada con la electroforesis de campo pulsátil, en la que el campo eléctrico en el gel de agarosa cambia periódicamente de orientación. De esta manera pueden separarse moléculas tan grandes como los cromosomas de levaduras (200 a 3000 kb) (ver VII.-1). La separación de las moléculas de este tamaño depende de su relativa facilidad o dificultad para reorientarse en respuesta a direcciones cambiantes de un campo eléctrico. Esta técnica puede utilizarse para hacer mapas físicos a gran escala.

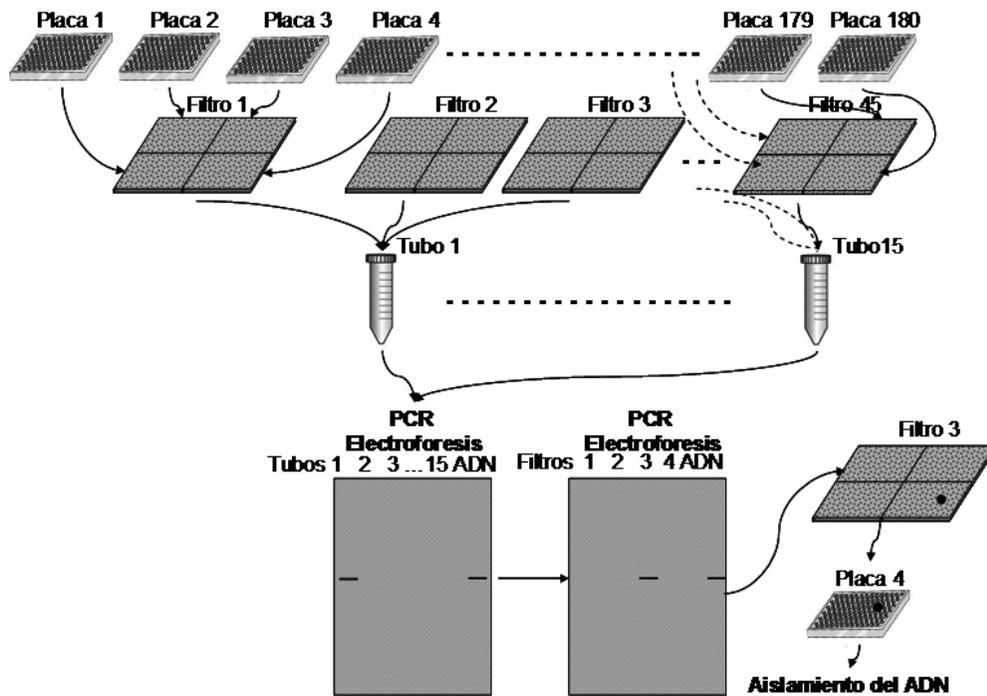
#### **Preparación de una genoteca de BACs.**

Como todo proceso de clonación, consiste en la preparación del vector, aislamiento y digestión del ADN, selección de los fragmentos por tamaño, transformación de las bacterias y ensamblado de la genoteca.

El vector debe estar purificado, digerido y defosforilado (cuando se corta con una sola enzima, se eliminan los fosfatos terminales, para evitar la recircularización). La digestión del ADN es crítica para obtener fragmentos adecuados para clonar. Los fragmentos son seleccionados por electroforesis de campo pulsátil y se separan de la matriz de agarosa por digestión de la misma con agarosa o gelasa, o por elusión del ADN desde el gel. Esto permite construir genotecas con fragmentos de tamaños similares, de aproximadamente 150 kb.

Estas genotecas de grandes insertos son consideradas recursos de largo plazo para investigación genómica y deben ser mantenidas continuamente, especialmente cuando son utilizadas para proyectos de secuenciación y mapeo a gran escala. En general, cuando se va a construir una genoteca debe elegirse cuidadosamente el genotipo de la especie utilizada como fuente de ADN, el tipo de inserto, etc., dado que la misma puede utilizarse para varios propósitos.

En la Fig. 8 se esquematiza la búsqueda de un gen en una genoteca de BACs. Más detalles acerca del ensamblado de los distintos fragmentos clonados puede encontrarse en el capítulo correspondiente a genómica.



**Figura 8.** Búsqueda de un gen específico en una genoteca de BACs mediante PCR. Las 180 placas de cultivo, de 96 pocillos cada una, contienen un genoma vegetal completo. Estas placas se replican, agrupadas de a 4, en filtros o membranas. Se realiza luego una reacción de PCR con el cebador específico al ADN obtenido de un tubo que contiene los clones correspondientes a tres filtros, representando 12 placas, que fueron previamente agrupados. Como control positivo se utiliza el ADN vegetal total, también amplificado por PCR. La reacción se analiza sobre un gel de agarosa. Un resultado positivo indica la presencia del gen en el ADN genómico del vegetal y también en uno de los tubos, conteniendo el ADN de 12 placas de 96 pocillos, en la figura, corresponde al tubo 1, que contiene los clones de las placas 1, 2 y 3. Se realiza en mismo procedimiento con tubos que contienen los clones de cada placa. Cuando se chequean los conjuntos individuales se encuentra que el clon positivo pertenece al filtro 3. El producto de PCR se hibrida luego con este filtro y esto nos dará la posición exacta en la placa de 96 pocillos donde se encuentra el gen de interés.

### C) Genotecas de cromosomas específicos o de segmentos de cromosomas

Cuando se busca un gen que se sabe está ubicado en un cromosoma específico simplifica mucho el trabajo hacer una genoteca de ese cromosoma solamente. Se trata de una técnica que permite separar cromosomas metafásicos teñidos con el colorante fluorescente Hoechst 33258 para regiones ricas en AT y cromomicina A, para regiones ricas en GC. Los cromosomas teñidos fluorescerán cuando se expongan a luz UV (de longitudes de onda de 361 y 363 nm (Hoechst 33258) o 458 nm (cromomicina A). Las cantidades y proporciones de colorantes varían para cada cromosoma y una computa-

dora reconoce el patrón fluorescente típico de cada cromosoma. Algunos cromosomas son muy similares por lo que no pueden ser separados. Se necesita un millón de cromosomas para hacer una genoteca. Equipos automáticos pueden hacer este trabajo en unas pocas horas.

También puede microdisectarse un cromosoma, tomando un segmento del mismo que tenga la región de interés. En principio se utilizó esta técnica para cromosomas grandes, fácilmente distinguibles por microscopía de contraste de fase. Actualmente, la combinación con PCR posibilita la aplicación de la técnica a cualquier cromosoma. Estos son teñidos con

Giemsa-tripsina y las bandas deseadas son cortadas con agujas de vidrio ultrafinas y clonadas. Cebadores posicionados en el vector permiten amplificar secuencias desconocidas. El ADN amplificado se clona en un vector apropiado y la localización cromosómica de cada clon se determina utilizando hibridación *in situ* (ver III.5)

## 7. SECUENCIACIÓN DEL ADN CLONADO

Una vez que se ha obtenido el clon con el fragmento de interés, el paso siguiente es secuenciarlo. La determinación de la secuencia puede realizarse por el método químico de Maxam y Gilbert o por el método enzimático de Sanger. El primero consiste en la utilización de compuestos que destruyen selectivamente una o dos de las bases que constituyen el ADN, partiendo de moléculas de ADN marcadas radiactivamente en un extremo. La destrucción no debe ser total, sino que en promedio cada molécula de ADN individual, debe ser clivada una vez. En función del compuesto químico utilizado y de las condiciones de reacción, se obtienen fragmentos de ADN hidrolizados a la altura de las bases G, A+G, T+C o C. Se generan fragmentos de distinto tamaño en función de la distancia de la base destruida al extremo marcado del fragmento a secuenciar. Los productos de las 4 reacciones se corren en un gel de poliacrilamida que es revelado por autorradiografía y la secuencia del fragmento queda determinada por el patrón de bandas.

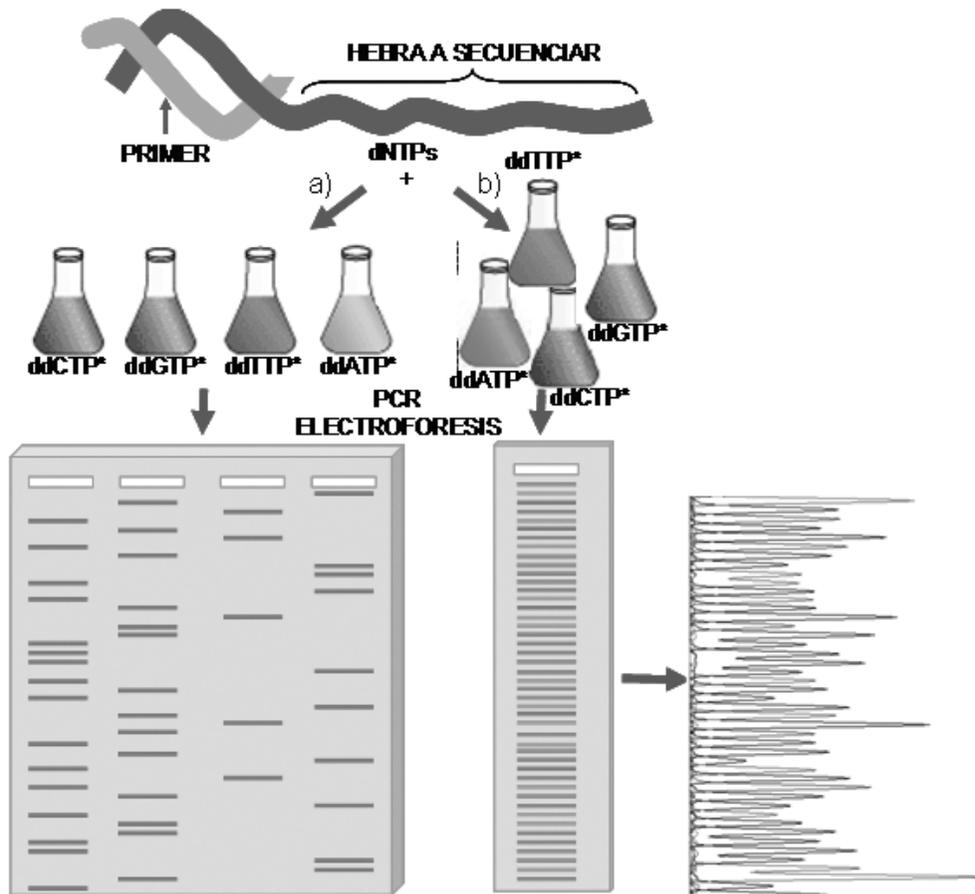
El método de Sanger, se basa en la interrupción controlada de la replicación enzimática del ADN, añadiendo a la reacción de síntesis junto con los cuatro desoxinucleótidos, un 2',3'-didesoxinucleótido (ddNTP) marcado. Los ddNTPs pueden ser incorporados al ADN por la ADN polimerasa pero interrumpe la síntesis de la cadena por carecer del OH en posición 3', necesario para la formación del enlace con la base siguiente. Se preparan cuatro reacciones con el ADN molde, el cebador, los cuatro dNTPs, y en cada una de ellas diferencialmente un ddNTP. Con la proporción ddNTP:dNTP adecuada, se generan fragmentos de distinta longitud, dependiendo de la distancia de la última base incorporada al extremo marcado del ADN. Estos fragmentos son separados en un gel de polia-

crilamida donde, mediante autorradiografía, se determina el patrón de bandas, que refleja la secuencia del ADN (Fig 9). Actualmente se utilizan digitalizadores de imágenes para la lectura de la autorradiografía.

Messing desarrolló una serie de vectores de clonado basados en el fago filamentoso M13, muy útiles para el secuenciado enzimático, con los que se utiliza un cebador universal que se posiciona en un lugar adyacente al sitio de policlono del vector. La ventaja es que solo una de las cadenas del vector es empaquetada en las cabezas del fago. Este ADN de simple cadena es ideal para utilizar con el método de Sanger. Actualmente se utilizan fasmidos. La adición de un "fago ayudante" ("helper phage") a las células que lo contienen, hace que el ADN del mismo, de simple cadena, sea replicado, empaquetado y extruído de la célula. Al ser de menor tamaño (aprox. 3000 bp) que M13 permite la inserción de fragmentos de ADN más largos.

Los proyectos de secuenciación genómica han requerido métodos de secuenciado más veloces y económicos. Para ello se ha desarrollado una técnica llamada Multiplex, que permite manejar mayor cantidad de muestras en menor tiempo. Se utilizan 20 vectores plasmídicos que permiten construir 20 genotecas diferentes a partir del mismo ADN. Cada vector lleva dos secuencias únicas que flanquean al sitio de clonado, que pueden ser usadas como etiquetas ("tags") para el ADN clonado y estarán presentes sobre cada fragmento a ser secuenciado. Se toma un clon de cada una de las genotecas plaqueadas (20 colonias) y se mezclan. Se hace crecer el cultivo, se aísla el ADN y se secuencian los plásmidos utilizando el método Maxam y Gilbert. Los productos de 12 conjuntos de reacciones de secuenciación se corren en un gel y se transfieren a una membrana de nylon. El filtro se hibrida secuencialmente (es decir, se despegan una sonda para luego hibridar con la siguiente y así sucesivamente), utilizando como sonda la etiqueta de cada uno de los 20 vectores. En cada caso se van leyendo las secuencias correspondientes a cada uno de los distintos vectores. De esta manera, una sola reacción produce datos de 20 clones a la vez.

Actualmente existen equipos robotizados que pueden llevar a cabo dos de las reacciones más limitantes en el proceso de secuen-



**Figura 9.** Esquema de la secuenciación de ADN por el método de dideoxinucleótidos, de manera a) manual, usando marcación radiactiva ó b) automática, utilizando ddNTPs que emiten color. Se representa a la hebra de ADN a secuenciar con el cebador unido por complementariedad. La reacción de síntesis del ADN complementario por la enzima ADN polimerasa (PCR) se desarrolla: a) en cuatro tubos individuales, cada uno con los cuatro desoxinucleótidos (dNTPs) y un dideoxinucleótido particular (ddNTP). En general, el dNTP marcado radiactivamente es dATP; ó b) en un único tubo, con cada ddNTP unido a un marcador fluorescente diferente. Luego, realiza la corrida electroforética en gel de poliacrilamida y la secuencia es leída desde la banda más corta (extremo 5') hacia el de mayor longitud (extremo 3') de manera: a) manual ó b) digital, luego de la excitación con láser y emisión de fluorescencia, obteniéndose un cromatograma.

ciado: la detección de las bandas de ADN y la traducción de un patrón de bandas en uno de secuencia. Estos equipos utilizan nucleótidos marcados con fluorocromos. Se utilizan 4 colorantes diferentes, que al ser excitados por láser emiten luz de diferente longitud de onda. Los colorantes pueden utilizarse para marcar el *primer* universal de secuenciación de M13 o cada uno de los 4 terminadores de cadena dideoxi. En este caso, cada mezcla de reacción con un terminador diferente se marca con un colorante diferente. Cuando la reacción es

completada los productos de las 4 reacciones se mezclan y se corren en una sola calle de un gel, en un secuenciador automático. A medida que los fragmentos pasan a través del láser, sus marcas fluorescentes se excitan y emiten luz que se detecta mediante un fotomultiplicador. Después de ser procesada por una computadora, la secuencia es mostrada como una serie de picos o cromatograma, donde cada uno de los 4 colores representa a un nucleótido diferente (rojo: timina, verde: adenina, negro: guanina y azul: citosina) (Fig. 9).

## Lecturas Recomendadas

- Madigan M. T., Martinko J. M. and Parker J. 1999 Brock, *Biología de los Microorganismos*. Octava edición revisada. Prentice Hall Iberia. Madrid, España.
- Singleton P. 2004. *Bacterias en Biología, Biotecnología y Medicina*. 5ta Edición. Acribia.
- Sambrook J., Fritsch E. F. and Maniatis T. 1989. *Molecular cloning: a laboratory manual*, second edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Southern E. M. 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, **98**, 503-517
- Fütterer J., Gisel A., Iglesias V., Kloti A., Kost B., Mittelsten Scheid O., Neuhaus G., Neuhaus-Url G., Schrott M., Shillito R., Spangenberg G. and Wang Z. Y. 1995. Standard Molecular Techniques for the Analysis of Transgenic Plants. En *Gene Transfer to Plants*. Potrykus Y, Spangenberg G (eds.) Springer Lab Manual.
- Watson J. D., Baker T. A., Bell S. P., Gann A., Levine M. and Losick Y. R. 2006. *Biología Molecular del Gen*. 5ª Edición, Ed. Médica Panamericana. Madrid.

# I. CAPÍTULO 5.

## Marcadores Moleculares

María Carolina Martínez, Marcelo Helguera, Alicia Carrera.

### 5.1. Introducción

Desde sus comienzos, el objetivo del mejoramiento vegetal ha sido seleccionar genotipos superiores a partir de la identificación de fenotipos superiores. El grado de éxito en este proceso depende de: i) el número de genes involucrados en el control genético del carácter (herencia monogénica o poligénica) y las relaciones interalélicas (dominancia o aditividad), ii) la influencia del ambiente, que se mide normalmente a través del parámetro heredabilidad. El proceso de caracterización y/o selección se vuelve más eficiente a través del uso de marcadores genéticos, definidos como caracteres que presentan polimorfismo o variabilidad experimentalmente detectable en individuos de una población segregante y un tipo de herencia predecible según las leyes de Mendel. Esta variación puede considerarse a diferentes niveles biológicos, desde cambios fenotípicos heredables significativos (marcador morfológico) hasta la variación de un solo nucleótido de ADN (marcador molecular). El marcador ideal debería ser altamente polimórfico (dentro y entre especies), de herencia mendeliana no epistática, insensible a los efectos ambientales, codominante (capaz de diferenciar individuos heterocigotas de homocigotas), de rápida identificación y simple análisis, y de detección en los estadios tempranos del desarrollo de la planta.

En la actualidad existe una gran variedad de marcadores genéticos y a lo largo de este capítulo desarrollaremos aquellos más frecuentemente utilizados en plantas. Para una mejor comprensión se propone clasificar los marcadores en las siguientes categorías: i) marcadores morfológicos, ii) marcadores bioquímicos, iii) marcadores moleculares, iv) marcadores funcionales o de expresión, vi) marcadores basados en SNPs.

### 5.2 Marcadores Morfológicos

Son características fenotípicas de sencilla

identificación visual tales como forma de hoja, pubescencia, color de fruto, etc. En todos los casos debe tratarse de caracteres de herencia monogénica y predecible según las leyes de Mendel. Muchos de ellos se convierten en importantes descriptores a la hora de inscribir nuevas variedades. Por ejemplo, alrededor de 50 caracteres de plántula, tallo, hoja, espiga, espiguilla y cariopse se utilizan para inscribir e identificar variedades de trigo en la Secretaría de Agricultura, Ganadería, Pesca y Alimentación de la Argentina. Este tipo de marcadores contribuyó significativamente al desarrollo teórico del concepto de ligamiento genético y a la construcción de los primeros mapas genéticos de plantas (como por ejemplo, en tomate y maíz). En el mejoramiento de girasol, los primeros híbridos se obtuvieron utilizando "color de hipocótilo" como marcador ligado al gen *ms* de androesterilidad; de este modo las plantas verdes que eran androestériles se utilizaban como líneas maternas y las plantas con antocianinas que eran fértiles se eliminaban del lote de producción al comienzo de su desarrollo.

Las principales limitaciones de los marcadores morfológicos son: i) se encuentran disponibles en un número restringido de especies vegetales, utilizadas como sistemas modelo para estudios genéticos, tales como el maíz, el tomate y la arveja, ii) bajo nivel de polimorfismo, iii) pueden producir alteraciones fenotípicas que dificultan el desarrollo de la planta (albinismo), iv) generalmente de herencia dominante y en algunos casos poligénica y, v) muchos de ellos se expresan en estadio de planta adulta, lo cual prolonga los tiempos de evaluación en los programas de mejoramiento.

Un enorme espectro de especies vegetales carece de información a este nivel. No obstante, los marcadores morfológicos permanecen como caracteres útiles en la identificación de materiales dado que representan un conjunto de genes que pueden ser evaluados con métodos sencillos y a bajo costo.

### 5.3 Marcadores Bioquímicos

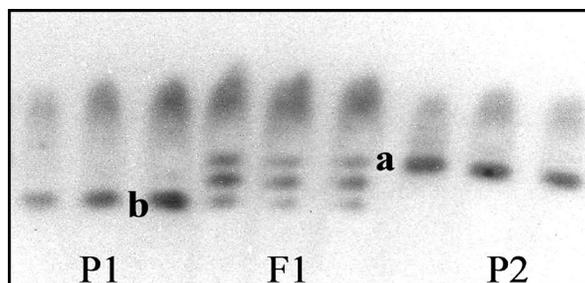
Son polimorfismos presentes en ciertas proteínas detectados a través de técnicas bioquímicas. El desarrollo de los marcadores bioquímicos produjo una revolución en los estudios

genéticos en plantas, que hasta el momento habían contado con un limitado número de marcadores morfológicos. La técnica permitía incluir potencialmente a todas las especies de plantas. En esta parte del capítulo nos referiremos a las Isoenzimas y a las Proteínas de reserva.

### 5.3.1 Isoenzimas

Las isoenzimas se definen como diferentes formas moleculares de una enzima, que poseen una actividad catalítica común, es decir actúan sobre el mismo sustrato. Ciertos cambios en el ADN que codifica estas enzimas (mutaciones) pueden resultar en cambios en la composición de aminoácidos, originando proteínas con la misma actividad biológica pero con diferente carga neta y por lo tanto con diferentes velocidades de migración en un campo eléctrico. Estas diferencias determinan patrones característicos de migración electroforética de las formas iso-enzimáticas. Varios factores definen el patrón de bandas o zimograma: i) Número de genes que las codifican: la presencia de varios genes codificando para una misma enzima ha sido atribuida a procesos de duplicación génica y subsecuente divergencia a través de mutaciones diferentes en cada caso; ii) Estados alélicos: el proceso más simple de generación de nuevas formas enzimáticas es la mutación de un gen estructural, las variantes alélicas se denominan aloenzimas. Estos marcadores muestran codominancia (en un individuo diploide ambos alelos de un locus son expresados y visualizados). A modo de ejemplo, la enzima alcohol dehidrogenasa (ADH), puede comprender varios loci y alelos con denominación *Adh-1a*, *Adh-1b*, *Adh-2a*, *Adh-2b*, *Adh-2c*, etc.; iii) Estructura cuaternaria de los productos proteicos: la enzima funcional puede estar compuesta por un número variable de sub-unidades. En las formas más simples o monoméricas los individuos homocigotas presentan una banda y los heterocigotas simplemente la suma de ambas. Cuando la estructura cuaternaria se vuelve más compleja, encontramos enzimas activas compuestas por dímeros, tetrámeros, etc. En este caso el individuo heterocigota presenta bandas adicionales no presentes en los homocigotas, que se generan por la combinación de sub-unidades co-

dificadas por los distintos alelos o loci (Figura 1). Finalmente, iv) Compartimentalización subcelular: se encuentran formas enzimáticas localizadas en citoplasma, cloroplasto o mitocondria, codificadas todas por genes nucleares pero con diferentes velocidades de migración.



**Figura 1.** Locus isoenzimático Pgd-3 de girasol en el que se destacan las variantes alélicas a y b. P1 (calles 1 a 3 de izquierda a derecha) y P2 (calles 7 a 9) son individuos homocigotas y F1 (calles 4 a 6) son individuos heterocigotas.

Las isoenzimas se extraen por homogeneización del tejido (semilla, raíz, hoja) en condiciones no desnaturalizantes (que preservan la actividad catalítica de la enzima) y se analizan mediante electroforesis en un soporte sólido, generalmente almidón. El gel puede ser cortado en capas horizontales y cada una se destina a una solución de revelado específica que consta de un sustrato, un colorante y cofactores, disueltos en un buffer apropiado. La reacción que ocurre entre las enzimas presentes en el gel y el correspondiente sustrato generan productos coloreados que conforman el patrón de bandas. Las metodologías empleadas en la extracción y electroforesis son relativamente sencillas y rápidas y el equipamiento es de bajo costo.

Las isoenzimas han tenido un rol prominente en estudios de poblaciones vegetales para determinar variabilidad y estructura genética, sistemática y biología evolutiva así como en descripción de germoplasma e identificación de variedades. Su aplicación en la construcción de mapas se ha visto limitada por el número de marcadores isoenzimáticos disponibles (en general menor a 50), y por su reducido polimorfismo (2-4 alelos). Por otro lado, las formas enzimáticas extraídas de hoja o raíz (no así las

de semilla) presentan variaciones en relación a las condiciones ambientales de crecimiento y a la edad del tejido, lo cual afecta la reproducibilidad de los zimogramas. No obstante, estos loci representan importantes marcas de referencia para relacionar mapas obtenidos a partir de distintos marcadores de ADN.

### 5.3.2 Proteínas de reserva

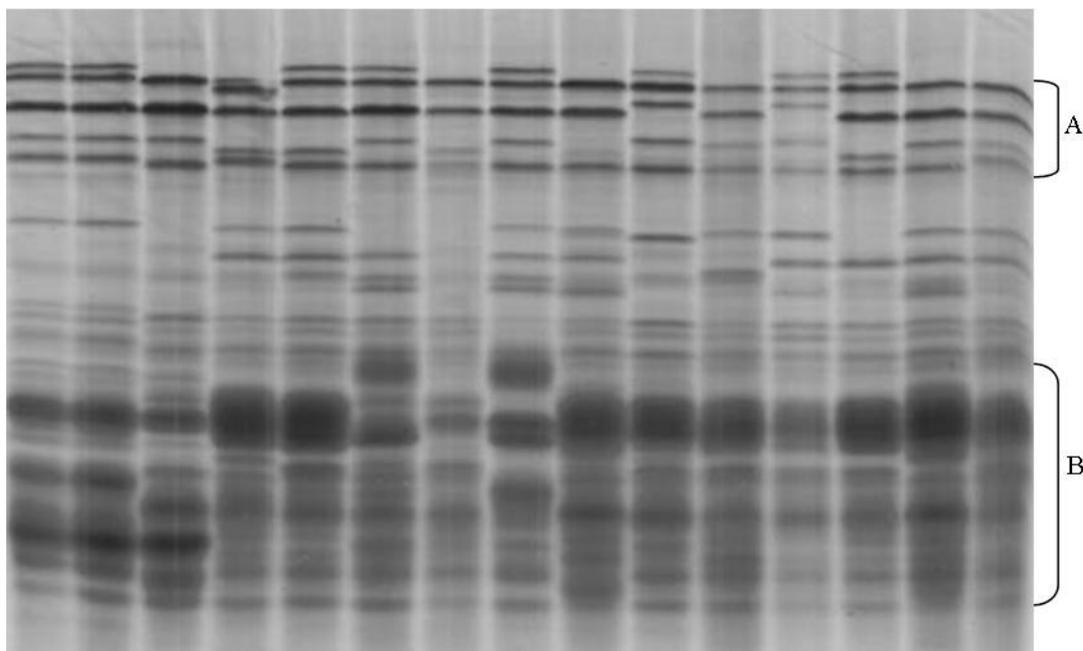
Las proteínas de reserva son un grupo heterogéneo de proteínas presentes en la semilla de la planta que tienen por función proveer de energía al embrión en los primeros estadios de crecimiento. Estas proteínas han sido extensamente estudiadas en cereales y oleaginosas y se relacionan directamente con la calidad del grano.

Las proteínas de reserva se extraen mediante procedimientos sencillos a partir de las semillas y se separan mediante electroforesis en geles de poliacrilamida en condiciones desnaturizantes (con detergentes). No se requiere detectar actividad enzimática, las proteínas se visualizan directamente por tinción con Azul de Coomassie, evidenciándose los polimorfismos como diferencias en la movilidad electroforética.

En la mayoría de los cereales, estas proteínas están codificadas por varios genes relacio-

nados conformando familias. Varios estudios demuestran que durante la evolución de estos genes se han detectado duplicaciones, translocaciones, inserciones de elementos móviles, entre otros rearrreglos cromosómicos, que serían los responsables de generar un alto nivel de polimorfismo proteico entre y dentro de especies.

Por ejemplo, en el caso del trigo, las principales proteínas de reserva del grano se denominan gluteninas y gliadinas. Estas proteínas son capaces de polimerizar durante el amasado de la harina formando una red denominada gluten de importancia fundamental en la elaboración del pan. Desde un punto de vista genético, las gluteninas se hayan codificadas en los loci *Glu-1* y *Glu-3* y las gliadinas en *Gli-1* y *Gli-2* pudiendo tener cada uno de estos loci 1, 2 o más genes; mutaciones en estos genes generan nuevas variantes alélicas. Numerosos estudios electroforéticos han revelado alto grado de polimorfismo en el número y la movilidad electroforética de estas proteínas y muchos de estos polimorfismos son utilizados como marcadores genéticos para mejorar la calidad panadera del trigo, principalmente en el caso de las gluteninas (Figura 2). Las gliadinas muestran patrones electroforéticos más complejos que las



**Figura 2.** Gluteninas de cultivares de trigo separadas mediante electroforesis en geles de poliacrilamida al 12% en condiciones desnaturizantes (SDS-PAGE). A: fracción correspondiente a las gluteninas de alto peso molecular, codificadas por el locus *Glu-1*. B: fracción correspondiente principalmente a las gluteninas de bajo peso molecular codificadas por el locus *Glu-3*. Fotografía gentileza de L. Vanzetti

gluteninas y han sido utilizadas en la descripción de germoplasma e identificación de variedades. Como en el caso de las isoenzimas, la aplicación de proteínas de reserva en la construcción de mapas genéticos es limitada por el número de marcadores genéticos disponibles, sin embargo estos loci representan importantes marcas de referencia, por ejemplo, para los cromosomas 1 y 6 de trigo, donde están presentes los principales genes de gluteninas y gliadinas.

Para más información sobre estructura, propiedades y rol de las proteínas de reserva en grano se recomienda la lectura de la revisión de Branlard et al. (2001).

La principal ventaja de los marcadores bioquímicos es su fácil implementación en el laboratorio ya que involucran metodologías sencillas, rápidas y de bajo costo. Como desventajas se pueden citar: baja cobertura del genoma, dificultades en la interpretación de los resultados (principalmente para las Isoenzimas) y, en el caso de las proteínas de reserva, dado que muchos de los genes codificantes suelen estar estrechamente ligados (familias multigénicas), no se puede asumir independencia entre los loci.

#### **5.4 Marcadores moleculares**

Un marcador molecular es simplemente un segmento de ADN con una ubicación específica en un cromosoma (punto de referencia) cuya herencia puede seguirse en individuos de una población. La secuencia puede pertenecer a regiones codificantes (genes) o sin función conocida.

Con el advenimiento de las técnicas modernas de biología molecular (para más detalle, ver Parte I, Capítulo 4 "Herramientas básicas de Ingeniería Genética"), surgieron diversos métodos de detección de polimorfismo genético directamente a nivel de ADN. En la actualidad, se puede obtener un número prácticamente ilimitado de marcadores en cualquier organismo vivo que permiten analizar la totalidad de la información genética (genoma) de un organismo.

Para describir los distintos tipos de marcadores moleculares, se seguirá una clasificación arbitraria en base a las metodologías que emplean para su detección. Todas estas técnicas parten, como primer paso, de extraer ADN genómico de la planta bajo estudio.

#### **5.4.1 Marcadores basados en la hibridación del ADN**

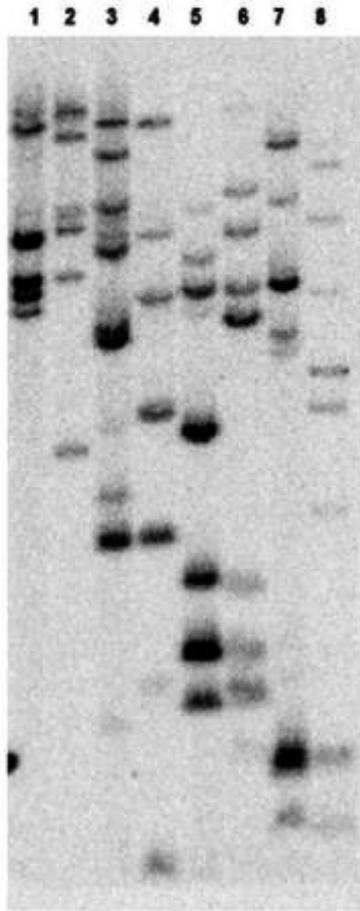
5.4.1.1 RFLP (*Restriction Fragment Length Polymorphism*) o polimorfismo en la longitud de fragmentos de restricción (Botstein et al., 1980)

Son los marcadores moleculares más antiguos y aún son usados para algunas aplicaciones importantes como mapeo comparativo y estudios de sintenia entre especies (ver Parte III, Capítulo 2 "Aplicaciones de los Marcadores Moleculares").

Este marcador se basa en la comparación de perfiles de bandas generados a partir del ADN de distintos individuos por digestión con enzimas de restricción. Los fragmentos obtenidos se separan por su tamaño mediante electroforesis y se transfieren a una membrana donde son desnaturalizados y luego hibridados con una sonda. Esta sonda es un fragmento de ADN de cadena simple (de secuencia conocida o no) que está marcado radioactivamente. La sonda se unirá a aquellos fragmentos de restricción con secuencia complementaria a la misma (ver enzimas de restricción y método de *Southern Blot* en Parte I, Capítulo 4 "Herramientas básicas de Ingeniería Genética").

Para la visualización, la membrana se expone a una placa radiográfica. En cuanto a la sonda empleada esta puede ser genómica (con alta proporción de ADN no codificante, no perteneciente a ningún gen) o de ADNc (cuando proviene de un transcripto). También puede obtenerse a partir de ADN de organelas (mitocondrias o cloroplastos). Pueden emplearse sondas aisladas de especies relacionadas (por ejemplo, sondas aisladas de tomate pueden emplearse en papa y viceversa).

La variabilidad genética presente en el marcador molecular proviene de diferencias en la secuencia del ADN genómico debidas a mutaciones puntuales en los sitios de restricción, a duplicaciones, deleciones, inserciones, etc., que modifican la distancia entre pares de sitios de restricción generando fragmentos polimórficos (de diferentes tamaños) (Figura 3). Un locus RFLP está definido por la combinación de una enzima de restricción y una sonda. Individuos homocigotas poseen el mismo



**Figura 3.** Análisis de RFLP empleando como sonda el marcador RGA(584). Calles impares: *S. pennellii*. Calles pares: *S. lycopersicum*. 1 y 2: ADN genómico digerido con *EcoRV*, 3 y 4: *HindIII*, 5 y 6: *EcoRI*, 7 y 8: *BamHI*. Fotografía de M. C. Martínez. y 6: *EcoRI*, 7 y 8: *BamHI*. Fotografía de M. C. Martínez.”

patrón de restricción en ambos cromosomas homólogos y por lo tanto producen una sola banda. Ejemplos de nomenclatura utilizada para loci RFLP son *XNpb136* de arroz o *Xpsr912-2A* de trigo, que hacen referencia al laboratorio de origen y a las sondas utilizadas.

Las ventajas de los RFLPs radican en que son altamente reproducibles, codominantes y multialélicos. Por otro lado, cuentan con las desventajas de ser muy laboriosos, difíciles de automatizar, se necesita disponer de las sondas para la especie en estudio y requieren de infraestructura adecuada para mantener las sondas y trabajar con radiactivo, lo que los hace relativamente costosos.

#### 5.4.1.2 VNTR (*Variable Number of Tandem Repeats*) o repeticiones en tandem de número variable

Los VNTR, también conocidos como minisatélites, son repeticiones de secuencias de 9 a 100 pares de bases. El número de repeticiones es variable pero en general es menor a 1000. Existen dos formas de detectar los VNTR: por hibridación o por técnicas de PCR. En el primer caso, el ADN genómico es digerido con enzimas de restricción que reconocen sitios adyacentes a la región repetida. Los fragmentos se separan por electroforesis, se inmovilizan en una membrana y se detectan mediante sondas. La longitud de los fragmentos de restricción producidos en diferentes individuos varía de acuerdo al número de repeticiones del minisatélite. La diferencia básica entre RFLP y VNTR reside en el tipo de sonda utilizada. En la técnica de VNTR las sondas están constituidas por secuencias homólogas a las secuencias repetidas de los minisatélites, por lo tanto todos los loci hipervariables del genoma son detectados simultáneamente, mientras que en RFLP, las sondas son homólogas a secuencias únicas del genoma, detectando así uno o pocos loci cada vez. De esta manera, en el autoradiograma, al contrario del patrón simple de bandas obtenido por RFLP, para minisatélites se obtiene un perfil complejo de bandas múltiples. Los VNTR tienen la ventaja que además de explorar el polimorfismo en la longitud de los fragmentos de restricción en cada locus hipervariable, utilizan también el polimorfismo en el número y distribución de estos loci a lo largo del genoma, posibilitando así la visualización simultánea de diversas regiones del mismo. Las limitaciones para esta técnica, son las mismas descritas oportunamente para RFLP.

Los minisatélites también pueden ser detectados mediante la amplificación por PCR de los segmentos conteniendo diferente número de repeticiones, utilizando iniciadores o "primers" que flanqueen los VNTR y luego su visualización mediante electroforesis y tinción con bromuro de etidio. Esta opción se ha vuelto más frecuente con el aumento de información de secuencias en bases de datos que permiten el diseño de iniciadores flanqueantes a determinados minisatélites. Sin embargo estos marcadores han sido poco utilizados en plantas.

#### 5.4.2 Marcadores basados en la amplificación arbitraria o semi arbitraria del ADN

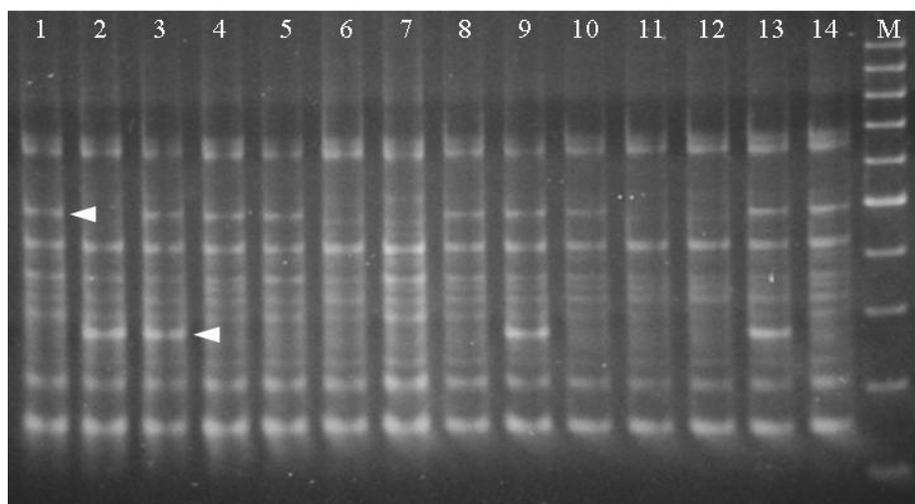
5.4.2.1 RAPDs (*Random Amplified Polymorphic DNAs*) o ADN polimórfico amplificado aleatoriamente (Williams et al. 1990, Welsh y McClelland 1990)

La metodología de RAPD fue desarrollada de forma independiente por dos grupos de Estados Unidos. Un grupo la denominó RAPD y el otro AP-PCR (*Arbitrarily Primed-Polymerase Chain Reaction* o reacción en cadena de la polimerasa con iniciadores arbitrarios).

La técnica de RAPD es una variante de PCR que utiliza un solo oligonucleótido de 10 bp que hibrida al azar con el ADN en estudio. Para que se genere un fragmento RAPD es necesario que el oligonucleótido iniciador hibride en las dos cadenas del ADN en orientaciones opuestas suficientemente cercanas (menos de 3000 bp) como para permitir la amplificación. La secuencia del oligonucleótido es aleatoria al igual que los sitios de hibridación en el ADN, por lo tanto la secuencia amplificada es desconocida. El iniciador puede hibridar en distintos sitios del ADN sin necesidad de complementariedad exacta de bases, dado que la PCR se hace con temperaturas de unión del iniciador bajas, generando varios fragmentos que son resueltos mediante electroforesis y posterior tinción (Figura 4). Cada banda del patrón de amplificación es considerada un locus RAPD, definido por la secuencia del iniciador y el peso molecular. El polimorfismo que se observa en-

tre distintos individuos consiste en la presencia o ausencia de fragmentos de ADN amplificado debido a mutaciones en el sitio de reconocimiento del iniciador, deleciones, inserciones. Este marcador no permite diferenciar individuos heterocigotas, por lo que es un marcador dominante. Los marcadores RAPD están basados en una técnica sencilla, de bajo costo de implementación, automatizable y no radioactiva. A diferencia de los marcadores basados en PCR convencional no se requiere información nucleotídica previa para su desarrollo, se dispone de un número ilimitado de marcadores y el nivel de loci polimórficos es alto. La principal desventaja de esta técnica es su baja reproducibilidad entre laboratorios, en el sentido de que pequeñas modificaciones en la técnica como por ejemplo, concentración de ADN inicial, modelo de termociclador, origen de la enzima ADN Polimerasa termoestable, etc., pueden alterar el patrón de fragmentos de ADN generados por RAPD de una muestra.

Relacionados con los RAPDs encontramos a los marcadores DAF (*DNA Amplification Fingerprinting*) (Caetano Anollés et al. 1991) y AP-PCR (*Arbitrary Primed PCR*) (Welsh y McClelland 1990). DAF y AP-PCR son técnicas similares a RAPD. DAF involucra el uso de iniciadores arbitrarios de 5 pares de bases de longitud. Esto incrementa la probabilidad de apareamiento con el ADN molde respecto al



**Figura 4.** Análisis de RAPD utilizando ADN genómico de trigo candeal definidos por electroforesis vertical en gels de agarosa. 1: cultivar Kofa, 2: línea experimental UC1113, 3 a 14: individuos RILs (líneas endocriadas recombinantes F8) generados del cruzamiento Kofa x UC1113, M: estándar de peso molecular de ADN (100 bp ladder, Invitrogen). Las flechas blancas indican los fragmentos polimórficos. Fotografía gentileza de Mercedes Nisi.

RAPD y, por lo tanto, resulta en un perfil más complejo de bandas. La visualización se lleva a cabo por medio de geles de poliacrilamida teñidos con plata. AP-PCR utiliza iniciadores ligeramente más largos que la técnica anterior (aprox. 20 pares de bases). Los productos de amplificación son marcados radiactivamente y también pueden ser resueltos por electroforesis en geles de poliacrilamida. Las ventajas y limitaciones de estas técnicas son similares a las descritas para los RAPD.

#### 5.4.2.2 AFLPs (*Amplified Fragment Length Polymorphism*) o polimorfismos en la longitud de fragmentos amplificados (Vos et al. 1995)

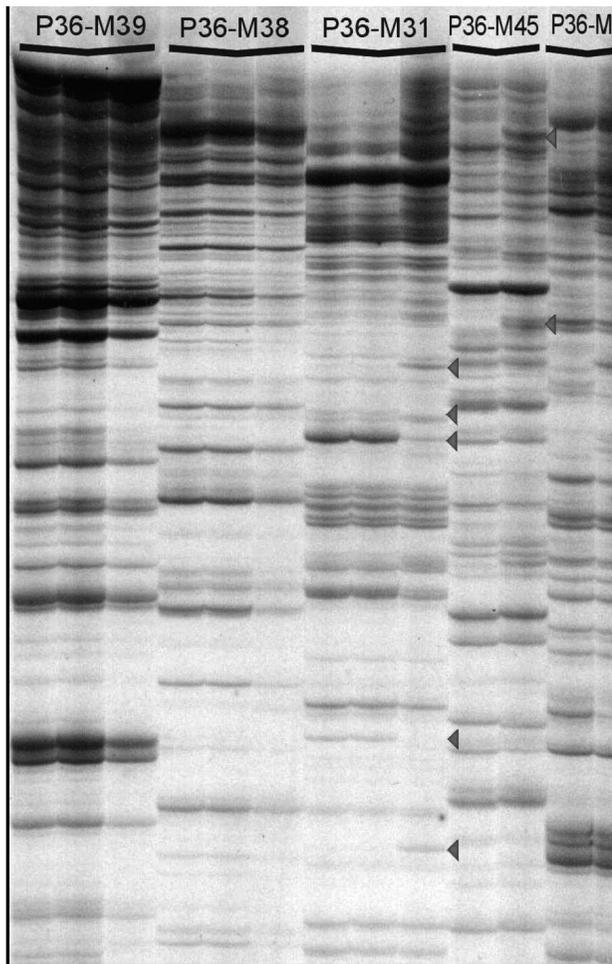
Los marcadores AFLP constituyen una herramienta poderosa de análisis del genoma dado que poseen un alto poder de detección de la variabilidad genética. El ensayo de AFLP combina la especificidad, resolución y poder de muestreo de la digestión con enzimas de restricción con la velocidad y practicidad de la detección de polimorfismos mediante PCR, sin necesidad de disponer de información previa del genoma a estudiar. Desde su desarrollo y divulgación esta técnica se utiliza ampliamente en el análisis de plantas.

Consiste esencialmente en cuatro etapas: i) el ADN genómico es cortado con dos enzimas de restricción. Generalmente una es de corte raro (ej. *EcoRI*), que reconoce de 6 a 8 pares de bases y otra es de corte frecuente (ej. *MseI*) que reconoce 4 pares de bases; ii) fragmentos de ADN doble cadena de 20 a 30 pares de bases llamados adaptadores se ligan en forma específica a los extremos de los fragmentos obtenidos en el paso anterior, generando así el molde para la amplificación posterior del ADN; iii) se amplifican selectivamente fragmentos por PCR. En esta etapa, se utilizan iniciadores de aproximadamente 20 nucleótidos que contienen una secuencia específica complementaria a la secuencia de los adaptadores y además, 1 a 3 nucleótidos selectivos adicionales de secuencia arbitraria en su extremo 3'. Dado que sólo una subpoblación de los fragmentos originales es amplificada, se obtiene un patrón de bandas que permite un registro adecuado. La amplificación descrita en iii) se realiza en dos etapas: una primera amplificación selecti-

va empleando un nucleótido arbitrario (amplificación +1 o preamplificación) y luego, este producto de amplificación obtenido es empleado como molde en una nueva amplificación empleando iniciadores que poseen 2 nucleótidos selectivos adicionales al anterior (amplificación +3 o amplificación final); iv) el análisis de los fragmentos así amplificados se realiza mediante electroforesis en geles de poliacrilamida desnaturizantes. Si uno de los iniciadores empleados está marcado radiactivamente, se visualizará mediante autorradiografía, si uno de los iniciadores está marcado con un compuesto fluorescente, puede ser resuelto empleando un secuenciador automático. Alternativamente, se puede visualizar mediante tinción con nitrato de plata (Figura. 5).

La base genética del polimorfismo de AFLP es la ausencia o presencia de fragmentos amplificados de un tamaño determinado dado por mutaciones puntuales, inversiones, inserciones y deleciones que llevan a la pérdida o ganancia de un sitio de restricción o la alteración de la secuencia reconocida o amplificada por los iniciadores. Al igual que en los RAPDs, no es posible distinguir individuos heterocigotas por lo que se trata de un marcador dominante. Una banda AFLP se interpreta como un locus, definido por las dos enzimas de restricción, una combinación de primers que incluyen las bases selectivas (Ej. *EcoRI-ATC/MseI-AAG*) y un peso molecular.

Su implementación en laboratorio requiere de una infraestructura considerable, son relativamente laboriosos para su obtención y el costo es medio a alto. Estos marcadores presentan un alto poder de detección de la variabilidad genética, ya que se explora simultáneamente el polimorfismo de ausencia/presencia de sitios de restricción (como los RFLP) y la ocurrencia o no de amplificación a partir de secuencias arbitrarias (como los RAPDs). Es un marcador mucho más robusto que los RAPDs, ya que en la amplificación se utilizan oligonucleótidos más largos, que aumentan significativamente la especificidad de la reacción sin perder las ventajas de la amplificación de secuencias al azar (no requiere información previa de secuencia de ADN). Asimismo, se pueden emplear distintas enzimas de restric-



**Figura 5.** Análisis de AFLP utilizando ADN de trigo candeal definidos por electroforesis en geles de secuenciación. En la figura se evaluaron 5 combinaciones de diferentes iniciadores para las enzimas de restricción *Pst* I y *Mse* I sobre el cultivar Kofa (K) y la línea experimental UC1113 (D). Las flechas indican presencia de fragmentos polimórficos. Fotografía gentileza de A. Picca.

ción e iniciadores selectivos, resultando en una ilimitada posibilidad de generar polimorfismos. Otra ventaja de los AFLPs es el número de fragmentos (marcadores) obtenidos por reacción y resueltos por electroforesis (oscila entre 30 - 50 contra los 4 - 10 de RAPDs).

Una variante de esta técnica es la llamada cDNA-AFLP que, en vez de ADN genómico, parte de ARN mensajero que es copiado a ADNc. Esta metodología es empleada en análisis de expresión diferencial de genes, estudios del transcriptoma y genómica funcional.

#### 5.4.3. Marcadores moleculares basados en la amplificación sitio-específica del ADN

En contraste con los marcadores basados en la amplificación de secuencias arbitrarias, los marcadores incluidos en la presente categoría requieren del diseño de iniciadores específicos para la amplificación de un locus en particular.

##### 5.4.3.1 Microsatélites (Litt y Luty 1989)

Los microsatélites son regiones hipervariables del genoma que contienen arreglos de secuencias simples en tandem de mono, di, tri, tetra o pentanucleótidos que se repiten entre 10 y 100 veces. Los marcadores microsatélites, también denominados *Simple Sequence Length Polymorphism* (SSLP), *Simple Sequence Repeat Polymorphism* (SSRP), *Simple Sequence Repeats* (SSR) o *Sequence-Tagged Microsatellite Sites* (STMS) se encuentran distribuidos por todo el genoma de la mayoría de las especies eucariotas.

Los microsatélites se detectan mediante su amplificación por PCR usando iniciadores específicos de 20 a 30 pb de longitud que hibridan en la región que flanquea al tandem de repeticiones (microsatélite). Estos marcadores se resuelven por electroforesis en geles de poliacrilamida en condiciones desnaturizantes, mediante tinción con plata o por autoradiografía (en el caso de usar un iniciador marcado radioactivamente). Si se dispone de un secuenciador automático, se pueden resolver por tamaño en este equipo mediante el empleo de un iniciador marcado con un fluoróforo, posibilitando un análisis automatizado.

La base genética del polimorfismo detectado en microsatélites se basa en la variabilidad del número de repeticiones en tandem y, consecuentemente, en el tamaño del microsatélite amplificado en individuos de una especie. Estas diferencias son originadas durante la replicación del ADN debido a fallas en la acción de la ADN polimerasa durante el copiado de una región repetida donde incorpora o elimina repeticiones. Otro mecanismo responsable de la variación es el entrecruzamiento desigual entre cromosomas homólogos. En este caso se generan alelos con diferencias mayores en el número de repeticiones.

El desarrollo de marcadores microsatélites requiere del conocimiento de las secuencias

flanqueantes adecuadas para el diseño de los iniciadores específicos. Los primeros SSR se desarrollaron a través de un proceso experimental complejo que implicó la obtención de genotecas enriquecidas en microsatélites, secuenciación, diseño de iniciadores aptos para amplificar microsatélites por PCR y selección de loci con patrones de herencia simple. Actualmente, las secuencias flanqueantes pueden obtenerse a partir de las secuencias depositadas en las bases de datos (búsqueda *in silico*). El uso de programas informáticos especializados en la búsqueda de microsatélites ha revelado que existen secuencias con estas características formando parte de secuencias que se expresan (genes), denominándose SSR génicos o EST-SSRs. Aunque tendrían un nivel de polimorfismo menor, serían más fácilmente transferibles entre especies relacionadas.

Un locus SSR está definido por las secuencias de los iniciadores flanqueantes al microsatélite. Por el alto polimorfismo que suelen presentar por locus (multiallelismo) se los considera los marcadores ideales para el mejoramiento en especies autóгамas como el trigo. Estos marcadores son codominantes (ambos alelos de un individuo heterocigota pueden ser visualizados), genoma-específicos y altamente polimórficos en comparación con los RFLPs y RAPDs. Su implementación en un laboratorio requiere de mayor infraestructura y presupuesto que los RAPDs, siendo semejantes en este aspecto a los AFLPs.

#### 5.4.3.2 Otros marcadores basados en los microsatélites

Una de las limitaciones para la utilización de los marcadores microsatélites es la necesidad de conocer las regiones que flanquean a las repeticiones para poder desarrollar los iniciadores específicos. Así, se desarrollaron marcadores moleculares buscando explorar las repeticiones microsatélites sin necesidad de secuenciar el ADN. Entre esta clase de marcadores podemos encontrar:

MP-PCR (*Microsatellite-Primed PCR*): en el desarrollo de este marcador molecular, se utiliza un iniciador que contiene las repeticiones de un microsatélite específico para amplificar el ADN. En las regiones del ADN que poseen

las dos secuencias inversamente orientadas de este microsatélite específico hibridará el iniciador y amplificará el fragmento de ADN localizado entre ellas. Conceptualmente, sería similar a RAPD empleando un iniciador con la secuencia repetida del microsatélite.

ISSR (*Inter-SSR amplification*): se basan en la amplificación por PCR empleando iniciadores que contienen en su secuencia repeticiones de di o trinucleótidos junto con 4 bases nucleotídicas determinadas en uno de sus extremos. Esto permite la amplificación parcial de todas las regiones (potencialmente) amplificadas por el marcador MP-PCR, descrito anteriormente, aumentando la reproducibilidad, que es una de las limitaciones del uso de los MP-PCR. Al tener en el iniciador esas 4 bases extras, a este tipo de marcador se lo conoce también como microsatélite anclado (AMP-PCR, *Anchored Microsatellite-Primed PCR*).

SAMPLE (*Selective Amplification of Microsatellite Polymorphic Loci*) o amplificación selectiva de loci microsatélites polimórficos. En esta metodología se amplifican loci microsatélites a partir de un iniciador arbitrario empleando dos técnicas: microsatélites y AFLP. Se realizan todas las etapas de AFLP, es decir, digestión con dos enzimas de restricción, ligación de adaptadores, preamplificación y amplificación selectiva del ADN. Sólo que en la amplificación selectiva final, se emplea el iniciador de AFLP con sus tres nucleótidos selectivos en combinación con un iniciador SAMPLE. Este iniciador, posee en su secuencia bases complementarias a un microsatélite. La banda polimórfica resultante de la amplificación del ADN con el marcador SAMPLE puede ser convertida en un marcador microsatélite convencional, a partir del clonado y secuenciación de la misma.

5.4.3.3. STS (*Sequence-Tagged Sites*) o sitios marcados por secuencias (Olson et al. 1989)

STS es un término general dado a un locus en particular definido por las secuencias de sus iniciadores específicos. Un STS puede ser generado para cualquier sitio del genoma siempre y cuando ese sitio (locus) pueda ser clonado y secuenciado. Un STS debe satisfacer dos condiciones: se debe conocer su secuencia,

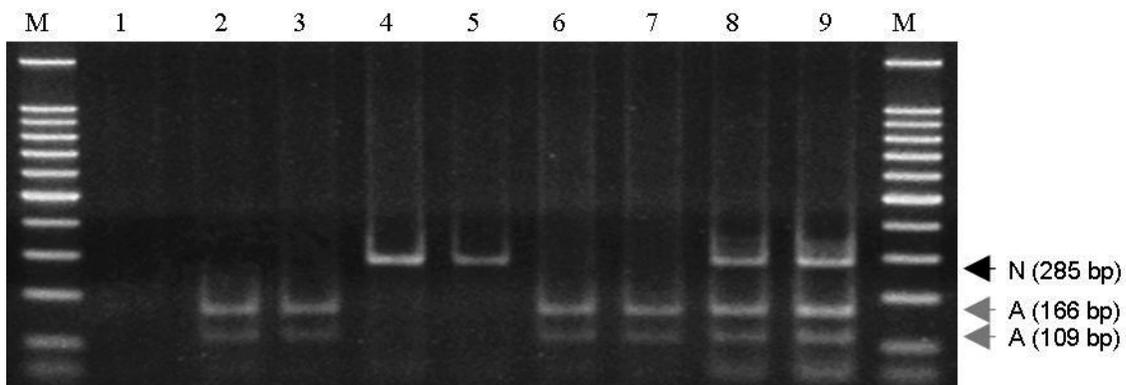
que permite diseñar iniciadores específicos que permitirán su amplificación por PCR, y tener una localización única en el cromosoma o genoma estudiado. Estos marcadores son sumamente usados en mapeo físico detallado de genomas grandes para ensamblar, por ejemplo, clones de BAC entre sí.

Alternativamente, para una utilización más eficiente de marcadores moleculares en programas de mejoramiento existe la posibilidad de convertir marcadores complejos y costosos o inestables como RFLPs, AFLPs y RAPDs, en marcadores PCR alelo-específicos, que son económicos, robustos y de sencilla implementación en cualquier laboratorio de biología molecular (los marcadores PCR alelo-específicos también serían STS en el sentido de que se trata de una secuencia de ADN corta que identifica un locus específico del genoma y puede ser amplificada por PCR). La estrategia consiste en aislar el fragmento de ADN polimórfico del gel, se clona, se secuencia y se diseñan iniciadores específicos de alrededor de 20 pb, para amplificar por PCR dicho fragmento. Cuando se parte de un fragmento RAPD, el marcador derivado mediante este proceso ha sido descrito como SCAR (*Sequence Characterized Amplified Region*, Paran y Michelmore 1993).

El problema que enfrentan estas estrategias suele ser el bajo nivel de polimorfismo entre

variedades de una misma especie (trigo, soja, etc), lo que disminuye las posibilidades de encontrar mutaciones útiles para desarrollar estos marcadores. De todos modos, existen antecedentes exitosos de desarrollo de marcadores de PCR alelo-específicos en trigo para caracteres de interés agronómico como pueden ser alelos de gluteninas, genes de resistencia a patógenos, genes vinculados a adaptación, etc.

Una segunda variante de los STS son los marcadores CAPS (*Cleavage Amplified Polymorphic Sequence*) o "secuencia polimórfica amplificada y clivada", (Konieczny y Ausubel 1993). En esta técnica, también conocida como PCR-RFLP, un fragmento de ADN es amplificado por PCR, luego digerido con enzimas de restricción y resuelto por electroforesis en agarosa o poliacrilamida. Este método aprovecha la capacidad que posee la técnica de PCR de amplificar segmentos específicos de ADN con la capacidad que poseen las enzimas de restricción de cortar el ADN en secuencias blanco de la enzima de restricción. La base genética del polimorfismo detectado por el marcador CAPS es la presencia/ausencia de sitios de restricción en la secuencia amplificada por PCR. Este marcador permite identificar individuos heterocigotas, por lo que se comporta como marcador codominante (Figura 6).



**Figura 6.** Marcador CAPS para seleccionar gen Lr37 de resistencia a roya de la hoja en trigo. En todas las muestras los fragmentos de PCR amplificados con los iniciadores URIC-LN2 fueron digeridos con la enzima Dpn II y sometidos a electroforesis en geles de agarosa 2%. La flecha negra (285bp) indica el fragmento de ADN asociado al gen Lr37 y las flechas grises (166 bp y 109 bp) indican fragmentos de ADN asociados a la ausencia del gen Lr37. Las calles 1, 2 y 3 son líneas nulísómicas de trigo para los cromosomas 2A, 2B y 2D; las calles 4 a 9 corresponden a individuos de una población F2 segregante para Lr37 (calles 4 y 5 individuos portadores del gen Lr37 en forma homocigota; calles 6, 7 individuos sin el gen Lr37; calles 8 y 9, individuos portadores del gen Lr37 en forma heterocigota). M: estándar de peso molecular de ADN. Fotografía de M. Helguera.

Los marcadores STS y sus variantes SCAR y CAPS, tienen la ventaja de requerir un ensayo metodológico sencillo y altamente reproducible por los que son fácilmente transferibles a otros laboratorios, permitiendo que sean muy usados en selección asistida por marcadores. Asimismo, según las características de su diseño, pueden ser codominantes y ser analizados en simultáneo varios loci STS en un mismo gel.

Actualmente, dada la disponibilidad de secuencias en base de datos es posible también diseñar marcadores STS directamente a partir de secuencias de interés (secuencias genómicas, secuencias ESTs, etc) evitándose de este modo el trabajo previo de secuenciación del STS.

## 5.5 Marcadores Funcionales o de Expresión

En la actualidad, las tecnologías de marcadores moleculares en plantas superiores han sufrido un cambio. Así, la mayoría de los marcadores descritos anteriormente derivan del ADN genómico, y por lo tanto pueden pertenecer a regiones transcritas y no transcritas del genoma. Estos marcadores basados en ADN derivado de cualquier región del genoma, han sido descritos como marcadores de ADN al azar (RDMs: *Random DNA Markers*). Sin embargo, durante los últimos años se ha observado una fuerte tendencia hacia el desarrollo de marcadores moleculares derivados de la región transcrita del genoma (genes). Esto ha sido posible gracias a la disponibilidad de un gran número de secuencias de clones de ADNc (ARNm transcrito a ADN empleando transcriptasa reversa, también conocidos como *Expressed Sequence Tags* o ESTs), en bases de datos públicas (TIGR, NCBI; EBI, etc., ver Parte I, Capítulo 12, "Análisis informático de secuencias moleculares").

Los tipos de marcadores desarrollados a partir de la región expresada del genoma son numerosos, complejos y variados. Sólo se describirán en detalle aquellos marcadores más relevantes y que no fueron enunciados anteriormente, recomendándose la lectura de la revisión de Gupta y Rustgi (2004) para información adicional.

### 5.5.1 Marcadores COS (*Conserved Orthologue Set*)

Los marcadores COS representan genes funcionales conservados en un amplio rango de plantas dicotiledóneas. Inicialmente, estos marcadores fueron desarrollados a partir de la comparación de la secuencia genómica de *Arabidopsis* con la base de datos de ESTs de tomate. Se postula que estos marcadores podrán ser aplicados en el mapeo comparativo entre genomas emparentados y, por lo tanto, serán útiles para estudios taxonómicos y en la deducción de las relaciones filogenéticas.

### 5.5.2 Marcadores GTMs (*Gene Targeted Markers*)

Los GTMs se basan en polimorfismos dentro de genes, independientemente de si se conocen o no sus funciones. Estos marcadores se pueden desarrollar a partir de secuencias disponibles en las bases de datos (cDNA/EST/secuencias genómicas que representan genes) o a partir de genotecas de ADN genómico enriquecidas para secuencias de genes. Un ejemplo de GTMs son los Análogos a Genes de Resistencia (*Resistance Gene Analogs*, RGAs) que son marcadores basados en secuencias derivadas de genes de resistencia a patógenos. Las secuencias RGAs son útiles para la identificación de nuevos genes de resistencia a patógenos en plantas.

### 5.5.3 Marcadores Funcionales propiamente dichos (FM: *Functional Markers*)

El desarrollo reciente de varios proyectos de genómica estructural y funcional en especies cultivadas, ha generado un enorme caudal de información a nivel de secuencias y esto ha posibilitado el desarrollo de un nuevo tipo de marcadores denominados marcadores funcionales (FM). Estos marcadores derivan de motivos polimórficos (mutaciones) dentro de los genes, y estos polimorfismos afectan directamente la expresión fenotípica del carácter asociado al gen (Andersen y Lübberstedt 2003). Los FM reciben también el nombre de marcadores diagnósticos, marcadores de genes blanco o marcadores perfectos (*perfect markers*).

El desarrollo de FM requiere de secuencias alélicas de genes caracterizados funcional-

mente, a partir de los cuales puedan identificarse motivos polimórficos funcionalmente responsables de afectar el fenotipo de la planta.

El primer paso en el desarrollo de FM es disponer de la secuencia de un gen con una función asignada. En *Arabidopsis*, menos del 10% de sus aproximadamente 25.000 genes han sido caracterizados funcionalmente. En otras especies, el número de secuencias caracterizadas funcionalmente es sustancialmente menor. Sin embargo, considerando la identidad de secuencias es posible asignar funciones putativas (probables) a un 30-50% de las secuencias expresadas (ESTs) en otras especies. Esta estrategia de considerar genes candidatos y las relaciones sinténicas (similitud) entre genomas de plantas ha sido explotada exitosamente para identificar genes agrónomicamente importantes.

El segundo paso en el desarrollo de FM es la búsqueda de secuencias polimórficas o alelos del gen en cuestión. Para ello se deben analizar las secuencias del gen en más de un genotipo por especie. Las estrategias de desarrollo de marcadores funcionales varían desde la detección de polimorfismos de un solo nucleótido (SNPs), el descubrimiento en regiones génicas de secuencias tipo microsatélites (SSRs), la identificación de sitios de restricción que diferencien alelos (CAPS), la detección de inserciones y/o deleciones (InDels), las variaciones a nivel de secuencias que se pueden revelar empleando la metodología de conformación de ADN simple cadena en electroforesis parcialmente desnaturalizantes (SSCP), etc.

Posteriormente, se debe probar la existencia de variación fenotípica asociada al alelo en cuestión. Esto puede realizarse indirectamente, por estudios de asociación en poblaciones segregantes, o directamente, mediante la comparación de genotipos isogénicos generados por mutagénesis específica en la secuencia del gen mediante TILLING, *Targeting-Induced Local Lesions In Genomes* (McCallum et al., 2003 Parte II; Capítulo 4 “Mutagénesis, TILLING y EcoTILLING”), etc.

En varias aplicaciones, la gran ventaja que presentan los FM es su ligamiento completo con motivos (mutaciones) funcionales. En contraste con los marcadores tradicionales (distribuidos al azar en el genoma), los FM permiten:

i) la aplicación confiable de marcadores en poblaciones sin necesidad de mapeo previo; ii) el uso de marcadores en poblaciones segregantes sin el riesgo de perder información debido a la recombinación; iii) una mejor representación de la variación genética en poblaciones naturales o de mejoramiento.

En general, los FM son útiles para: i) la fijación más eficiente de alelos en cualquier población segregante; ii) el análisis de alelos tanto en poblaciones naturales como de mejoramiento; iii) la combinación de alelos de FM que afectan idénticos o diferentes caracteres en mejoramiento de plantas; iv) la construcción de haplotipos de FM ligados.

En la actualidad se encuentran disponibles varios genes con potencial para el desarrollo de FM sobre caracteres de importancia agrónómica, por ejemplo, altura de planta en cereales; tiempo de floración en maíz, requerimientos de vernalización en *Brassica* y trigo, tamaño de fruto en tomate, calidad de alimento en arroz; resistencia a enfermedades y respuesta a stress en varias especies, etc. Sin embargo, aún en especies modelo, sólo el 10% de los genes han sido caracterizados funcionalmente mientras que para especies no modelo, este número es probablemente menor al 5%.

Los FM derivan de estudios de asociación realizados en grandes colecciones de genotipos o de la comparación de líneas isogénicas, las cuales tienen un costo alto para su desarrollo, limitando el estudio inicial a uno o unos pocos fondos genéticos. Luego estos FM deben ser evaluados en otros fondos genéticos, con lo cual, para todo esto es necesario desarrollar un marco experimental, estadístico y bioinformático para la acumulación de datos y las evaluaciones a lo largo de diferentes estudios.

Teniendo en cuenta estos factores y el laborioso desarrollo de los FM, la generación de FM debería concentrarse en genes que confieran una variación fenotípica sustancial. Así, la selección de “genes claves” para caracteres de interés agrónómico será crucial y el paso limitante para el desarrollo de FM útiles.

#### 5.5.4 Arreglos de ADN para estudios de expresión de genes

Los arreglos de ADN (*arrays*) permiten, en-

tre otras cosas, analizar sistemáticamente el patrón de expresión génica de un organismo en escala genómica. De este modo, es posible examinar la expresión simultánea de cientos a miles de genes en varios estadios del desarrollo, en respuesta a diferentes condiciones ambientales y en tejidos específicos.

Los arreglos de ADN son soportes sólidos, comúnmente vidrio o *nylon*, en los cuales están fijadas de forma ordenada gran cantidad de secuencias de ADN de interés (marcadores). Estas secuencias pueden ser genes completos o parciales. El empleo de estos arreglos como herramienta para el estudio de expresión de genes a escala genómica sigue típicamente los siguientes pasos: i) la construcción del arreglo de ADN; ii) la preparación de sondas (a partir de ARNm de las distintas situaciones a comparar); iii) la hibridización del arreglo; iv) la detección de la señal y análisis de los datos por herramientas computacionales apropiadas.

Existen diferentes sistemas de arreglos de ADN, dependiendo de la fuente de ADN de interés, naturaleza del soporte sólido y el sistema de detección de la hibridización. El sistema más simple es el que requiere menos inversión, comúnmente se lo llama macroarreglo (*macroarray*) y emplea una membrana de *nylon* como soporte en combinación con sondas radiactivas. El sistema más sofisticado utiliza láminas de vidrio y sondas fluorescentes y se lo llama microarreglo (*microarray*). Asimismo, en los macroarreglos se depositan centenas de secuencias de ADN mientras que en los microarreglos la densidad de secuencias es al menos un orden de magnitud mayor. Es importante destacar la existencia de distintos términos empleados para describir estos arreglos, de modo que *glass array*, *chips* de ADN y *biochips* son denominaciones usadas para los microarreglos sobre placas de vidrio, en cuanto que *nylon array* y *high density membranes* son para los macroarreglos sobre *nylon*.

En el área vegetal, la utilización de arreglos de ADN para el análisis de la expresión génica aumentó significativamente de la mano de los proyectos de secuenciación completa del genoma de *Arabidopsis thaliana* y *Oryza sativa* (arroz) (para más detalles ver Rensink y Buell, 2005).

Además de la contribución para el entendimiento de la expresión génica a gran escala, se vislumbra la posibilidad de emplear la técnica de arreglos de ADN en mejoramiento asistido por marcadores, *fingerprinting*, selección de accesiones de germoplasma, desarrollo de plantas tolerantes a diferentes tipos de estrés, etc. Asimismo, la identificación de genes expresados diferencialmente, en especial aquellos cuya expresión no es abundante o están involucrados en cascadas de transducción de señales, ampliará enormemente los horizontes una vez que esos genes puedan ser empleados como marcadores para selección asistida (Galbraith, 2006).

### 5.6 Marcadores SNPs (Single Nucleotide Polymorphisms)

Si bien, son una clase particular de marcadores moleculares que podrían ser incluidos dentro del punto 5.4, se los decidió tratar aparte dado que los SNPs pueden derivar tanto de secuencias genómicas codificantes (que se expresan) como no codificantes.

Recientemente, de la mano de los proyectos de secuenciación de genomas enteros de plantas, una nueva clase de marcadores moleculares denominados SNPs (polimorfismo de un nucleótido), está siendo utilizada en distintos estudios genéticos. Estos marcadores se basan en la detección de polimorfismos resultantes de la alteración de una única base en una secuencia de ADN. En la figura 7 se muestran secuencias de ADN ejemplificando algunas variaciones de punto, las cuales se definen como SNPs. Algunos autores consideran SNPs sólo cuando ocurre una sustitución de base, otros consideran también la inserción/delección de una base (InDel).

Los SNPs pasaron a tener mayor importancia a partir de la secuenciación del genoma humano, al descubrirse que el 90% de polimorfismo encontrado en el genoma eran SNPs. Recientemente en el área vegetal, la secuenciación a gran escala de genomas completos y de secuencias expresadas (ESTs), ha permitido evidenciar la presencia de SNPs en especies como *Arabidopsis thaliana*, melón, soja, girasol, arroz, maíz, cebada, trigo, y caña de azúcar entre otras.

```

CACTTATTCAATATTAATTATGATAATGAAATAAAGGTTAAGGTGGATGGAAC
CACTTATTCAATATTAAGTTATGATAATGAAATAAAGGTTAAGGTGGATGGAAC

ATGTATATGTATATGTATATGTATATGTATATGTATATGTATACTCATCAATCAAAACCT
ATGTATATGTATATGTATATGTATATGTATA-----CTCATCAATCAAAACCT

```

**Figura 7.** Ocurrencia de SNPs por la comparación de secuencias de ADN de dos individuos diferentes de girasol (variedades PAC2 y RHA266). De izquierda a derecha, obsérvese el primer SNP en la posición 17 de la primera secuencia (A-G) y el segundo SNP (inserción/delección TGTATA) a partir de la posición 31 de la segunda secuencia. Gentileza C.Fusari.

Los marcadores genéticos SNP poseen naturaleza bialélica y son muy abundantes en el genoma, siendo la frecuencia de SNPs de 1 cada 100-300 pb para los genomas de plantas. Los SNPs pueden localizarse en regiones codificantes, regulatorias y no codificantes. Cuando están presentes en regiones codificantes, muestran 100% de asociación con el carácter de interés, por lo que son muy útiles en MAS (*marker assisted selection* o selección asistida por marcadores) y en el aislamiento de genes.

Debido a su frecuencia de distribución, los SNPs surgen como importantes marcadores para la obtención de mapas genéticos de alta resolución. Estudios llevados a cabo utilizando *Arabidopsis* como organismo modelo, han demostrado que estos marcadores posibilitan la obtención de mapas de una resolución cerca de 100 veces superior a la obtenida con los marcadores convencionales.

Los SNPs son estables desde el punto de vista evolutivo, lo que facilita su empleo en estudios de poblaciones. Otra característica importante es que la genotipificación de los SNPs no está basada en la medida del tamaño de los alelos como ocurre con los otros marcadores moleculares, y la distinción de los alelos puede ser automatizada. Por lo tanto, los SNPs, por tener una tasa de mutación relativamente baja, ser mucho más frecuentes en el genoma y ser detectables en forma automatizada, surgen como importantes marcadores genotípicos.

En tiempos recientes se han desarrollado varias metodologías para detectar SNPs que utilizan diferentes estrategias para comparar regiones específicas del ADN obtenidas de varios individuos. La elección de uno de los métodos dependerá de muchos factores: costos,

potencial para el procesamiento de los datos generados, los equipamientos necesarios y la dificultad de los ensayos.

Inicialmente, los abordajes para la detección de SNPs consistían en amplificar y secuenciar fragmentos genómicos equivalentes de regiones génicas específicas del ADN de varios individuos y comparar sus secuencias buscando SNPs. La adopción de esa estrategia involucra el diseño de iniciadores para amplificar segmentos de ADN de entre 400 a 700 pb, derivados frecuentemente de genes de interés o provenientes de ESTs. Para la amplificación se emplean ADNs de varios individuos, preferencialmente representativos de la diversidad de la población de interés. Estos productos de amplificación resultantes se secuencian directamente y las secuencias resultantes se alinean, empleando programas bioinformáticos apropiados para la identificación de polimorfismos (SNPs).

Actualmente, la secuenciación a gran escala del genoma de varios organismos permite la identificación de SNPs mediante la comparación de millares de secuencias depositadas en las bases de datos, incluyendo clones genómicos y, principalmente, secuencias de ADNc y/o ESTs.

Independientemente del método utilizado para identificar los SNPs es indiscutible la contribución de la bioinformática en dicho proceso. Por otra parte, la necesidad de la validación experimental de los datos es indispensable. Actualmente, están disponibles distintos métodos de validación y genotipado (detección), uno de ellos son los chips de ADN o microarreglos de ADN. Como se dijo anteriormente, estos chips consisten en arreglos de oligonucleó-

tidos de secuencia conocida depositados sobre un soporte sólido. Para el caso de la detección de SNPs, los oligonucleótidos difieren entre sí en sitios específicos en nucleótidos individuales (en el sitio del SNP). La técnica es apropiada para analizar varios SNPs en paralelo a partir de cada muestra (en forma multiplex). En una columna del arreglo cuatro oligonucleótidos diferirán solamente en el sitio del SNP (tendrá cada uno, una de las cuatro bases en la posición del SNP). Cuando este arreglo es hibridado con productos de PCR o fragmentos de ADN genómico obtenidos por nebulización (romper el ADN en fragmentos de un tamaño determinado y al azar), estos se unirán solo a los oligonucleótidos perfectamente complementarios y los productos no perfectamente complementarios (con *mismatch*) serán lavados. La unión perfecta de cada caso podrá ser tomada por un sistema de detección. Así, en un único microarreglo (chip) es posible analizar miles de SNPs, correspondientes a distintos loci en simultáneo.

Otros métodos de análisis de SNPs incluyen: cromatografía líquida de alta resolución desnaturalizante (DHPLC), espectrometría de masa, ensayos con extensión de iniciadores, PCR en tiempo real, minisequenciación, secuenciación de una única base, digestión por enzimas de restricción (RFLP), CAPS, entre otros. Puede encontrarse una revisión detallada de estos métodos en Kwok (2001), Tsuchihashi y Dracopoli (2002).

En cuanto a las aplicaciones y perspectivas de los marcadores SNPs cabe destacar que han sido fundamentales para el análisis de genes y el descubrimiento de la base genética molecular de importantes características agronómico-industriales. Como por ejemplo en arroz, el descubrimiento de SNPs involucrados en la calidad de cocción, procesamiento y aroma del grano (Larkin y Park, 2003).

La comparación directa de SNPs también ha sido empleada en estudios de genética de poblaciones y filogenia, identificación de variedades, construcción de mapas genéticos y físicos, análisis funcionales, análisis de asociación en mejoramiento genético vegetal etc.

El uso de métodos masivos de análisis, como los arreglos de alta densidad de oligonu-

cleótidos, se están empleando en la actualidad para identificar *Single Feature Polymorphisms* (SFPs, Polimorfismos de Característica Única) como una alternativa atractiva para detectar SNPs y otro tipo de mutaciones entre genotipos. Los SFPs son detectados sobre arreglos de alta densidad de oligonucleótidos (chips) y representan polimorfismos de secuencia de ADN entre dos genotipos dentro de un oligonucleótido individual, el cual se detecta por diferencia en la afinidad de hibridación. El término "característica (*feature*)" refiere a la señal de hibridación dada por una sonda (oligonucleótido) en el arreglo (Zhu y Salmeron, 2007).

El descubrimiento de los SNPs, asociado a la posibilidad de utilizarlos como marcadores, permite a los investigadores explorar nuevas hipótesis. Se considera que muchos de estos SNPs están localizados en el interior de secuencias génicas, esto puede significar una importante reducción en el tiempo y costos para el descubrimiento de genes de interés, comparado con los marcadores actualmente disponibles. Asimismo, el desarrollo constante de tecnologías para su aplicación permitirá automatizar el mapeo de alta densidad y, consecuentemente, reducir los costos de su empleo en estudios moleculares a gran escala.

## 5.7 Lecturas recomendadas

- Andersen J.P & Lübberstedt T. 2003. Functional markers in plants. *Trends Pl. Sc.* 8: 554-560.
- Botstein D., White R.L., Skolnick M. & Davis R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331. 1980.
- Branlard G., Dardevet M., Saccomano R., Lagoutte F. & Gourdon J. 2001. Genetic diversity of wheat storage proteins and bread wheat quality. *Euphytica* 119: 59-67.
- Caetano-Anollés, G., Bassam, B.J. & Gresshoff, P.M. 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Bio/Technology* 9: 553-557.
- Galbraith D.W. 2006. DNA Microarray Analysis in Higher Plants. *OMICS: A Journal of Integrative Biology.* 10(4): 455-473.
- Gupta P.K. & Rustgi S. 2004. Molecular markers from the transcribed/expressed region of the ge-

- nome in higher plants. *Funct. Integr. Genomics* 4:139-162.
- Konieczny A., & F. Ausubel. 1993. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR based markers. *Plant J.* 4:403-410.
- Kwok P-Y. 2001. Methods for genotyping Single Nucleotide Polymorphisms. *Annu. Rev. Genomics Hum. Genet.* 2: 235-258.
- Larkin P.D. & Park W.D. 2003. Association of waxy gene single nucleotide polymorphsim with starch characteristics in rice (*Oryza sativa* L.). *Mol Breed* 12:335-339.
- Litt M. & Luty J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44:398-401. 1989.
- McCallum C.M., Comai L., Greene E.A. & Henikoff S. 2003. Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123:439-442.
- Olson M., Hood L., Cantor C. & Dotstein D. 1989. A common language for physical mapping of the human genome. *Science* 245:1434-1435.
- Paran I. & Michelmore R.W. 1993. Development of reliable PCR based markers linked to downy mildew resistance genes in lettuce. *Theor. Appl. Genet.* 85:985-993.
- Rensink W.A. & Buell C.R. 2005. Microarray expression profiling resources for plant genomics. *Trends Pl. Sc.* 10(12): 603-609.
- Steinmetz L.M. 2000. Combining genome sequences and new technologies for dissecting the genetics of complex genotypes. *Trends Pl. Sc.* 5(9): 397-401.
- Tsuchihashi Z. & Dracopoli N.C. 2002. Progress in high throughput SNP genotyping methods. *Pharmacogenomics* 2: 103-110.
- Vos P., Hogers M., Bleeker M., Rijans M., Van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M. & Zabeau M. 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* 23:4407-4414.
- Welsh J. & McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* 18:7213-7218.
- Williams J.G., Kubelik A.R., Livak K.J., Rafalski L.A. & Tingey S.D. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18:6531-6535.
- Zhu T. & Salmeron J. 2007. High-definition genome profiling for genetic marker discovery. *Trends Pl. Sc.* 12(5): 196-202.

## I CAPITULO 6

### Construcción de mapas de ligamiento genético, localización de genes y regiones cromosómicas asociadas a caracteres de interés en plantas

Gerardo D L Cervigni,  
Juan Pablo A Ortiz, Sergio E Feingold

#### 1. Introducción

La posibilidad de construir mapas de ligamiento genético en especies vegetales o animales es una de las contribuciones de mayor impacto de las técnicas de marcadores moleculares en las ciencias biológicas. Los marcadores moleculares utilizados pueden corresponder a regiones intergénicas no codificantes o a segmentos génicos, en cuyo caso son denominados *funcionales* y constituyen marcadores “ideales” a los efectos de selección de genotipos. Un mapa genético establece de manera probabilística el arreglo lineal de un grupo marcadores (o genes) sobre el genoma de una especie. Si bien el concepto data de principios de siglo XX, a partir de los trabajos de Morgan y Bridges con mutantes de *Drosophila*, recién a partir del advenimiento de los marcadores moleculares fue técnicamente posible construir mapas genéticos saturados en la mayoría de las especies vegetales de interés agronómico. Usando este tipo de mapas es posible identificar la posición y el efecto de genes sobre caracteres de importancia mediante asociaciones estadísticas entre los valores fenotípicos y las variantes alélicas de los marcadores. La disponibilidad de mapas genéticos permite también la selección indirecta de genotipos deseables, comúnmente denominada *MAS* (del inglés *Marker Assisted Selection*), mediante el seguimiento de marcadores localizados en regiones genómicas determinadas. La utilización de marcadores comunes permite comparar la estructura del genoma de diferentes especies (*comparative mapping*) e identificar rearrreglos cromosómicos a pequeña y gran escala (micro y macro *sintenia*) para estudios de filogenia y evolución molecular. Por otro lado, el desarro-

llo de mapas genéticos altamente saturados permite el clonado posicional de genes.

#### 2. Introducción a la construcción de mapas genéticos en vegetales

Para comprender mejor las bases teóricas y metodológicas empleadas en el mapeo genético es necesario recordar algunos conceptos básicos de genética Mendeliana. Gregor Mendel realizó sus experimentos en *Pisum sativum* (arveja cultivada) efectuando cruza- mientos controlados entre individuos que se diferenciaban en determinadas características (como por ejemplo el color de la flor, tipo de tegumento de las semillas y largo de los tallos), y estudió la forma en que estos caracteres se transmitían a la descendencia. A partir de datos experimentales formuló una hipótesis simple para explicarlos y posteriormente diseñó expe- rimentos para contrastarla. Sus trabajos permiti- eron inferir la existencia de “factores” hereditarios y sus mecanismos de transmisión antes de conocer la existencia del ADN como mate- rial de transmisión genético. En base a sus da- tos experimentales Mendel formuló su prime- ra ley, o ley de la segregación. Ella establece que los miembros de un par génico (alelos) se segregan (se separan) uno de otro durante la formación de las gametas. Como consecuen- cia, la mitad de las mismas portan un alelo y la otra mitad el otro. La progenie es luego for- mada por la combinación al azar de las game- tas de ambos progenitores. La segunda ley de Mendel, o ley de la segregación independiente, indica que la segregación de los alelos de un gen es independiente de la segregación de los alelos de otro gen. Estas dos leyes fueron los puntos de partida sobre los cuales se elaboró toda la teoría de herencia que hoy conocemos. Incluso aún en el presente, los análisis de las frecuencias en que aparecen los caracteres en las progenies de cruzamientos controlados son una parte fundamental de los análisis ge- néticos. Los caracteres genéticos pueden estar controlados por uno o pocos genes, o poseer un control complejo que involucra muchos ge- nes, denominados *loci* de caracteres cuantitati- vos o *QTLs* (del inglés *Quantitative Traits Loci*), a menudo afectados por el ambiente. La dis- ponibilidad de un gran número de marcadores

(*loci*) polimórficos, no afectados por el medio ambiente, neutros, sin efectos deletéreos y cuya herencia puede determinarse con relativa facilidad, simplificó enormemente los análisis genéticos en especies vegetales y animales. La construcción de un mapa de ligamiento genético es esencialmente una representación gráfica lineal del orden más probable de marcadores moleculares o genes a partir de los valores de recombinación observados entre ellos. Cada arreglo lineal es denominado “grupo de ligamiento”. En teoría un mapa saturado debe contener el mismo número de grupos de ligamiento que número básico de cromosomas.

En este Capítulo se desarrollarán los conceptos básicos del mapeo genético en vegetales y algunas de sus aplicaciones. De ninguna manera se pretende abordar el tema en toda su extensión debido a que la misma supera largamente los objetivos de este libro. Aquellos lectores que deseen ampliar y profundizar los temas que aquí se desarrollan pueden consultar los textos que se citan en la lista de referencias bibliográficas al final del Capítulo.

Los pasos fundamentales a seguir en la construcción de un mapa genético de una determinada especie o el mapeo específico de un gen (*locus*) de interés pueden resumirse en: **1)** selección de progenitores y desarrollo de una población segregante (población de mapeo), **2)** generación de un número suficiente de marcadores y registro de los mismos en cada uno de los individuos de la población (*genotipado*), **3)** determinación de las frecuencias de recombinación entre marcadores y determinación de grupos de ligamiento y **4)** determinación del orden de los marcadores y conversión de la frecuencia de recombinación (*r*) en unidades de mapeo o centiMorgan (cM).

### 2.1. Poblaciones de mapeo

El primer paso en todo proyecto de mapeo es el desarrollo de una población segregante para una o más características de interés. La selección de los progenitores es un punto importante debido a que no sólo deben ser contrastante para la característica en estudio, sino que además deben presentar diferencias a nivel de secuencias del ADN y así obtener suficientes marcadores polimórficos que per-

mitan la construcción del mapa genético. En ausencia de polimorfismos genéticos los análisis de segregación y ligamiento son impracticables. En general los cruzamientos entre especies alógamas presentan mayor grado de polimorfismo que entre autógamas. La falta de polimorfismos es común en poblaciones derivadas de cruzamientos entre progenitores de base genética estrecha, como los realizados entre distintos cultivares de la misma especie o entre individuos de especies silvestres derivados de la misma población natural. En teoría es posible desarrollar varios tipos de poblaciones segregantes y la decisión de cual utilizar depende de la resolución que se desee alcanzar con el mapa y la posibilidad y/o facilidad de desarrollarla. En la Figura 1 se muestra un esquema del desarrollo de 5 poblaciones básicas de mapeo. Las poblaciones de tipo retrocruza (*backcross* o  $BC_1$ ) y  $F_2$  son generalmente adecuadas y fáciles de generar en la mayoría de las especies cultivadas como: trigo, maíz y soja. Las poblaciones  $F_2$  proveen mayor información genética que las retrocruzas para el mismo número de individuos debido a que pueden evaluarse dos cromosomas recombinantes por planta. En especies autoincompatibles y altamente heterocigotas las poblaciones  $F_1$  también son segregantes y pueden emplearse para mapeo. Estos 3 tipos de familias son efímeras, es decir, que cada individuo segre-

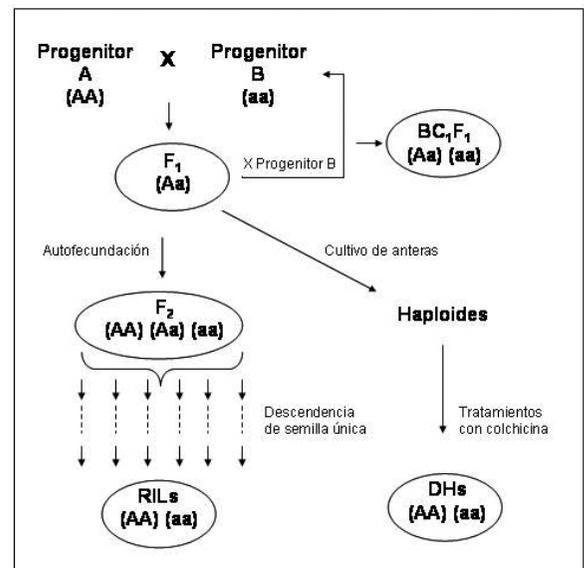


Figura 1.

gará las combinaciones génicas que porta en la siguiente generación. Si no es posible mantenerlas clonalmente como por ejemplo a partir de órganos vegetativos o por cultivo *in vitro*, la información obtenida en ellas debe ser transferida a una nueva población utilizando algunos de los marcadores que fueron previamente localizados (mapeados) de manera de conservar puntos de referencia entre ambas.

En el caso de requerirse múltiples ensayos, (repeticiones intra experimento, ensayos en distintas localidades o en varios años), es necesario desarrollar poblaciones de mapeo inmortal como las Líneas Endocriadas Recombinantes (RILs por *Recombinant Inbred Lines*) o poblaciones de Haploides Duplicados (DHs). Ambas poblaciones se componen de un conjunto de líneas homocigotas que pueden ser replicadas por semillas varias veces y por varios ciclos de cultivo. Estas características las hacen especialmente adecuadas para mapear QTLs. Las RILs son generadas a partir de una población  $F_2$  por descendencia de semilla única de un número determinado de individuos durante cinco o más generaciones. Finalmente, cada línea representará los diferentes bloques de ligamiento presentes en un individuo  $F_2$  de la población original, pero en estado homocigota. Una vez estabilizadas, las RILs pueden ser mantenidas y multiplicadas por semillas año tras año. Una de las desventajas de este método es que algunas regiones del genoma pueden seguir en el estado heterocigota a pesar de las múltiples autofecundaciones. Por otro lado, no es posible desarrollarlas en especies autoincompatibles. Las poblaciones DH se generan a partir del cultivo de tejidos de anteras de individuos  $F_1$ . En algunas especies es posible regenerar plantas enteras a partir de microgametofitos haploides (grano de polen en el estado uninucleado). Si la planta dada es una  $F_1$ , sus granos de polen portarán las gametas derivadas de la meiosis con diferentes combinaciones alélicas. Posteriormente, por tratamientos de duplicación cromosómica con colchicina, las plantas haploides obtenidas pueden diploidizarse llevando todos los *loci* al estado homocigota. Como resultado se obtiene una serie de líneas que representan las combinaciones génicas presentes en las gametas

de la  $F_1$ . La desventaja de este método es que es necesario disponer de un protocolo eficiente para la regeneración de plantas *in vitro* a partir del cultivo de anteras y la posterior duplicación cromosómica.

Una vez establecido el tipo de población a utilizar, el número de individuos es también un punto importante debido a que la resolución del mapa y el ordenamiento de los marcadores en los grupos de ligamiento dependen de éste. En general, poblaciones de 60 a 100 individuos son adecuadas para trabajos de tipo académico, como la localización de un determinado *locus* en el genoma o el mapeo de un carácter cualitativo. Sin embargo, para proyectos de mapeo de QTLs o la construcción de mapas de alta resolución son necesarias poblaciones de 500 a 1000 individuos.

## 2.2. Registro de marcadores moleculares y clasificación de la población

Los marcadores moleculares útiles para mapeo deben existir en 2 o más formas alélicas distinguibles, de manera que los genotipos de cada individuo puedan ser claramente identificados. Los mismo pueden ser de tipo dominante (presencia o ausencia de una banda como en RAPD y AFLP) por lo cual no es posible distinguir entre el genotipo homocigota (AA) del heterocigota (Aa) o co-dominantes (todos los alelos pueden identificarse, por ej., usando SSR y RFLP), donde los 3 genotipos de un mismo *locus* dialélico pueden ser discriminados (AA, Aa y aa). Como criterio general una vez que se dispone de entre 100-200 marcadores en una determinada población es posible identificar grupos de ligamiento, regiones genómicas específicas y mapear genes o QTLs. Asimismo, prácticamente cualquier nuevo marcador que se agregue puede integrarse a los preexistentes. El genotipo de cada individuo de la población, para cada uno de los marcadores debe ser determinado (genotipado). En general se toman como marcadores informativos aquellos que muestran relaciones de segregación, por presencia/ausencia, que se ajusten a las frecuencias esperadas considerando el tipo de población y marcador molecular utilizado (por ej.: 1 Aa: 1 aa en una retrocruza o 3 A\_: 1aa en una  $F_2$ ), por medio de la prueba del ji-cuadrado ( $X^2$ ).

### 2.3. Estimación de las frecuencias de recombinación y detección de grupos de ligamiento

#### 2.3.1. Concepto de ligamiento

Dos *loci* que se encuentran muy cercanos en el mismo cromosoma no segregan independientemente durante la meiosis. Esta relación se denomina ligamiento genético y explica la aparición de una proporción mayor de individuos portando combinaciones alélicas parentales que lo esperado según la Segunda Ley de Mendel. Los genes ligados pueden ser separados por un intercambio físico de partes de cromosomas (cromátidas hermanas) durante la meiosis por el proceso de entrecruzamiento o "crossing-over" (CO), el cual genera gametas recombinantes, distintas de las parentales (Figura 2).

para que el número de plantas recombinantes sea igual al de no recombinantes, se dice que los mismos segregan independientemente. Por otro lado, si el número de plantas recombinantes es menor al esperado (50%) se considera que ambos *loci* están ligados genéticamente. De esta manera, determinando la frecuencia de aparición de individuos con combinaciones alélicas recombinantes es posible realizar una medida relativa de la distancia entre ellos. **2.3.2**

#### Detección del ligamiento

El primer paso en el estudio del ligamiento entre 2 *loci* (por ej., A y B) consiste en determinar si los mismos segregan individualmente de acuerdo al modelo esperado (1:1, 3:1, 1:2:1, etc.), y posteriormente, si lo hacen en forma independientemente o no. En ambos casos la prueba estadística adecuada es el  $X^2$

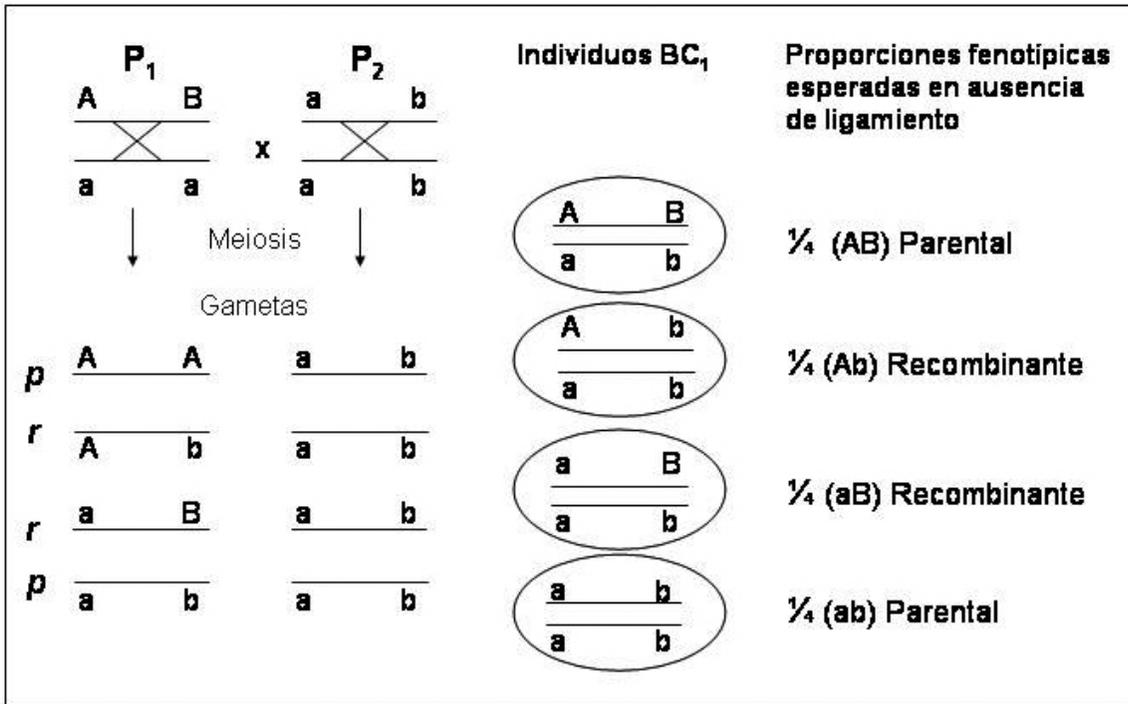


Figura 2.

Existe una relación entre la distancia física a la que se encuentran 2 *loci* y la probabilidad de que se produzca un CO entre ellos. Cuanto más alejados estén uno del otro mayor será la probabilidad de que ocurra un entrecruzamiento y se generen gametas recombinantes. Consecuentemente, si dos *loci* están lo suficientemente separados en el cromosoma como

considerando un nivel de significancia predeterminado ( $p \geq 0,05$  o  $0,01$ ), aunque también es posible utilizar el *LOD score* (ver punto 2.4). Si la probabilidad asociada al  $X^2$  obtenido para cada *locus* en forma individual es superior a la predeterminada, se considera que ambos marcadores segregan de acuerdo a lo esperado. En el segundo paso, el número de progenies

observadas en cada clase genotípica se contrasta con los valores esperados para una segregación independiente (Tabla 1).

**Tabla 1:** Segregaciones esperadas para 2 loci independientes en distintas poblaciones de mapeo.

Tipo de población	Marcador dominante	Marcador co-dominante
BC <sub>1</sub>	1:1:1:1	1:1:1:1
F <sub>2</sub>	9:3:3:1	1:2:1:2:4:2:1:2:1
RILs	1:1:1:1	1:1:1:1
DH	1:1:1:1	1:1:1:1

BC<sub>1</sub>: población de retrocruzamiento; F<sub>2</sub>: población tipo F<sub>2</sub> derivada del cruzamiento entre 2 progenitores heterocigotas; RIL: población de líneas endocriadas recombinantes (RILs por Recombinant Inbred Lines) y DH: población de haploides duplicados.

La hipótesis nula (H<sub>0</sub>) es que ambos loci segregan independientemente y la hipótesis alternativa (H<sub>A</sub>) es que ambos loci están ligados. Si los desvíos obtenidos entre los valores observados y esperados tienen una probabilidad de ocurrencia menores que 1/20 o 1/100 (p < 0,05-0,01), se considera que los loci A y B no segregan independientemente y por lo tanto se acepta H<sub>A</sub> que indica la existencia de ligamiento entre ellos (ver ejemplo Figura 3).

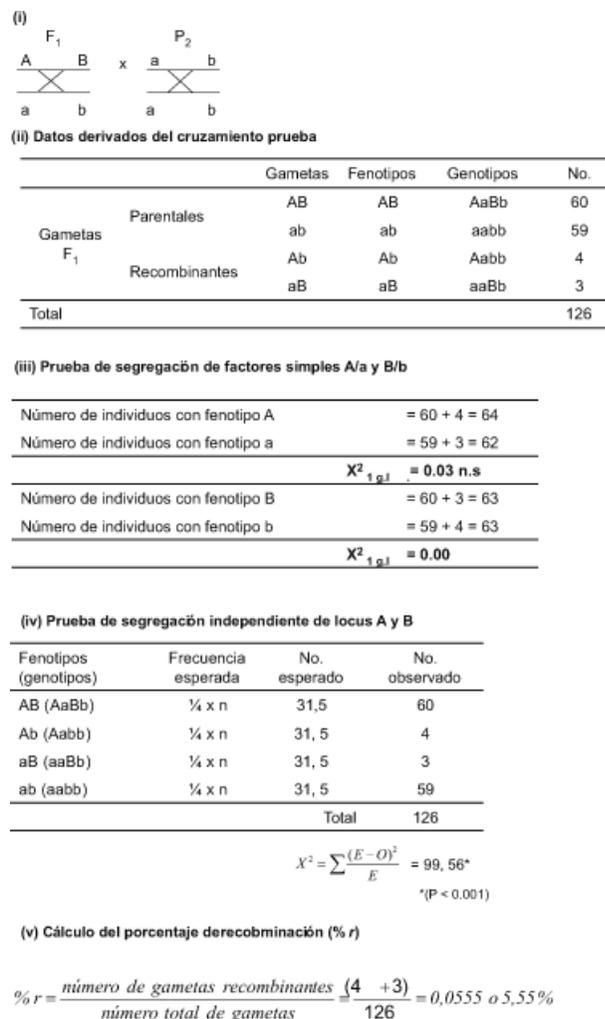
### 2.3.3. Estimación de la frecuencia de recombinación

Una vez verificada la existencia de ligamiento entre 2 loci es posible estimar la frecuencia de recombinación (r) entre ellos. En el caso más simple de una retrocruza, todas las clases fenotípicas (o genotípicas) pueden ser identificadas (Figura 3). El valor de r expresado en porcentaje se calcula:

$$r(\%) = \frac{\text{número de gametas recombinantes}}{\text{número total de gametas}} \times 100$$

Una estimación de r = 0,01 (1 %) es definida como una unidad de mapeo (u.m), y corresponde a la distancia entre dos loci en la cual

1 producto de cada 100 meiosis es del tipo recombinante. Un ejemplo de los pasos a seguir en la estimación del ligamiento entre dos genes se muestra en la Figura 3.



**Figura 3.**

En este caso la frecuencia de recombinación entre A y B es de 5,55 %. Como la frecuencia de recombinación es una proporción, su variancia (derivada de la distribución binomial) está dada por la expresión: R (1-R)/N, donde R es la frecuencia de recombinación y N es el número total de los individuos de la población. Por lo tanto, un valor de r = 5,55 % (como el de nuestro ejemplo), posee el siguiente error estándar:

$$\sqrt{[0,055 \times (1 - 0,055) / 126]} = 0,020$$

De esta manera es posible establecer el grado de ligamiento entre 2 loci. Sin embargo, en algunos tipos de poblaciones como las  $F_2$ , no todas las clases genotípicas pueden ser identificadas fácilmente, especialmente cuando se trabaja con marcadores co-dominantes. Esto es debido a que no es posible diferenciar si los dobles heterocigotas derivan de 1 o 2 gametas recombinantes. Para estos casos se han desarrollado procedimientos alternativos para estimar las distancias genéticas. El método de máxima probabilidad (*maximum likelihood method*) es el más utilizado. Debido a que se trata de un método estadístico de estimación de probabilidades las fórmulas empleadas en los cálculos son generalmente complejas. Allard (1956) desarrolló el método de los *scores* para derivar las fórmulas, estimar las distancias genéticas y calcular las desviaciones estándares para varios tipos de poblaciones. Afortunadamente en el presente se dispone de programas de computación que utilizan estas fórmulas y permiten trabajar con grandes archivos de datos derivados de poblaciones de un gran número de individuos. Entre los más populares se encuentran el Mapmaker (Lander 1987) y el JoinMap (Stam et al. 1993).

#### 2.3.4. Ordenamiento de los marcadores y conversión de frecuencias a unidades de mapeo

Cuando se estudia el ligamiento simultáneo entre 3 o más marcadores (por ej. A, B y C), se puede establecer un orden en base a las distancias relativas entre ellos (mapeo de 3 puntos). Suponiendo que el orden sea ABC, la distancia estimada en la prueba de 2 puntos entre A y C será aproximadamente igual a la suma de la distancia entre A-B + B-C. Sin embargo, frecuentemente la distancia entre A y C (medida como prueba de dos puntos) suele ser menor a la de A-B + B-C. Esto se debe a la presencia de entrecruzamientos dobles entre A y C cuyo resultado final es la formación de gametos parentales AC y ac que se computan como "no recombinantes", cuando en realidad sí lo son. Si no hubiéramos mapeado el gen

B nunca hubiéramos detectado estos dobles entrecruzamientos. Extendiendo este razonamiento, en realidad tampoco sabemos si entre A-B hubo entrecruzamientos dobles y, por lo tanto, haber subestimado la proporción de recombinantes entre ellos. Cuanto más grande es la distancia entre 2 marcadores, más inexacta será la medida de la distancia. Asimismo, otro hecho que debe considerarse es el fenómeno de interferencia (I). La interferencia se relaciona con la imposibilidad física de que adyacente a un punto de entrecruzamiento se produzca otro en la misma meiosis debido a un impedimento físico entre ellos (se "interfieren"). Estas características hacen que las frecuencias de recombinación no sean magnitudes aditivas para grandes segmentos cromosómicos. Para solucionar esto, los valores de *r* se convierten en unidades de mapeo (centMorgan – cM) mediante la aplicación de las funciones de mapeo. Las ecuaciones para convertir *r* en cM derivan de funciones de distribución y son fundamentalmente dos, la función de Haldane (1919) y la de Kosambi (1944). La diferencia básica entre ellas es el supuesto que asumen con respecto a la interferencia en la ocurrencia de dobles CO. La función de Haldane considera interferencia nula (I=0) mientras que la función de Kosambi estima el valor de I. Estos valores generalmente se encuentran entre 0 y 1 (interferencia completa). El valor real de interferencia se ubicará entonces en algún lugar en medio de ambos valores. Los valores de cM estimados con ambas funciones son equivalentes cuando se consideran valores de *r* de hasta 10 %, para valores mayores la transformación de Kosambi es más exacta.

#### 2.4. Programas de mapeo

La utilización de programas de mapeo como el Mapmaker 3.0 (Lander et al. 1987; <http://www.nslj-genetics.org/soft/mapmaker>) o JoinMap 1.4 - 4.0 (Stamp 1993; <http://www.kyazma.nl>), permiten el análisis simultáneo de muchos marcadores en poblaciones de gran número de individuos, mediante la utilización de métodos de estimación de parámetros como el de máxima probabilidad o mínimos cuadrados. Estos programas poseen diferentes procedimientos, o rutinas, que permiten la

estimación de  $r$  entre pares de marcadores, la identificación de los grupos de ligamiento y la determinación del orden de los marcadores dentro de cada grupo. El programa Mapmaker utiliza el test estadístico LOD score para evaluar el valor de  $r$  estimado y la posición más probable de un marcador en el grupo de ligamiento. El LOD score (*logaritmo de los odds*) es el logaritmo en base 10 del cociente entre la probabilidad de obtener los datos observados en caso de que los *loci* considerados estuvieran ligados, sobre la probabilidad de obtener los datos en ausencia de ligamiento. Un LOD = 3 para un par de marcadores indica que es mil veces más probable que los *loci* estén ligados respecto a que no lo estén. El programa Mapmaker también utiliza LOD score para comparar diferentes órdenes posibles entre varios genes. Al orden más probable le asigna un 0 y a los restantes le asigna valores negativos. Por su parte, el programa JoinMap (Stam, 1993) permite trabajar con distintos tipos de poblaciones e inclusive con distintos tipos de segregación dentro de la misma población. Este programa fue especialmente diseñado para estimar datos de ligamiento genético en poblaciones derivadas de especies altamente heterocigotas y/o donde no es posible obtener líneas puras. Otra de sus principales ventajas es que permite “combinar” mapas derivados de distintas poblaciones (o datos bibliográficos) si disponen marcadores en común (Stam 1993). Debido a que las unidades de mapeo genético están basadas en frecuencia de recombinación su significado en cuanto a distancia física (número de nucleótidos) entre dos *loci* varía según la especie considerada, el cromosoma y la localización dentro del mismo. Es conocido que existen regiones de alta frecuencia de recombinación localizadas sobre los brazos cromosomales (comúnmente llamados *hotspots*) y regiones donde la recombinación está fuertemente restringida como por ejemplo cerca de los centrómeros o telómeros. Aunque se corre el riesgo de cometer un error grosero puede considerarse que en promedio 1 cM equivale aproximadamente a 1 Mpb (un millón de pares de bases) en humanos y entre 500 y 750 kpb (mil pares de bases) en plantas superiores (como el arroz y tomate, respectivamente).

### 3. Mapeo de QTL

Muchas de las características de importancia agronómica presentan una distribución continua de valores, esto es, dentro de una población determinada no existe una clara distinción en clases fenotípicas. La base genética de estas características fue esclarecida poco después del redescubrimiento de las leyes de Mendel, cuando Yule en 1902 propuso que esa variación era la consecuencia de la acción aditiva de muchos genes. Entre éstos se suele identificar a genes de efectos pronunciados, denominados *genes mayores* y genes con pequeños efectos que fueron inicialmente denominados *poligenes*. Sin embargo, la distinción entre *poligenes* y *genes mayores* (también llamados *oligogenes*) no resultó satisfactoria ya que el efecto de un gene depende del entorno genético, el ambiente y del alelo presente en el *locus* del gen. Geldermann (1975) propuso denominar a estos *loci* controladores de características cuantitativas como QTLs (del inglés *Quantitative Traits Loci*). Esta nomenclatura resultó más adecuada ya que no asocia al gen con la magnitud de su efecto.

Si bien la genética cuantitativa ha propuesto y utilizados modelos biométricos con gran éxito a lo largo de la historia del mejoramiento de plantas y animales, estos modelos se basan en la estimación de los efectos génicos y no tanto en el número de genes involucrados. Utilizando marcadores moleculares ligados a los QTLs, es posible estimar el genotipo, los efectos genéticos y tener una aproximación menos sesgada del número de o genes/QTLs determinantes del carácter. Es indispensable que el lector no considere la metodología de marcadores moleculares sustituta de de los modelos genéticos-estadísticos utilizados por la genética cuantitativa, por el contrario, ellos son la base de los estudios de mapeo de genes/QTL utilizando marcadores moleculares.

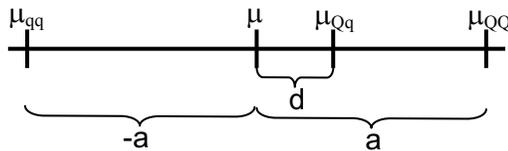
Una de las técnicas de identificación de QTLs es la denominada asociación con marcas simples o marcadores individuales. En este caso, la utilización de un mapa genético no es necesaria, aunque un mapa genético ayuda a la interpretación de los resultados. Las posibilidad de identificar un QTL usando marcadores

moleculares depende de la distancia genética existente entre el marcador y el QTL, y de la magnitud de los efectos genéticos, aditivo ( $a$ ) y de dominancia ( $d$ ) (ver más abajo), del propio QTL. La estimación de estos efectos es indirecta pues se realiza con base en el marcador ligado al QTL. Además, la estimación de estos efectos será más o menos exacta dependiendo de la técnica de mapeo utilizada, siendo la más adecuada aquellas técnicas derivadas del *mapeo por intervalo*.

Consideremos  $M/m$  el *locus* del marcador molecular y  $Q/q$  el *locus* del QTL. Para saber si  $M$  y  $Q$  están ligados, son necesarias las siguientes etapas:

- obtener una población segregante para ambos *loci*  $M$  y  $Q$  (población de mapeo),
- medir la característica fenotípica ( $Y$ ) controlada por el QTL ( $Q$ ),
- obtener el fenotipo del marcador ( $M$ ) en cada planta de esta población, y registrarlo en una matriz de datos.

Considerando el caso de un único *locus* ( $Q/q$ ), en que el alelo  $Q$  aumenta el valor de la característica, y haciendo  $\mu_{QQ}$ ,  $\mu_{Qq}$  y  $\mu_{qq}$  las medias de los individuos con genotipos  $QQ$ ,  $Qq$  y  $qq$ , respectivamente, y  $\mu = (\mu_{QQ} + \mu_{qq})/2$  la media entre los genotipos homocigotos, los siguientes efectos genéticos pueden ser obtenidos una población  $F_2$ :



$a$ ) efecto aditivo ( $a$ ) y ( $d$ ) efecto de la dominancia donde:

$$a = \frac{\mu_{QQ} - \mu_{qq}}{2} \quad d = \mu_{Qq} - \frac{\mu_{QQ} + \mu_{qq}}{2}$$

En una generación  $F_2$  cuyos genotipos marcadores  $MM$ ,  $Mm$  y  $mm$  estarán presente en la frecuencia de  $1/4$ ,  $1/2$  y  $1/4$ , respectivamente. En este caso los efectos  $a$ ,  $d$  y  $d/a$  son estimados con base en el marcador ligado al QTL, como sigue:

Efecto aditivo

$$a = \frac{\mu_{MM} - \mu_{mm}}{2(1 - 2r)}$$

Efecto de dominancia:

$$d = \frac{\mu_{Mm} - \frac{\mu_{MM} + \mu_{mm}}{2}}{(1 - 2r)^2}$$

Razón  $d/a$  (grado medio de dominancia):

$$\frac{d}{a} = \frac{2\mu_{Mm} - \mu_{MM} - \mu_{mm}}{(1 - 2r)(\mu_{MM} - \mu_{mm})}$$

Si el QTL no estuviera ligado al marcador, ( $r = 0,5$ ) las diferencias entre las tres medias no serán estadísticamente significativas. De la misma forma, si  $a$  y  $d$  fuesen iguales a cero (ausencia de efecto), no se detectarían diferencias estadísticas entre las medias. De esta manera, sólo si  $r < 0,5$  es posible detectar asociaciones entre  $M$  y  $Q$ , dependiendo esta detección del grado de ligamiento o magnitud de  $r$  (pequeños valores de  $r$  facilitan la detección) y del efecto del alelo  $Q$  sobre el carácter.

Formalmente tenemos:  $H_0: \mu_{MM} = \mu_{Mm} = \mu_{mm}$ ;  $H_{A1}: \mu_{MM} \neq \mu_{Mm}$ ;  $H_{A2}: \mu_{MM} \neq \mu_{mm}$ ;  $H_{A3}: \mu_{Mm} \neq \mu_{mm}$ .

Medias de los genotipos

Valores genotípicos

La asociación Marcador-QTL se confirma si las diferencias entre las medias fenotípicas de grupos de marcadores son estadísticamente diferentes de cero (se rechaza de  $H_0$ ). Las diferencias pueden ser evaluadas por medio del análisis de la varianza (ANOVA) o prueba  $t$ , eligiendo un nivel de significancia de  $P \leq 0,05$  como se ejemplifica en la Tabla 2.

En una población  $F_2$ , la asociación entre el marcador y el QTL puede ser valorada a través de la regresión entre los genotipos y su corres-

**Tabla 2:** ANOVA entre los tratamientos MM, Mm y mm obtenidos de una población  $F_2$

FV	GL	SC	CM	F	P
Tratamiento	2	202791,97	101395,98	54,94	0,0000
Residuo	142	262030,95	1845,28		
Total	144	464822,93			

FV: fuente de variación, GL: grados de libertad, SC: suma de cuadrados, CM: cuadrados medios, F: valor del estadístico F (CM Tratamiento/CM Residuo) y P: probabilidad de rechazar una  $H_0$  verdadera. Cuando  $P < 0,05$  indica asociación entre el marcador (M) y el QTL (Q).v

pendiente fenotipo. En este caso, es necesario codificar los tres genotipos (MM, Mm y mm) como 1, 0 y -1 para el modelo aditivo, y 0, 1 y 0 para el modelo dominante.

El modelo de regresión adoptado es:

$$Y_i = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \varepsilon_j$$

donde:

$Y_i$  = valor de la característica;  $X_{1j}$  = código del marcador para efecto aditivo (MM = 1, Mm = 0, mm = -1);  $X_{2j}$  = código del marcador para efecto dominante (MM = 1, Mm = 0, mm = -1);  $\beta_0$  = intercepto de la regresión (media de la característica);  $\beta_1$  = inclinación de la recta para el modelo aditivo;  $\beta_2$  = inclinación de la recta para el modelo dominante y  $\varepsilon_j$  = error aleatorio para  $j$ -ésimo individuo.

Una vez evaluado y confirmado el modelo completo (modelo aditivo + dominante), debe analizarse la importancia relativa de cada uno de los modelos, por ejemplo mediante la metodología de regresión múltiple de eliminación por paso (stepwise).

En poblaciones de retrocruzamiento, sólo el genotipo heterocigoto (Qq) y uno de los homocigotos (QQ o qq) están presentes por lo que el efecto aditivo y dominante quedan confundidos, en este apartado denominados  $g_1$  y  $g_2$ , respectivamente, y estimados como sigue:

$$g_1 = \mu_{MM} - \mu_{mm} = (a-d)(1-2r)$$

$$g_2 = \mu_{Mm} - \mu_{mm} = (a+d)(1-2r)$$

Observe que:

$$a = g_1 - g_2$$

$$d = g_2 + g_1$$

por lo que para estimar los efectos  $a$  y  $d$  en una población de retrocruzamiento debe retrocruzarse la  $F_1$  con cada uno de los progenitores. Un problema adicional es que  $a$  y  $d$  pueden anularse. Si existiera dominancia completa de Q sobre q

( $d = a$ ),  $g_1 = 0$  y si existiera dominancia completa de q sobre Q ( $d = -a$ ),  $g_2 = 0$ .

El análisis de ANOVA y la regresión son técnicas que permiten la identificación de marcadores ligados a QTLs, pero poseen ciertas limitaciones:

- 1) no indican si hay uno o más QTLs asociados al marcador,
- 2) no estiman la posición del QTL y
- 3) el efecto del QTL puede ser subestimado debido a que es confundido con la frecuencia de recombinación.

### 3.2. Mapeo por intervalos

El mapeo por intervalo (*interval mapping*) fue propuesto originalmente por Lander y Botstein (1989) y es la base de todos los métodos de mapeo de QTL utilizados actualmente. Para conceptualizar esta metodología podemos considerar el cruzamiento entre dos progenitores homocigotos cuya  $F_1$  posee la constitución genética que se ejemplifica en la Figura 4.

A diferencia del mapeo por marcas simples, en el mapeo por intervalo es indispensable poseer un mapa genético ya que la unidad de análisis es el intervalo genético y no un marcador individual. De manera resumida el mapeo por intervalo utiliza la siguiente estrategia: **1)** como la distancia entre cada par de marcadores es conocida, el método estima los parámetros del modelo a incrementos de distancias genéticas predefinidos, **2)** los parámetros del modelo pueden ser estimados mediante la función de verosimilitud o regresión, **3)** la hipótesis  $H_0$ : ausencia de QTL/  $H_A$ : QTL en un intervalo entre marcadores, son contrastadas a través



**Tabla 3:** Esperanza de los efectos genotípicos medios para un QTL para todos los posibles genotipos de los marcadores flanqueantes, en una población  $F_2$ , cuando no se considera la ocurrencia de recombinaciones dobles en el intervalo.

Genotipo Marcador	Probabilidad condicional			Variables indicadoras	
	QQ	Qq	qq	(a)x'	(d)z'
$M_1M_1M_2M_2$	1	0	0	1	0
$M_1M_1M_2m_2$	$1-p$	$p$	0	$1-p$	$p$
$M_1M_1m_2m_2$	$(1-p)^2$	$2p(1-p)$	$p^2$	$1-2p$	$2p(1-p)$
$M_1m_1M_2M_2$	$p$	$1-p$	0	$p$	$1-p$
$M_1m_1M_2m_2$	$cp(1-p)$	$1-2cp(1-p)$	$Cp(1-p)$	0	$1-2cp(1-p)$
$M_1m_1m_2m_2$	0	$1-p$	$p$	$-p$	$1-p$
$M_1m_1M_2M_2$	$p^2$	$2p(1-p)$	$(1-p)^2$	$-(1-2p)$	$2p(1-p)$
$m_1m_1M_2m_2$	0	$p$	$1-p$	$-(1-p)$	$p$
$m_1m_1m_2m_2$	0	0	1	-1	0

$$p=r_a/r ; (1-p)=r_b/r ; c=r^2/[r^2+(1-r)^2] \text{ (Tomado de Schuster \& Cruz, 2004).}$$

ligado contribuye a la variación fenotípica; **2)** interfiere en la estimación del efecto del QTL y **3)** removiendo el efecto de la segregación del otro QTL se reduce la varianza residual del intervalo en consideración, aumentando el poder de detección y precisión. El mapeo por intervalo compuesto (Jansen 1993, 1994; Zeng 1993, 1994), combina el mapeo por intervalo con la regresión múltiple. Esta última, permite incluir en el mapeo marcadores ubicados por fuera del intervalo como cofactores, incrementando el poder de detección y precisión en la estimativa de la posición del QTL. La pregunta de cuántos cofactores deben utilizarse no es de fácil respuesta. En principio los dos marcadores que flanqueen el intervalo deben ser incluidos, al igual que aquellos que resulten significativos en una regresión múltiple. Definidos los marcadores que se utilizarán como cofactores, la metodología de mapeo por intervalo compuesto se aplica exactamente igual a la de mapeo por intervalo simple.

### 3.4. Test de hipótesis y nivel de significancia

La formulación de hipótesis adecuadas y utilización de estadísticos apropiados, para la evaluación de las mismas, es esencial para declarar la presencia de un QTL. En una po-

blación  $F_2$  pueden ser estimados los efectos aditivo y de dominancia, en un RC el efecto  $g_1$  o  $g_2$  y sólo el efecto aditivo en poblaciones de RILs y DH. Para determinar si el valor de cada efecto estimado es diferente de cero se usan las estadísticas LR o LOD. En muchos trabajos de mapeo un valor de LOD superior a 2 – 3 es utilizado como criterio para declarar la presencia de un QTL. Esta forma de determinar el valor umbral o crítico es arbitrario y la mayoría de los autores no lo consideran el más adecuado, pues la probabilidad de detectar al menos un falso positivo (declarar la real la presencia de un QTL inexistente) en alguna región del genoma es prácticamente 1 (100 %). Se propuso disminuir este sesgo determinando un nivel de significancia para cada cromosoma. En este caso, el nivel de corte para cada cromosoma correspondería al del genoma completo dividido el número de cromosoma o grupos de ligamientos obtenidos. También es posible definir un nivel de significancia para cada intervalo, dividiendo la significancia del cromosoma por el número de intervalos en el grupo de ligamiento (número de marcadores menos uno).

Finalmente, el valor crítico de corte puede establecerse empíricamente mediante el test de permutaciones que es actualmente el más usado. Para realizar este test deben numerar-

se los individuos de 1 a  $n$ . Los valores fenotípicos de las características son tomados al azar y atribuidos a los individuos conforme vayan siendo retirados. El primer valor fenotípico retirado es asignado al individuo 1, el segundo valor fenotípico al individuo 2, así sucesivamente, se asignan todos los valores fenotípicos a cada uno de los individuos. Los datos permutados son analizados para detectar la presencia de QTL, almacenando el mayor valor de LOD obtenido. El proceso se repite de forma completa  $N$  veces, finalmente, los valores de LOD retenidos, son ordenados por orden de magnitud permitiendo estimar el valor crítico de cada test. El valor crítico de cada intervalo tendrá la magnitud  $N(1 - \alpha)$ , o sea, si se realizaron 1000 permutaciones, la significancia de 5% será el 950 ° valor ordenado. Se procede de la misma forma para obtener el valor de corte para un determinado grupo de ligamiento o para el genoma completo (conjunto de grupos de ligamientos) considerando uno u otro en el test de permutación. Cuanto mayor el número de permutaciones, mayor la precisión en el valor crítico.

#### 4. Mapas y marcadores funcionales

En la historia del desarrollo de los marcadores moleculares se ha priorizado la detección de polimorfismos en regiones no codificantes del genoma. Esto ha sido en función de asegurar la neutralidad fenotípica de los marcadores y posibilidad de maximizar la variación entre diferentes genotipos. Esta última característica está sustentada en que las regiones no génicas del genoma no han sufrido presión de selección sobre las mutaciones que pudieran haber ocurrido en el transcurso de la evolución y por consiguiente ser más variables. La asociación de los marcadores con caracteres de interés, como se ha visto anteriormente, está determinada por la distancia de recombinación (o física) del marcador con el gen responsable del fenotipo. De esta manera, los mapas genéticos basados en marcadores moleculares han posibilitado la localización de regiones cromosómicas estadísticamente asociadas a caracteres de tipo cualitativos (monogénicos) o cuantitativos (poligénicos o QTLs). Como ya se ha visto, la determinación de QTLs permite

establecer el número, posición y efecto individual de cada uno de los *loci* involucrados. Sin embargo, la identificación de QTLs en diversos cultivos no ha concluido con la identificación de los genes responsables del fenotipo, cuya existencia se presume en el intervalo comprendido entre marcadores flanqueantes de los mismos. Como consecuencia, la función biológica (o el modo de acción) de los genes responsables de los efectos de los QTLs es aún desconocida en la mayoría de los casos reportados. Asimismo, la posibilidad de realizar un "caminado cromosómico" (*chromosome walking*) desde el marcador flanqueante al QTL hacia la región física donde reside el gen responsable con el fin de aislarlo, depende de la distancia de recombinación entre el marcador y el gen, y de la relación de esa distancia con la distancia física real, que puede estar afectada por varios otros factores, como fuera mencionado antes en este capítulo.

La creciente disponibilidad pública de información de secuencias de transcritos de ADN mensajero (*Expressed Sequence Tags*, ESTs), junto con el grado de homología evidenciado entre especies, ha permitido la generación de los llamados marcadores génicos o funcionales, cuya característica relevante es la de estar localizados en regiones codificantes del genoma.

tipican en las llamadas mutaciones silenciosas. Esto es, cuando el cambio nucleotídico se produce en la tercera base del codón sin alterar el aminoácido para el cual codifica. La identificación de SNPs puede inferirse directamente a partir de datos de secuencias públicas, si estas corresponden a distintos genotipos, o por secuenciación de fragmentos de genes amplificados por PCR de individuos diversos.

Aunque históricamente se ha asumido que las secuencias tipo microsatélites son propias del ADN no codificante, hoy sabemos que las mismas pueden encontrarse también en regiones génicas. Si bien la frecuencia relativa de estos motivos es baja (< 5 %) la gran abundancia de secuencias de ESTs en diversos cultivos, ha ocasionado que varios estudios hayan usado esta estrategia para el desarrollo de marcadores moleculares. La selección de motivos de trinucleótidos, aumenta la posibilidad

de encontrar SSRs polimórficos en regiones codificantes, minimizando los posibles cambios en el marco de lectura ocasionados por una variación en el número de repeticiones de motivos de di, tetra o pentanucleótidos. Estos marcadores, denominados SSRs génicos o EST-SSRs, presentan las mismas ventajas que los SSRs neutros que han sido enunciadas en el Capítulo 5.

Los marcadores CAPs requieren la existencia de un sitio de restricción polimórfico en la secuencia génica, y si bien se han reportado casos, los mismos son de baja ocurrencia.

La aproximación más sencilla para el desarrollo de un marcador funcional es a través de una amplificación por PCR de un fragmento de un gen y estableciendo la existencia de polimorfismo entre genotipos a través de la resolución del producto en geles de SSCP. Los polimorfismos detectados mediante esta técnica se basa en que hebras simples de ADN con secuencias distintas, migrarán diferencialmente en un campo eléctrico, en función de un diferente plegamiento intra-cadena dado por la complementariedad de bases. Patrones de bandas distintos entre genotipos estarían indicando una variación de secuencia en el fragmento amplificado. Sin embargo, no puede asegurarse identidad de secuencia de dos fragmentos amplificados, por no observarse diferencias en los patrones de electroforesis, ya que la detección de las mismas es un balance entre el número de SNPs presentes y el tamaño del fragmento amplificado. Los polimorfismos presentados por los marcadores funcionales pueden entenderse como variantes alélicas, que pueden o no estar asociados a una variación a nivel proteico, ya sea en estructura o expresión. De todos modos, estos marcadores, análogamente a lo enunciado antes en este capítulo, pueden utilizarse para generar mapas genéticos, que al estar formados por marcadores funcionales se denominan mapas funcionales.

Los mapas funcionales en general están basados en mapas pre-existentes de marcadores neutros con el agregado de marcadores funcionales. Estos marcadores funcionales pueden corresponder a genes sin relación entre sí, como por ejemplo en mapas de ESTs-SSRs, o incluir análogos de genes de resistencia

(RGAs) o genes pertenecientes a rutas metabólicas particulares (por ejemplo: síntesis de carbohidratos), o a integrantes de una familia génica (por ejemplo: genes de la familia P450).

Los mapas genéticos funcionales posibilitan establecer una estrategia de genes candidatos en la identificación de factores responsables de QTLs a nivel molecular, a través de la coincidencia en la localización de QTLs y marcadores génicos. Un marcador funcional se convierte en gen candidato cuando está asociado al carácter fenotípico a través de un análisis de marcas simple, o cuando en un mapa su localización coincide con un QTL. En el primer caso, tanto el marcador como el carácter fenotípico deben estar evaluados en la misma población, mientras que en el segundo caso esto no es necesario, y es factible comparar la posición de un marcador funcional con QTLs de mapas de otras poblaciones de la misma u otra especie. Por ejemplo, se ha encontrado que el locus del gen *Lin5* del cromosoma IX del tomate correspondiente a una invertasa vacuolar (responsable de la degradación de sacarosa en glucosa y fructosa) localizada con el QTL *Brix9-2-5* de rendimiento de azúcar de fruto, mientras que en papa el gen ortólogo *invGE* está asociado a variaciones en el nivel de azúcares reductores en tubérculos de una población segregante, y su posición coincide con el QTL *Sug9a*, responsable del endulzamiento por frío en tubérculos de otra población de papa. Análogamente se ha encontrado que marcadores funcionales, basados en análogos a genes de resistencia (RGAs) se asocian en diferentes especies a resistencias a distintas enfermedades fúngicas y virales.

Asimismo, en caracteres complejos donde se hayan establecido múltiples QTLs, los marcadores funcionales permiten establecer la interacción entre éstos a la luz de las rutas metabólicas involucradas, pudiéndose identificar enzimas claves para la manifestación del carácter. Esta información genera un marco adecuado para estudios de ingeniería metabólica.

El mejoramiento genético asistido por marcadores moleculares (MAS) encuentra en los marcadores funcionales una situación ideal en la que la recombinación entre el gen responsable y el marcador es nula, ya que ambos guardan identidad.

Al estar diseñados sobre genes, la transferencia de estos marcadores a través de especies está beneficiada, dada la mayor conservación de las secuencias codificantes. Asimismo, la comparación de mapas funcionales aporta información en estudios de sintenia (conservación del tipo y orden de marcadores entre especies relacionadas), evolución y especiación entre diferentes organismos.

Del mismo modo, la utilización de marcadores funcionales en la conservación de los recursos genéticos permite ponderar la diversidad genética existente en base a las variantes alélicas intra e inter-específicas. Concordantemente, los estudios de genética de asociación (o mapeo de asociación), basados en el desequilibrio de ligamiento entre un marcador y un carácter fenotípico en una población de individuos no relacionados (como es el caso de los bancos de germoplasma) se presenta robustecido por el uso de marcadores génicos.

Finalmente, la correlación de los marcadores funcionales con datos fenotípicos o con QTLs sólo sugiere la función putativa de un gen y sus variantes alélicas. Es por esto que la confirmación definitiva de la función génica debe realizarse mediante estrategias más directas como ensayos de complementación y/o silenciamiento génico.

## 5. Bibliografía y lecturas recomendadas:

- ALLARD, R. W.** 1956. Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia* 24: 235–278.
- FEINGOLD, S. E.; LLOYD, J.; NORERO, N.; BONIERBALE, M. y LORENZEN, J.** 2005. Map location and diversity values for new potato SSRs developed from EST databases. *Theor. Appl. Genet.* 111: 456–466.
- GELDERMANN, H.** 1975. Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theor. Appl. Genet.* 46: 319–330.
- HALEY, C. S. y KNOTT, S.A.** 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: 315–324.
- JANSEN, R. C.** 1993. Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205–211.
- KEARSEY, M. J. y POONI, H.S.** 1996. Estimation of recombination frequency. En: *The genetic analysis of quantitative traits*. (First edition). Chapman & Hall. London. pp. 120–132
- LANDER, E. S.; GREEN, P.; ABRAHAMSON, J.; BARLOW, A.; DALY, M. J.; LINCOLN, S. E. y NEWBURG, L.** 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174–181.
- LANDER, E.S. y BOTSTEIN, D.** 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.
- ORTIZ, J. P. A.; PESSINO, S. C.; BATH, V.; HAYWARD, M. D. y QUARÍN, C. L.** 2001. A genetic linkage map of diploid *Paspalum notatum*. *Crop. Science* 41: 823–830.
- RUSSELL, P.J.** 1996. Mendelian Genetics. En: *Genetics*, Chapter 2 (4<sup>th</sup> Edition). Harper Collins College Publishers, New York. pp. 17–46
- SAX, K.** (1923) The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8:552–560
- SCHUSTER, I. y CRUZ, C. D.** 2004. *Estatística genômica aplicada a populações derivadas de cruzamentos controlados*. Primera Edición. Viçosa, Brasil. Ed. UFV. pp 568.
- STAM, P.** 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J.* 3:739–744.
- WEISING, K.; NYBON, H.; WOLFF, K. y KAHL, G.** 2005. Linkage analysis and genetic maps. En: *DNA Fingerprinting in plants. Principles, Methods, and Applications* (2<sup>nd</sup> Edition). CRC Press, Taylor & Francis Group, Boca Raton, Florida. pp. 277–289
- YOUNG, N.D.** 1994 Constructing a plant genetic linkage map with DNA markers. En: *DNA-based markers in plants. Advances in cellular and molecular biology of plants*. Vol. 1. (PHILLIPS, R. L. y VASIL, I. K. Editors). (1<sup>st</sup> edition). Kluwer Academic Publishers. Dordrecht, The Netherlands. pp. 39–57
- YULE, G.U.** 1902. Mendel's laws and their probable relation to intra-racial heredity. *New Phytol.* 1: 193–207.

# I - CAPÍTULO 7

## Genómica

Viviana Echenique; Juan P. Selva; Mauro Meier; Pablo Roncallo; Gustavo Schrauf

### 1 Introducción

La «**genómica**» se desarrolló en las últimas dos décadas como consecuencia de los avances realizados en biología molecular e Informática, dos áreas de la ciencia que han tenido un desarrollo tecnológico enorme, generando una revolución en el conocimiento y en la comprensión de los procesos biológicos.

El término fue acuñado en 1986 por Thomas Roderick para referirse a la **subdisciplina de la genética que se ocupa del mapeo, secuenciación y análisis de las funciones de genomas completos** y sirvió de nombre para una revista especializada en la publicación de los mencionados temas, “**Genomics**”. Consiste en la caracterización molecular de genomas enteros y aporta información acerca de la secuencia y de la función de cada sector del genoma en diferentes situaciones de desarrollo y bajo diferentes condiciones ambientales, así como de los mecanismos implicados en la regulación de la expresión e interacción génica.

Para ello, las herramientas que se utilizan para el análisis individual de genes o pequeñas regiones cromosómicas, se aplican al análisis global de genomas completos, estudiando en conjunto los miles de genes, proteínas y metabolitos que constituyen un organismo, así como las complicadas redes de interacciones que operan entre ellos. La información generada es enorme y es clave para la identificación y el aislamiento de genes de interés y permitirá interpretar, en términos moleculares, los procesos biológicos. Para ayudar en este proceso han surgido poderosas **herramientas bioinformáticas** que permiten almacenar e interpretar esta información.

Las aplicaciones de la genómica alcanzan a todos los ámbitos de la actividad humana relacionados con la biología, como la salud, la alimentación y el medio ambiente. Actualmente

se encuentran disponibles las secuencias de nucleótidos y aminoácidos de miles de genes y productos génicos codificados por los genomas de varios organismos complejos. Esta información se utiliza para el estudio de enfermedades humanas complejas y para comprender fenómenos tales como la fisiología celular, el desarrollo, la conducta, la evolución, etc. También aportan información acerca de todos los aspectos del crecimiento y desarrollo vegetal, diferenciación y respuesta a estreses bióticos y abióticos. Gracias al potencial de relacionar fenotipos biológica y económicamente importantes con los genes responsables de los mismos actualmente es posible identificar genes de interés para ser transferidos a otros organismos por medio de tecnología génica. Por otro lado, el conocimiento de la secuencia de los genes posibilita el desarrollo de marcadores perfectos, basados en la secuencia del mismo gen, lo cual facilita la selección y la transferencia de los mismos a variedades de interés. La conjunción entre tecnología génica, marcadores moleculares, genómica y bioinformática ha revolucionado el mejoramiento genético vegetal, dando origen a lo que se denomina **mejoramiento molecular** y se están desarrollando nuevos y promisorios cultivares (**Fig. 1**).

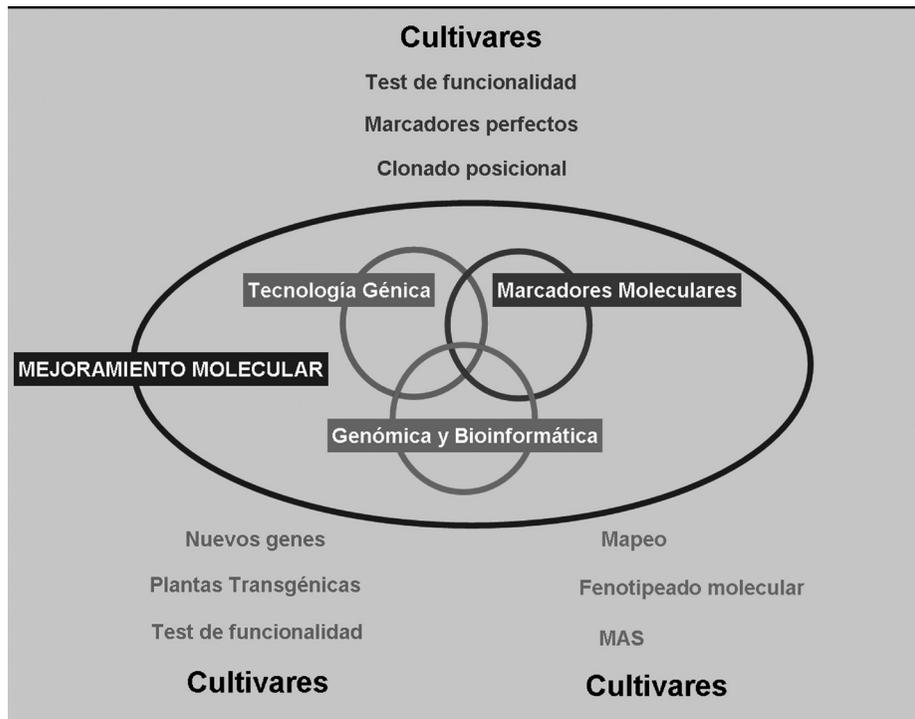
La genómica puede dividirse en:

- **Genómica estructural**, que se ocupa de la caracterización física de los genomas.

- **Genómica funcional**, que caracteriza el **transcriptoma**, que está constituido por el conjunto completo de transcritos producidos por un organismo, el **proteoma** o conjunto de proteínas codificadas por un genoma, y el **metaboloma** o conjunto total de metabolitos de una célula, consecuencia de la función de los ARN y proteínas.

El objetivo de la genómica es la dilucidación completa y exacta de la secuencia de ADN de un genoma haploide representativo de una especie. Cuando esta secuencia se conoce, abre la puerta a numerosas posibilidades.

Por análisis computacional de la misma y utilizando principios conocidos de genética es posible:



**Figura 1:** La conjunción entre tecnología génica, marcadores moleculares, genómica y bioinformática revolucionaron el mejoramiento vegetal, conduciendo al desarrollo de nuevos cultivos. Los nuevos genes hallados son introducidos en plantas por tecnología génica. Esto a su vez permite establecer la funcionalidad de los mismos. El conocimiento de la secuencia permite la obtención de mapas, marcadores perfectos y realizar selección asistida por marcadores moleculares (MAS), así como el fenotipado molecular. Gentileza de G. Spangenberg.

- **Comparar secuencias** similares presentes en diferentes entidades biológicas y comprender la función de las mismas.
- **Realizar predicciones** acerca de todas las proteínas codificadas por una especie.
- **Establecer las variaciones genéticas** entre distintas poblaciones de una misma especie.

**Comparar secuencias de diferentes especies y entender procesos evolutivos.**

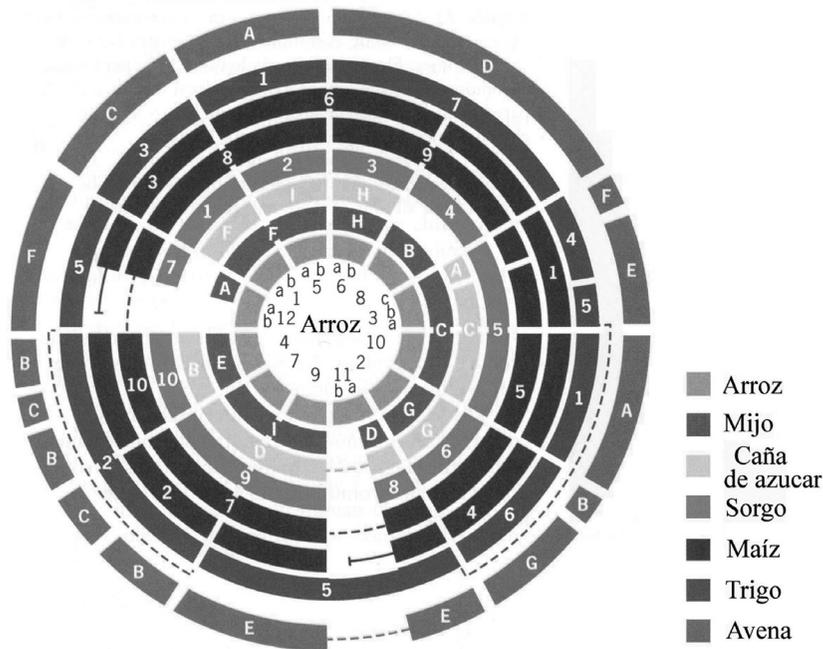
Esto ha dado origen a la **genómica comparativa** y ha demostrado que existe considerable sintenia, es decir, una localización conservada de genes en posiciones equivalentes en especies relacionadas (**Fig. 2**). También ha sido una excelente herramienta para identificar motivos de secuencia altamente conservados y, por lo tanto, funcional-

mente importantes en regiones codificantes y no codificantes del genoma. Catalogar las variaciones genéticas entre individuos ayudará a identificar aquellos cambios que conducen a enfermedades genéticas, investigar nuevas terapias y establecer la resistencia o susceptibilidad a diferentes factores. Para la comparación de secuencias se utiliza el algoritmo **BLAST** (**b**asic **l**ocal **a**lignment **s**earch **t**ool: herramienta de búsqueda de alineación local básica), para lo cual existen distintos programas que permiten hallar regiones similares entre diferentes genes (Parte I, Cap. 12).

En Febrero de 2001, coincidiendo con el cincuenta aniversario de la publicación del modelo estructural del DNA por Watson y Crick, la revista *Nature* publicó el primer borrador del **genoma humano**, que ha sido secuenciado completamente. También se han secuenciado otros genomas como maíz, colza, soja, trébol, raigrás, etc. (Tabla 1).

**Tabla 1.** Ejemplos de genomas secuenciados en los últimos años.

	<b>Especie</b>	<b>Tamaño del genoma</b>
<b>Bacterias</b>	<i>Haemophilus influenzae</i>	1.830.137 pb
	<i>Mycoplasma genitalium</i>	580.070 pb
	<i>Methanococcus jannaschii</i>	1.660.000 pb
	<i>Helicobacter pylori</i>	1.667.867 pb
	<i>Escherichia coli</i>	4.639.221 pb
	<i>Bacillus subtilis</i>	4.214.810 pb
	<i>Treponema pallidum</i>	1.138.006 pb
	Levadura ( <i>Saccharomyces cerevisiae</i> )	12.052.000 pb
<b>Organismos Eucariotas</b>	<i>Trichomonas vaginalis</i>	160 Mpb
	Nematodo ( <i>Caenorhabditis elegans</i> )	97 Mpb
	Mosca del vinagre ( <i>Drosophila melanogaster</i> )	120 Mpb
	Mosquito de la malaria ( <i>Anopheles gambiae</i> )	278 Mpb
	Mosquito del dengue y fiebre amarilla ( <i>Aedes aegypti</i> )	1.380 Mpb
	Abeja ( <i>Apis mellifera</i> )	236 Mpb
<b>Plantas</b>	Sorgo ( <i>Sorghum bicolor</i> )	700 Mbp
	Aroz ( <i>Oryza sativa</i> ssp. <i>japonica</i> e <i>indica</i> )	420 - 466 Mpb
<b>Vertebrados</b>	Pez globo ( <i>Tetraodon nigroviridis</i> )	300 Mpb
	Gallo rojo de la jungla ( <i>Gallus gallus</i> )	1.000 Mpb
	Zarigüeya (Marsupial, <i>Monodelphis domestica</i> )	3.475 Mpb
	Ratón ( <i>Mus musculus</i> )	2.500 Mpb
	Rata ( <i>Rattus norvegicus</i> )	2.750 Mpb
	Perro ( <i>Canis familiaris</i> )	2.411 Mpb
	Chimpancé ( <i>Pan troglodytes</i> )	2.843 Mpb
	Macaco rhesus ( <i>Macaca mulatta</i> )	2.870 Mpb
	Humano ( <i>Homo sapiens</i> )	3.000 Mpb



**Figura 2:** Sintenia entre los genomas de varias gramíneas. El genoma de arroz contiene grupos de marcadores que se hallan altamente conservados en muchos cereales, por lo cual dichos genomas pueden sintetizarse básicamente como compuestos por «bloques de ligamiento de arroz». En la figura se muestra el alineamiento de los cromosomas y segmentos de cromosomas (identificados con letras mayúsculas) de varias gramíneas con los del arroz (centro). El genoma del maíz tiene dos copias semejantes de cada bloque de genes por lo que está representado por dos anillos del círculo. Las líneas externas punteadas conectan sectores adyacentes de los cromosomas de trigo. Modificado de Gale y Devos (1998).

También es posible estudiar todos los genomas bacterianos asociados a genomas más complejos, sin necesidad de aislar y de cultivar especies microbianas particulares. Esto se denomina **metagenómica** (Parte I, Cap. 11). Para este estudio se extrae todo el ADN de una comunidad microbiana de un ambiente particular (suelo, agua de mar, intestino humano, etc.), se determina que genes se encuentran presentes, cuales son sus funciones y las diferencias que existen entre distintas comunidades a este nivel.

Más recientemente ha surgido la denominada **“Genómica Sintética”**, que pretende crear formas sintéticas de vida recreando genomas artificiales, células con vías metabólicas predeterminadas y propiedades catalíticas específicas. Involucra el rediseño y la fabricación de sistemas biológicos. Como ejemplo podemos citar el trasplante de un genoma de una célula bacteriana a otra.

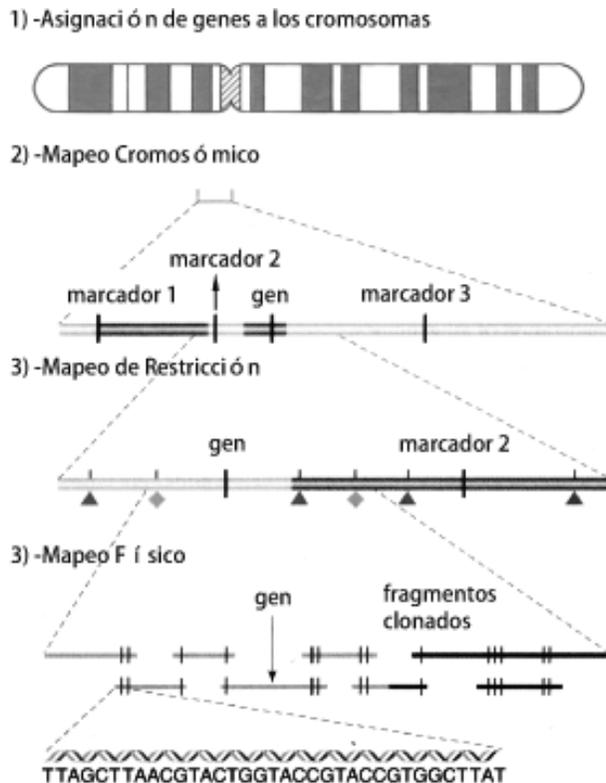
A continuación se brinda un panorama global acerca de la forma en que se articulan los proyectos de genómica, algunos resultados relevantes, las aplicaciones en agricultura y el estado de la investigación en Sudamérica.

## 2 Genómica estructural

Como su nombre lo indica, el objetivo de la misma es caracterizar la **estructura del genoma**. Conocer la estructura de un genoma individual puede ser de utilidad para manipular genes y segmentos de ADN en una especie particular. El análisis comparativo de diferentes genomas posibilitará deducir reglas generales que gobiernan la estructura de todos los genomas.

La genómica estructural procede a través de niveles incrementales de resolución analítica, comenzando con la asignación de genes y marcadores a cromosomas individuales, mapeando estos genes y marcadores dentro de los cromosomas, finalizando con la preparación de un mapa físico a través de la secuenciación. Es decir, que se procede desde un mapeo genético de baja resolución, basado en el análisis de recombinación meiótica, hacia uno de alta resolución, basado en marcadores moleculares, llegando a la realización de un mapa físico, basado en la secuencia de fragmentos clonados parcialmente solapantes (**Fig. 3**).

Los pasos a seguir son:



**Figura 3:** Niveles de complejidad de la genómica estructural. La genómica estructural procede desde un mapeo genético de baja resolución, basado en el análisis de recombinación meiótica, a uno de alta resolución, basado en marcadores moleculares hasta la realización de un mapa físico, basado en la secuencia de fragmentos clonados solapantes. Modificada de Griffiths *et al.* (2004).

- **Mapeo cromosómico de baja resolución** (ubicación de genes que producen fenotipos mutantes conocidos y algunos marcadores). Las distancias genéticas se basan en frecuencias de entrecruzamientos («crossing overs») y son medidas como porcentaje de recombinación en centimorgans (cM).

- **Mapeo genético de alta resolución** de cada cromosoma (con marcadores moleculares ubicados muy próximos entre sí).

- **Mapeo físico**, basado en distancias moleculares, resultantes de mapas de fragmentos de restricción parcialmente superpuestos. La distancia entre dos marcadores puede ser medida determinando el tamaño de los fragmentos de restricción que los contienen, siendo el fragmento más pequeño que lleva ambos marcadores una estimación de la distancia entre ellos. Utilizando combinaciones de sondas y enzimas de restricción puede construirse un mapa físico determinando qué fragmentos poseen marcadores en común. Las distancias físicas se miden en kilobases (kb) o megabases (Mb).

- **Anclado del mapa genético en el mapa físico.**

- **Secuenciación de ADN a gran escala**, que brinda un mapa completo de cada cromosoma.

- **Anclado del mapa genético y del mapa físico en el de secuencia.** Las distancias físicas generalmente no se correlacionan directamente con las distancias genéticas porque las frecuencias de recombinación no son siempre proporcionales a las distancias moleculares. Sin embargo, a menudo las dos correlacionan razonablemente bien en regiones eucromáticas de los cromosomas (aquellas menos condensadas, que se colorean de manera más difusa). En humanos, un cM es equivalente, en promedio, a aproximadamente 1 Mb de ADN.

## 2.1 Asignación de loci a cromosomas específicos.

**Se utilizan los siguientes métodos:**

- **Ligamiento a loci conocidos:** análisis tradicional basado en cruzamientos, en especies en las que es factible hacerlo.

- **Electroforesis de campo pulsátil:** se usa en el caso de especies que poseen cromosomas pequeños, separables por esta técnica, como sucede con los de levaduras. Las bandas en el gel corresponderían a cromosomas individuales y pueden utilizarse **para localizar nuevos genes por hibridación**. Primero, utilizando sondas de localización conocida

se establece que banda corresponde a cada cromosoma. Luego, un nuevo gen clonado de ubicación desconocida se utiliza como sonda para asignarle una posición en un cromosoma determinado.

- **Híbridos celulares interespecíficos**, por ejemplo humano-ratón. Se utiliza exclusivamente en mapeo del genoma humano o de especies animales. Se establecen bancos de líneas celulares que contengan todos los cromosomas del roedor y un cromosoma humano particular. Siempre incluye un cromosoma humano que lleva un alelo salvaje defectivo para una función bioquímica determinada en el genoma del ratón. Si sobreviven en un cultivo deficiente para el factor que no sintetizan es porque llevan el alelo humano equivalente, que complementa la función faltante.

## 2.2 Ubicación de los genes a lo largo del cromosoma

El próximo nivel de resolución consiste en determinar la posición de un gen o marcador molecular sobre el cromosoma. Este paso es importante porque los mapas genéticos pueden alinearse con los mapas físicos y utilizarse para validar los mismos.

- **Mapeo por recombinación:** en los casos en que ha sido factible hacerlo (organismos experimentales como levaduras, mosca de la fruta, ratón, Arabidopsis, etc.) ha posibilitado la construcción de mapas que a lo largo de los años aparecen repletos de genes con efectos fenotípicos determinados. Sin embargo, los intervalos recombinacionales entre genes conocidos contienen cantidades muy grandes de ADN. Estos intervalos o «gaps» no pueden ser completados utilizando análisis de ligamiento debido a la ausencia de marcadores en esas regiones. Para llenarlos y construir mapas de alta resolución surgieron los marcadores moleculares, entre los que se encuentran los RFLP y los SSLP (mini y microsatélites) (Parte I, Cap. 5). Estos marcadores permiten la construcción de mapas genéticos con una densidad de un marcador/centimorgan (cM). Aunque tal resolución es un logro muy importante, un centimorgan representa, en humanos, una megabase, lo que equivale a un millón de pares de bases o 1000 kb. Para lograr mayor resolución actualmen-

te existen los SNP, que son marcadores basados en el polimorfismo de un solo nucleótido.

- **Hibridación *in situ***: cuando se dispone de un gen clonado éste puede ser utilizado como para hibridar con los cromosomas *in situ*, para lo cual se lo marca radiactivamente o por fluorescencia. Puede, de ese modo, identificarse a los cromosomas portadores de la secuencia, determinar si son secuencias específicas de un cromosoma y determinar la posición del gen en el mismo. En el caso de fluorescencia, la técnica se denomina **FISH** («fluorescence *in situ* hybridization»). La localización del fragmento en el cromosoma se visualiza como un punto brillante (Parte I, Cap. 3).

- **Mapeo físico de los cromosomas**: aporta un nivel de resolución mayor. Se trata de identificar un conjunto de fragmentos de ADN clonados superpuestos que, juntos representen un cromosoma o un genoma completo. Los mapas así obtenidos se llaman **mapas físicos**, porque el ADN es el material físico del genoma.

### 2.3 Secuenciación a gran escala del genoma

#### - **Generación de bancos de EST (etiquetas de secuencias expresadas)**

La complejidad de los genomas eucariotas hace aconsejable, como primera aproximación, no abordar el estudio del genoma completo. Es preferible estudiar sólo una fracción del mismo, es decir, aquellos genes que se están expresando en un momento determinado de la vida del organismo. Para ello se obtienen poblaciones de ADNc (que reflejan sólo secuencias expresadas) que se secuencian de forma masiva para generar miles de secuencias parciales conocidas como ESTs (Parte I, Cap. 8). Estas secuencias de entre 300-500 pb. suelen ser suficientes para la identificación de los genes mediante comparación con las secuencias existentes en las bases de datos públicas (ej. Genbank, EMBL), utilizando para ello programas de análisis informático (Parte I, Cap.12). Si la información contenida en la secuencia parcial no es suficiente, habría que determinar la estructura del ADNc completo para estudiar su función por otros métodos.

#### - **Secuenciación genómica**

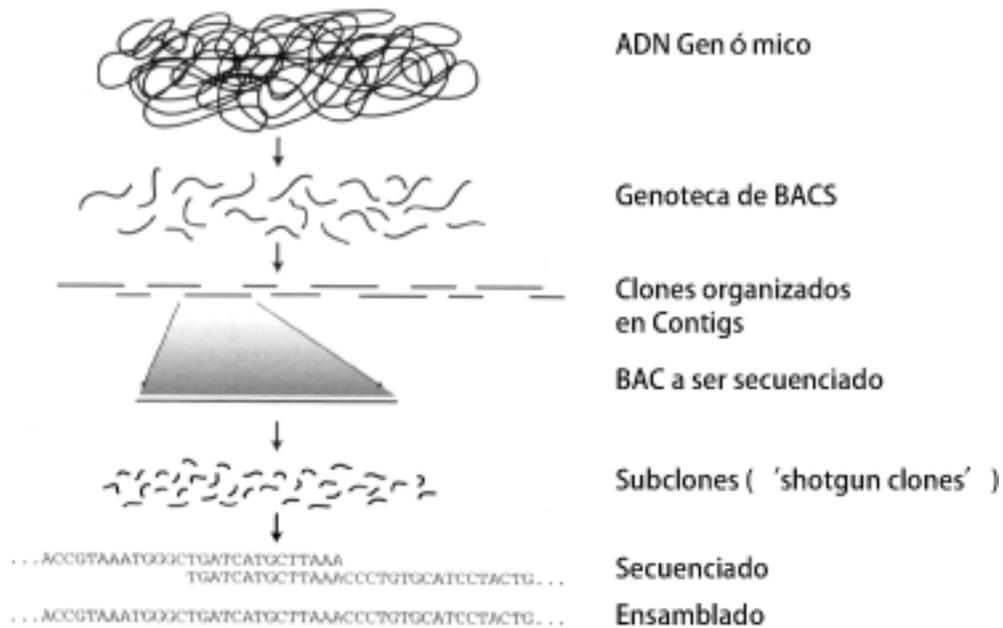
La aproximación anterior no permite identificar genes que se expresan a bajo nivel o en situaciones fisiológicas no consideradas o regiones no

representadas en el ARNm. Para obtener esta información debe secuenciarse el genoma completo. Para ello, el ADN genómico total se digiere con enzimas de restricción apropiadas o se fragmenta mecánicamente en fragmentos de gran tamaño (100-300 kb.), que se clonan en vectores apropiados, para construir genotecas que contienen, al menos, una representación completa del genoma, que puede estudiarse en detalle y secuenciarse (Parte I, Cap. 4). A fin de distinguir las regiones codificantes de las no codificantes, las secuencias se comparan con las de ESTs y ADNc previamente estudiados.

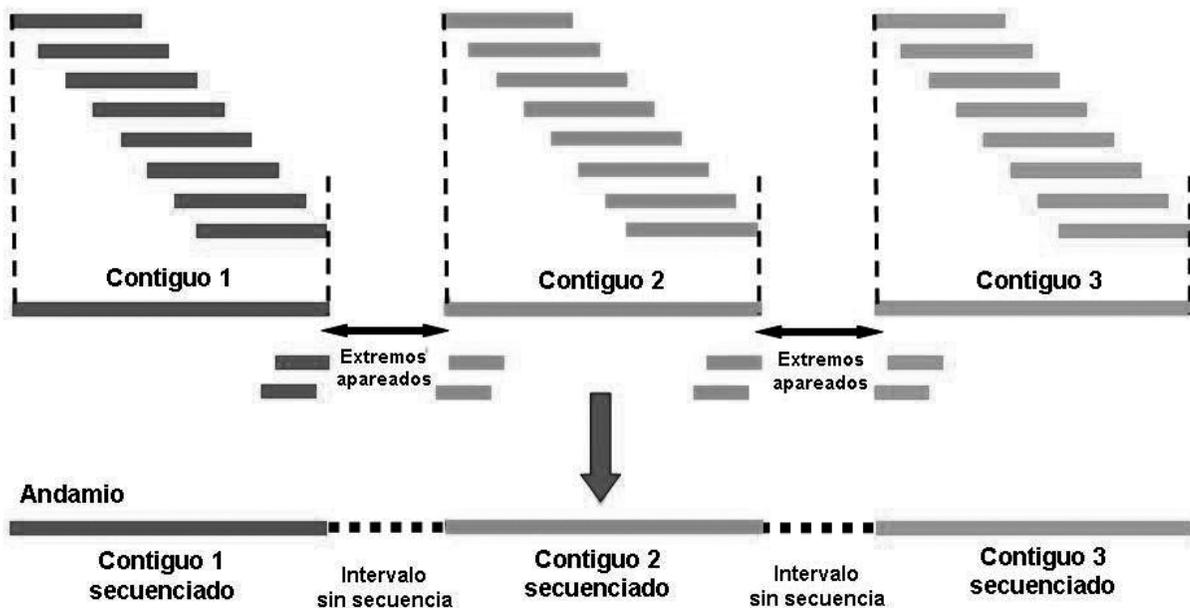
Para reconstruir la secuencia del genoma original a partir de los fragmentos clonados se utilizan las siguientes estrategias:

#### a) **Secuenciación de los clones y ensamblado de los fragmentos**

El clonado comienza obteniendo un número elevado de fragmentos al azar que se clonan en un vector apropiado. En la preparación de mapas físicos de genomas se utilizan vectores como los cósmidos, YACs, BACs y PACs, que aceptan insertos de gran tamaño (Parte I, Cap 4). El contenido de estos clones se caracteriza y se ordenan, buscando los puntos de solapamiento o superposición. Un conjunto de clones solapantes se llama contiguo. Para establecer el solapamiento de clones se secuencian regiones cortas del inserto proximas al sitio de clonado utilizando *primers* posicionados en el vector. Los intervalos sin secuencias (*gaps*) se completan utilizando el final de un fragmento clonado para llegar al siguiente (*primer walking*). A medida que se van caracterizando los clones, los contiguos se alargan y van convergiendo unos con otros. El proyecto termina cuando se tiene un conjunto de contiguos que equivale al número de cromosomas de la especie en estudio. En la Figura 4 se resume una estrategia de secuenciación jerárquica (*hierarchical shotgun sequencing*). Esta se usa para genomas grandes. Para genomas más pequeños se utiliza otra, llamada secuenciación de genomas completos (*whole genome shotgun sequencing*) (**Fig. 5**). Todos los estadios del análisis genómico como la preparación de clones, el aislamiento de ADN, la electroforesis y la secuenciación han sido bien adaptados a diferentes máquinas o robots, automatizando el trabajo.



**Figura 4:** Estrategia de secuenciado jerárquico («hierarchical shotgun sequencing strategy»). El primer paso consiste en construir una genoteca por fragmentación del ADN e introducción del mismo en vectores que aceptan grandes insertos, en este caso de BACs. Los fragmentos de ADN representados en la misma son organizados en un mapa físico y los clones individuales son seleccionados y secuenciados para lo cual se hace una nueva genoteca de trozos más pequeños («shotgun library»). La secuencia de los clones se ensambla finalmente para reconstruir la secuencia del genoma. Modificada de la publicación del International Human Genome Sequencing Consortium, 2001



**Figura 5:** Montaje de un genoma complejo mediante la secuenciación completa del genoma por tiros de escopeta («whole genome shotgun sequencing»). En primer lugar, se construyen los contiguos con las secuencias que se superponen. Finalmente se utilizan los extremos apareados para cubrir los intervalos sin secuencia y así orientar y ordenar los contiguos en unidades llamadas andamios (*scaffolds*) Modificado de Griffiths *et al.* (2004).

### b) Ordenamiento de los clones

Se utilizan varias técnicas para ordenar los clones en contiguos. Entre ellas:

- **FISH:** para localizar las posiciones aproximadas de insertos grandes.

- **Fenotipificación (fingerprinting):** se corta el clon con enzimas de restricción que generan un grupo de bandas que representan un «fingerprint» o «huella digital» del clon. Las bandas generadas por varios clones pueden alinearse visualmente o por un programa de computación para determinar si se superponen o solapan.

- **STS («sequence tag sites» o sitios de secuencia marcada):** son secuencias cortas de insertos grandes clonados. Se reúne un conjunto de clones al azar y se cortan en fragmentos más pequeños que se clonan en  $\lambda$  y se secuencian pequeñas regiones de cada uno. Para ello se diseñan *primers* que amplifican una secuencia corta de ADN (STS).

Actualmente existe una tecnología más moderna para el secuenciado de genomas, la pirosecuenciación. Esta técnica utiliza esferas atrapadas en una placa que contiene 1.600.000 microceldas. Cada esfera (una por microcelda) permite realizar una reacción de secuenciación individual, basada en la síntesis complementaria de un ADN de cadena sencilla, previamente multiplicado por PCR y alineado a cada esfera. La incorporación de cada nucleótido durante la síntesis de ADN libera una molécula de pirofosfato (PPi), que al interactuar con enzimas presentes en el medio (luciferasa, entre otras) produce una señal quimioluminiscente captada por una cámara de detección de fotones, la cual traducida en una computadora como la adición de un nucleótido determinado, generando una secuencia individual por cada celda. Esta tecnología, desarrollada por la empresa 454 Life Sciences, dio lugar a la creación del primer equipo de secuenciación masiva (Secuenciador GS 20™) diseñado para secuenciar genomas bacterianos, con una capacidad para descifrar 20 millones de bases por corrida (4hrs) en secuencias individuales de  $\pm 100$  bases. En el proyecto del género *Bacillus*, que tiene un genoma de aproximadamente 4 Mb, con seis corridas se descifraron 129.6 Mb,

con una cobertura mayor a 20x de su genoma. En comparación, por el método de Sanger, se logró una cobertura de 3x (12,5Mb).

### 3 Utilización de mapas genómicos para el análisis genético

Los mapas genéticos y los físicos son un importante punto de partida para varios tipos de análisis genético, incluyendo el aislamiento de genes y genómica funcional.

Por ejemplo:

- **Aislamiento de genes por clonado posicional:** para ubicar un gen cuya secuencia se desconoce se puede partir de un marcador conocido que esté estrechamente ligado al mismo. El mismo actúa como punto de partida para el caminado cromosómico (*chromosome walking*), donde los fragmentos finales del marcador ligado son utilizados como sondas para seleccionar otros clones de la genoteca. Del segundo grupo de clones (los que solapan con el inicial) se hacen mapas de restricción y los fragmentos obtenidos son utilizados para realizar una nueva ronda de selección de clones superpuestos. Así el proceso de «caminado» se mueve hacia ambos lados a partir del sitio inicial, que culmina cuando se llega al gen de interés. Existe otra técnica relacionada llamada salto cromosómico, que permite saltar a través de áreas distantes y potencialmente no clonables del ADN y genera «marcas», ampliamente espaciadas a lo largo de la secuencia, que pueden utilizarse como puntos de inicio para múltiples caminatas cromosómicas bidireccionales. Esta técnica consiste en crear fragmentos grandes (entre 80-150 kb) por restricción del ADN en la región que se cree contiene al gen de interés. Cada fragmento de ADN es luego circularizado, poniendo en contacto los extremos libres. Este procedimiento acerca puntos relativamente distantes de la región de genoma en estudio. Se trata de generar en esa unión un sitio que no sea cortado en una posterior restricción, de manera que esas dos regiones que estaban distantes en el genoma ahora estén unidas en un fragmento. El círculo se corta con una nueva enzima de restricción y los fragmentos obtenidos, entre los que se encuentran los sitios de unión, se clonan en fagos.

Esto es lo que se llama una «**genoteca de saltos**». Una sonda del sector de comienzo de la región que contiene al gen de interés puede utilizarse como punta de partida para comenzar a «saltar» por el genoma. Un salto puede, por ejemplo, avanzar unos 50 kb hacia el locus de interés. Una vez allí, el otro extremo de la unión del fragmento inicial se corta y se usa para buscar el siguiente punto de salto en la genoteca. O sea que a través de las uniones se va avanzando hacia el punto donde se encuentra el locus. Cada posición de salto es un punto de partida para una caminata cromosómica. Estas regiones se van secuenciando. Allí comienza la búsqueda de genes utilizando programas de predicción y se seleccionan las secuencias candidatas. Esta estrategia de saltos se utilizó para clonar el gen de la fibrosis quística, que es una enfermedad humana que resulta fatal y es causada por mutaciones en un gen de gran tamaño localizado en el cromosoma 7.

- **Estrategia del gen candidato:** la caracterización de una región cromosómica hace que aparezcan varios genes de función desconocida. Si un gen de interés es mapeado en esa región, estas secuencias pasan a ser «candidatas». Para cada una de ellas, o para la más probable, de acuerdo al análisis de secuencia, se diseñan experimentos tendientes a determinar en cuales tejidos se expresa, en qué condiciones, etc., utilizando Northern blot o RT-PCR. Si el patrón de expresión encontrado coincide con el patrón esperado es altamente probable que hayamos encontrado el gen de interés. El experimento ideal para confirmar esto, sería la transformación de un individuo mutante para esta función con el gen normal. Si se produce la reversión del fenotipo mutante (**complementación**), tendremos la prueba irrefutable de que estamos frente al gen buscado.

- **Genes con patrones complejos de herencia:** no todos los genes muestran patrones simples de herencia. Puede suceder que:

- La **variación fenotípica sea cuantitativa**, como peso o altura. Este tipo de variación se debe a la interacción acumulativa entre alelos + y - de varios genes y el ambiente.

La disponibilidad de miles de marcadores moleculares como los SSLP, dispuestos a lo largo del cromosoma, ha posibilitado el mapeo de al-

gunos de estos genes que contribuyen a la variación cuantitativa cuyos loci son llamados **QTL** (*quantitative trait loci*) (Parte I, Cap. 5 y 6). Para abordar este problema, se buscan dos tipos de líneas que muestren fenotipos contrastantes para un carácter cuantitativo y se cruzan para generar descendencia homocigota que contenga solo un segmento o un pequeño número de segmentos de una de las líneas. Estos individuos pueden caracterizarse por su fenotipo y puede estimarse la contribución de los segmentos específicos a la variación observada. Los SNP (polimorfismos de un nucleótido simple) aceleran el mapeo de caracteres complejos.

#### 4 Análisis funcional de los genes

Una vez secuenciado el genoma completo pueden identificarse la mayoría de los genes de una especie. Restaría entonces determinar cuál es la **función** de cada uno de ellos y cómo interaccionan para definir un fenotipo determinado. La **genómica funcional** intenta resolver esta cuestión a través del estudio de:

- **El transcriptoma:** se refiere al estudio de los **perfiles de expresión de todos los genes presentes en el genoma**. Para ello se extraen todos los ARNm contenidos en una célula, tejido u órgano en un determinado momento de desarrollo o situación fisiológica. El método más utilizado es el de micromatrices de ADN, que permite analizar simultáneamente la expresión de miles de genes, pudiéndose disponer entre 5000-50.000 genes (ESTs, ADNc, oligonucleótidos) sobre un portaobjetos utilizando un sistema robotizado. Sobre este *chip* de ADN se realizan experimentos de hibridación con muestras marcadas radiactivamente o con fluoróforos apropiados, y los resultados se cuantifican mediante análisis con *phosphorimager* o microscopía confocal. De esta manera se pueden identificar, de forma simultánea, los patrones de expresión de miles de genes en un momento del desarrollo o en respuesta a diferentes estímulos ambientales. El análisis bioinformático de estos datos permite asociar grupos de genes que se expresan de forma coordinada y proporciona información importante sobre la función de los mismos (Parte I, Cap. 8).

- **El proteoma:** comprende el **conjunto total de proteínas expresadas por un genoma**

**completo**, incluyendo las modificadas después de la traducción. El estudio del transcrito debe ser complementado con el análisis de las proteínas codificadas por los transcritos, puesto que estas macromoléculas están al final de la ruta de expresión génica. El método más utilizado para estudiar la abundancia relativa de cientos de proteínas es la electroforesis bidimensional en geles de poliacrilamida (2DPAGE), que permite separar con gran resolución la mayoría de los polipéptidos celulares combinando de forma secuencial diferencias en carga y en masa molecular. Utilizando sistemas computarizados de análisis de imagen se seleccionan las proteínas cuya abundancia relativa cambia durante el desarrollo o en respuesta a diferentes estímulos ambientales (Parte I, Cap.9).

- **El metaboloma:** comprende el **conjunto total de metabolitos de una célula**. En plantas, el metabolismo secundario (reacciones que no son vitales para el individuo y que conducen a la producción de compuestos que a veces ayudan a su supervivencia, como por ejemplo antraquinonas, alcaloides, digoxina, taumatina, vainillina, menta, etc.) produce una enorme variedad de compuestos diferentes. Se han identificado más de 40.000 y se estima que aún quedan por descubrir alrededor de 100.000. Los investigadores que trabajan en este nuevo campo de la química aplicada a la biología (metabolómica) consideran que sólo de esta forma se podría definir, en términos moleculares, un fenotipo concreto. En metabolómica se emplean herramientas analíticas muy sensibles (tales como espectrometría de masa, cromatografía líquida o gaseosa y resonancia magnética nuclear) para analizar los cambios metabólicos provocados por mutaciones génicas o por la expresión de transgenes. La metabolómica puede ser aplicada al monitoreo de estreses inducidos, para identificar pasos metabólicos limitantes, para el análisis de mutantes y hasta para realizar la evaluación de manejos agronómicos, tales como el efecto de fertilizantes sobre el metabolismo, etc. (Parte I, Cap. 10).

- **La bioinformática** intenta dar sentido a la información derivada de las técnicas anteriormente descritas. La importancia de esta cien-

cia resulta obvia cuando se considera que los genomas poseen miles de millones de pares de bases (Parte I, cap. 12).

## 5 Modelos

El mapeo comparativo ha demostrado que **la organización de los genes dentro de los genomas ha permanecido muy conservada** a través de la evolución, existiendo estrechas relaciones de **colinearidad** entre los genomas de casi todas las gramíneas cultivadas, entre las solanáceas, entre las brassicáceas cultivadas y *Arabidopsis*, entre los pinos, rosáceas y varias leguminosas. Por ello, teniendo en cuenta la envergadura de un proyecto de secuenciado, los emprendimientos genómicos tomaron especies modelo, representativas de un genoma vegetal. El primer modelo vegetal fue una dicotiledónea, *Arabidopsis thaliana*. Le siguió una monocotiledónea, el arroz, cuyo genoma es seis veces menor que el del maíz y 37 veces menor que el del trigo. El estudio de estas dos especies permitirá arribar a conocimientos clave para el mejoramiento vegetal.

*Arabidopsis thaliana* es una dicotiledónea que **posee uno de los genomas vegetales más pequeños**. Carece de importancia económica pero resulta ser un excelente organismo para la investigación puesto que es fácil de transformar y su ciclo de vida tarda sólo 7 semanas, de semilla a semilla. Alrededor de 12.000 científicos trabajan coordinadamente en esta pequeña planta, conformando el más avanzado sistema de experimentos en biología vegetal del mundo. Su secuencia genética es de libre acceso en el sitio <http://www.arabidopsis.org>. En un artículo publicado en Nature (2000) se analiza el genoma de *Arabidopsis* a partir de las 115.4 megabases secuenciadas (de un total de 125). Su evolución involucró una duplicación completa del genoma, seguida por una subsecuente pérdida y duplicación de genes, lo cual dio origen a un genoma dinámico, enriquecido por una transferencia lateral de genes de cianobacterias. Contiene 25.498 genes que codifican para 11.000 familias de proteínas, con una diversidad funcional similar a la encontrada en *Drosophila* y *Caenorhabditis elegans*. *Arabidopsis* posee varias familias de proteínas nuevas pero carece de otras co-

munes, indicando que los grupos proteicos en común han experimentado una expansión y contracción en estos tres grupos eucariotas.

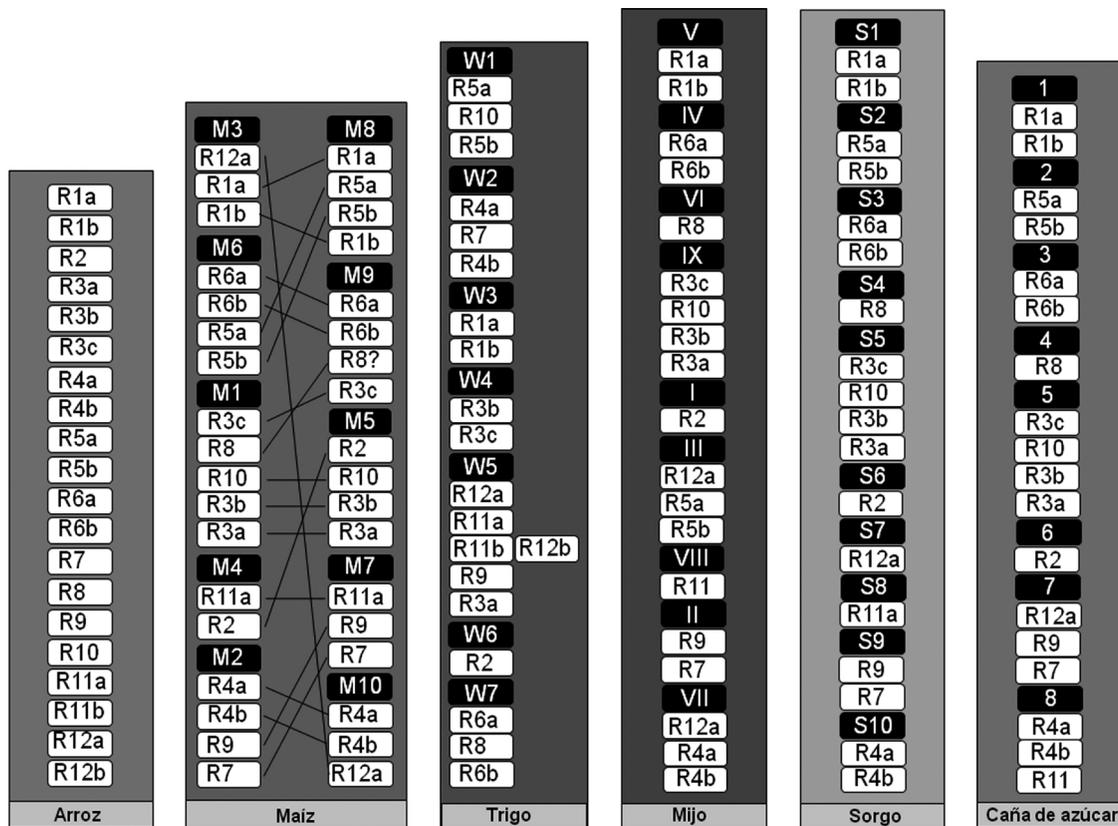
El genoma de arroz está compuesto por 19 bloques génicos que, reordenados como “*bloques de Lego*”, permiten reconstruir el genoma de las *Tritíceas*, maíz, *Setaria*, caña de azúcar y sorgo (Fig. 6). Considerados como una unidad genética representarían el genoma ancestral de las gramíneas, el cual tendría un único par de cromosomas.

La familia de las gramíneas es probablemente la mejor caracterizada en este aspecto, contando con una buena cantidad de *mapas genéticos comparativos* que demuestran las relaciones interespecíficas (Fig. 7). En términos genéticos, las gramíneas representan una familia muy diversa. Se estima que el genoma de las mismas divergió hace alrededor de 65 millones de años desde un antecesor común.

El nivel de ploidía y el número básico de cromosomas son muy variables, así como el tamaño del genoma.

### 6 Algunas generalizaciones acerca de los genomas

Los estudios en especies modelo han permitido arribar a varias generalizaciones. Una de ellas es que **el número de genes no es la base de la complejidad**. La mosca de la fruta tiene unos 13.000 genes, *Caenorhabditis elegans* 18.000, *Arabidopsis* 26.000 y los humanos 30.000. Si el número de genes no es muy diferente entre estas especies, ¿cuál es la base de la mayor complejidad en los humanos? La explicación estaría en el proteoma, más que en el genoma. Las proteínas codificadas por los genes pueden agruparse en familias en base a su similitud y muchas de estas familias proteicas son compartidas por todos los grupos



**Figura 6:** El genoma de arroz está compuesto por 19 bloques génicos que, reordenados como «*bloques de Lego*», permiten reconstruir el genoma de las *Tritíceas*, maíz, *Setaria*, caña de azúcar y sorgo. Estos bloques, considerados como una unidad genética, representarían el genoma ancestral de las gramíneas, el cual tendría un único par de cromosomas.

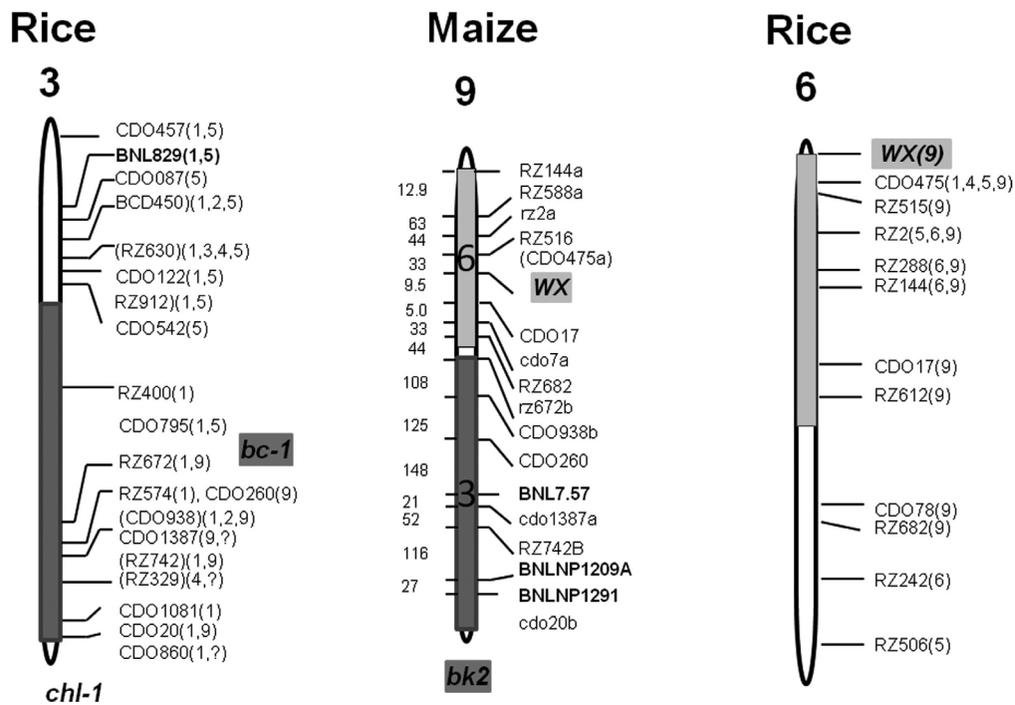


Figura 7: Alineamiento de cromosomas de arroz (3, 6) y maíz (9).

mencionados, aunque el número de miembros por familia es mayor en humanos.

Esto es particularmente evidente en aquellos genes involucrados en el desarrollo. Los humanos tenemos 30 genes para factores de crecimiento del fibroblasto, mientras que *Drosophila* y *Caenorhabditis elegans* tienen 2. Las nuevas proteínas surgen a través de cortes y empalmes alternativos de un mismo ARN mensajero (*alternative splicing*), lo que permite aumentar considerablemente la diversidad proteica a partir de los mismos mensajes. Las predicciones indican que el 60% de los genes humanos tiene dos o más alternativas de empalme. En *Caenorhabditis* el 22%.

Un elevado número de factores de transcripción y modificaciones postraduccionales también conducen a una mayor complejidad. En el caso de *Arabidopsis*, cerca del 9% de los genes son factores de transcripción (FT). Si se compara la proporción con los genomas de otras especies, se observa que *Arabidopsis* no sólo tiene más genes que algunos animales pequeños, sino que la proporción de FT es más alta que en cualquier clase de organismo

estudiado. Esto parece reflejar el hecho de que las plantas deben adaptarse a una amplia variedad de condiciones medioambientales, a diferencia de los animales, que pueden movilizarse y cambiar de lugar si este no les resulta propicio.

En humanos, los genes ocupan una cuarta parte del genoma, y sólo el 1,5% codifica para proteínas. Las secuencias correspondientes a los exones (secuencias del mensajero que se encuentran representadas en la proteína, las secuencias correspondientes a los intrones no lo están) comprenden un porcentaje bajo del genoma, mientras que los intrones comprenden el 24%. Los genes no están distribuidos en forma pareja. Algunos se encuentran agrupados en ciertas regiones del genoma mientras que otros se encuentran aislados. En la mosca, el nematodo y *Arabidopsis* la disposición de los genes es mucho más regular.

Aproximadamente **la mitad del genoma humano consiste de secuencias repetitivas**, siendo la mayoría **elementos transponibles** que se propagan replicándose e insertando una copia de sí mismos en otras regiones del

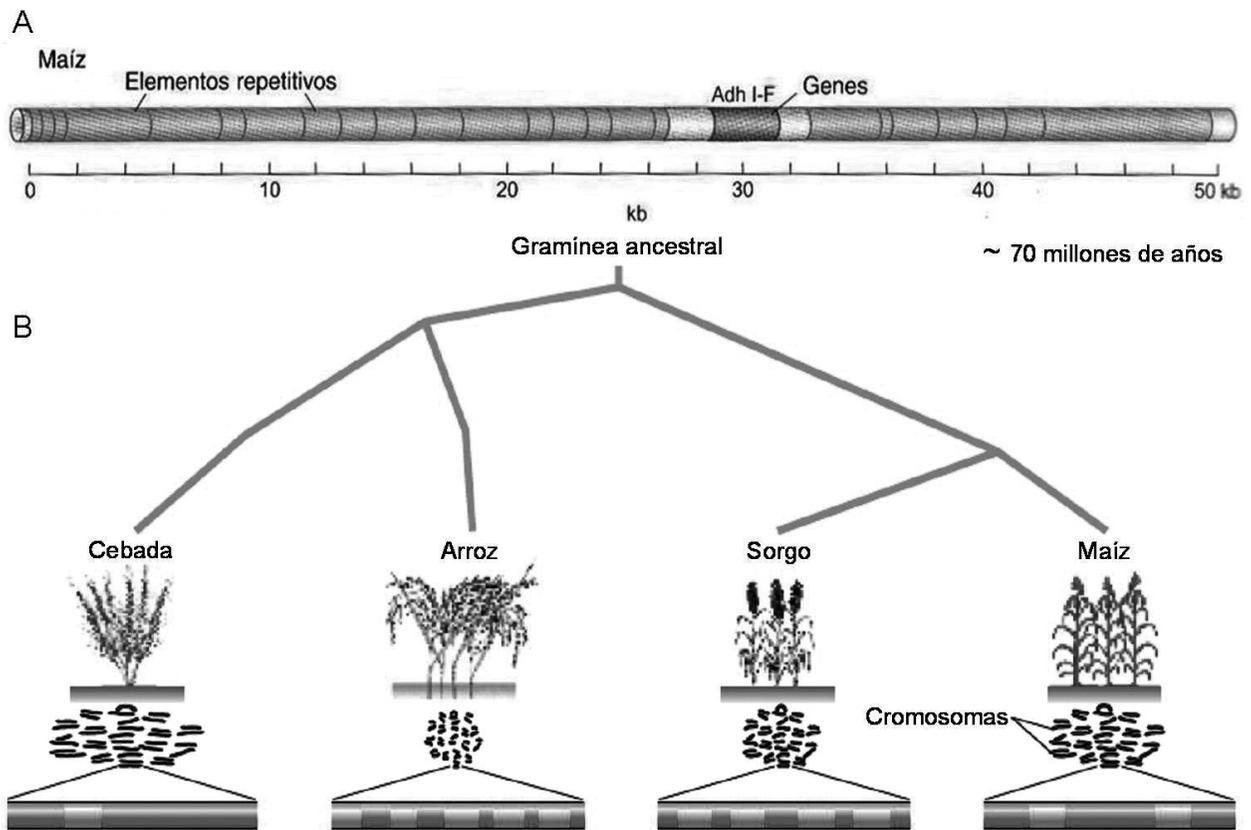
genoma. En la actualidad solo dos tipos de elementos son activos, Alu y LINE1. La mayoría de los mismos se encuentran en regiones ricas en A y T, mientras que los genes se encuentran en áreas con elevado contenido de G y C. Fragmentos de estos elementos también se encontraron en las secuencias regulatorias que controlan la expresión de varios genes, por lo que se ha sugerido que, de alguna manera, podrían afectar positivamente su expresión y evolución.

Los genomas vegetales también están plagados de secuencias repetitivas (transposones y retrotransposones), que constituyen un porcentaje muy importante del ADN nuclear y han contribuido, a lo largo de la evolución, a la ex-

pansión de los genomas. En maíz representan del 50 al 80% del genoma y en trigo el 80% (Fig. 8). De esta manera los genomas de los cereales pueden considerarse como islas de genes que se hallan dispersas en un mar de secuencias repetitivas.

### 7 Genómica y Agricultura

Especialistas en varias disciplinas han llegado a la conclusión de que el alcance de la genómica será mayor en lo que se refiere a la economía y la calidad de vida que en el ámbito de la salud humana. Gracias al aporte de grupos de investigación de todo el mundo se dispone de bases de datos públicas con información genómica de utilidad para los mejoradores. El



**Figura 8:** a) Elementos repetitivos en un fragmento cromosómico de maíz. b) Diagrama que muestra como los elementos transponibles en gramíneas son responsables del incremento en el tamaño del genoma. El arroz, sorgo, cebada y maíz derivaron de un ancestro común hace aproximadamente 70 millones de años. Desde entonces, los transposones y retrotransposones se han ido acumulando a diferentes niveles en las especies. Los cromosomas son más largos en maíz y cebada, cuyos genomas contienen grandes cantidades de retos con LTR (long terminal repeats). En verde se señalan los elementos transponibles y en naranja los genes. Modificada de Griffiths *et al.* (2000).

conocimiento de cómo actúan los genes en una especie puede ayudar a los mejoradores a perfeccionar su función en otra. Las secuencias de los genomas de *Arabidopsis* y de arroz proporcionan información relevante acerca de otros cultivos, como el maíz y el trigo. Dado que estos tres cereales representan más de la mitad de la producción alimentaria mundial y que el arroz es el alimento básico de más de la mitad de la población del planeta, la trascendencia de la genómica en la agricultura es indiscutible.

La similitud entre las especies de cereales también implica que cuando se trasladan genes de una especie a otra, tenderán a funcionar bien y en la misma forma con una mínima manipulación genética.

A través de esfuerzos públicos y privados se trabaja en la identificación de genes asociados con caracteres como rendimiento, resistencia a la sequía, calidad de los alimentos, resistencia a insectos y tolerancia a herbicidas.

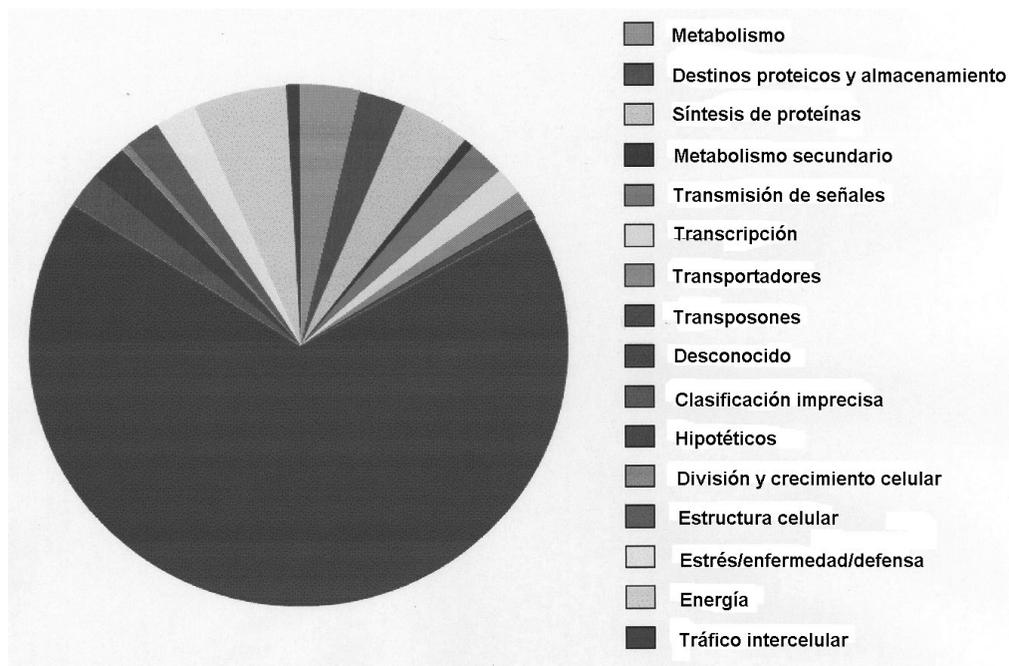
De esta manera se ha acelerado significativamente el aporte de nuevas variedades al mercado, ya sea por modificación genética o por selección con marcadores moleculares o

utilizando criterios de selección a partir de la información molecular.

La **genómica comparativa**, basada en el análisis de ESTs de diferentes plantas tolerantes a estreses abióticos, ha permitido la identificación de redes de genes comunes asociados con estreses ambientales, tales como salinidad, sequía, bajas y altas temperaturas. El análisis de especies tolerantes a situaciones extremas ha identificado genes involucrados en los mecanismos que las hacen tolerantes. Dichos genes podrán entonces ser transferidos a especies de cultivo.

Como ejemplos pueden citarse *Agrostis adamsonii* y *Agrostis robusta*, tolerantes a salinidad, *Microlaena stipoides*, tolerante a aluminio y *Deschampsia antarctica*, tolerante a bajas temperaturas (la única gramínea que crece en la Antártida) (**Fig. 9**).

También se ha mencionado que los factores de transcripción serían las "llaves" para desbloquear funciones génicas que aprovechen la diversidad de la naturaleza para producir mejores cultivos. Estos son genes que virtualmente controlan todo carácter importante, desde el



**Figura 9:** Categorización funcional de ESTs de *Deschampsia antarctica*. El objetivo de este proyecto es encontrar genes involucrados en la resistencia a bajas temperaturas para ser transferidos, por tecnología génica, a especies de cultivo. Gentileza de G. Spangenberg.

punto de vista agrícola, en las plantas, incluyendo rendimiento, resistencia a las enfermedades, protección contra las heladas y sequía, producción de químicos, proteínas usadas en farmacéutica, etc. La mayoría de los caracteres vegetales a los que interesa aplicar la ingeniería genética son multigénicos (ej. tolerancia al estrés hídrico) y los genes involucrados se expresan en cascadas de forma tal que genes primarios activan a genes secundarios que a su vez activan a genes terciarios. Los factores de transcripción serían los responsables de hacer funcionar a los genes primarios, obteniéndose largas cascadas de expresión de genes por la manipulación de un solo gen.

Durante los últimos años se ha obtenido una gran colección de FT de plantas, funcionalmente caracterizados, no sólo en *Arabidopsis*, sino también en especies como tomate, maíz y soja. En total hay cerca de 1900 FT en *Arabidopsis* pero en el contexto de la **genómica funcional** el interés recae en unas 60 familias que tienen funciones reconocidas. La herramienta primaria que se utiliza para conocer la función de los FT consiste en sobreexpresarlos y/o detener la expresión de algunos de ellos. Se puede, por ejemplo, medir el efecto de cada FT en la composición de los lípidos en las semillas aceiteras y la composición de lípidos en las hojas, la composición de azúcares, de esteroides, etc., estudiar procesos básicos, a través de la dilucidación de sus efectos en el desarrollo vegetal y la resistencia a las enfermedades, o tratar de identificar el FT que regula el consumo de nitrógeno (N). El N es uno de los insumos más costosos para la agricultura, por lo que identificar los genes que controlan y modifican la respuesta al N sería de gran valor.

También existen proyectos genómicos relacionados con leguminosas y con sus simbiontes fijadores de nitrógeno. Recientemente se han secuenciado los genomas de las especies de *Mesorhizobium* y *Sinorhizobium* y se han producido cientos de miles de EST de *Medicago truncatula*, *Lotus corniculatus* y soja. En el caso de *M. truncatula* no sólo se han disectado los pasos que controlan las señales entre la bacteria fijadora y la planta sino también entre la planta y otro simbionte, las micorrizas (aso-

ciación de un hongo con la raíz de una planta superior. Esta asociación no es específica de especie como el caso de *Rhizobium* y las leguminosas. El hongo puede ser de varios géneros y aporta fósforo a la planta). Recientemente se logró la identificación, en alfalfa, de un receptor requerido para el reconocimiento de señales bacterianas de nodulación, los llamados factores NOD.

Las plantas medicinales aportan el 25% de los compuestos activos utilizados actualmente por la industria farmacéutica. Entre estos compuestos se encuentran los citocromos P450, que participan en la biosíntesis de muchos compuestos anticancerígenos, alcaloides, fitoesteroides, antioxidantes y antimicrobianos. También tienen un rol muy importante en la detoxificación de xenobióticos (herbicidas y pesticidas). Se han identificado 1052 citocromos P450 y el objetivo actual es, utilizando herramientas de metabolómica, determinar su función bioquímica precisa.

Otro de los objetivos de la metabolómica es el estudio de la síntesis de las esencias que confieren perfume a las flores y el mejoramiento del sabor de algunas plantas utilizadas en alimentación humana, como por ejemplo mandioca, donde el objetivo sería la eliminación de los glucósidos cianogénicos que le confieren sabor amargo.

La devastación de los bosques es motivo suficiente para invertir en proyectos de genómica tendientes a la identificación de genes involucrados en perennidad, desarrollo, interacción con herbívoros, respuesta a estrés y síntesis de metabolitos secundarios como lignina y celulosa. Se han generado ESTs de álamo, abedul, abeto, pino y eucaliptos. El álamo se utiliza como sistema modelo para el análisis de los genes involucrados en la formación de madera. Otro objetivo en este campo es la búsqueda de estrategias para inducir floración temprana, que es un factor crítico en el mejoramiento de forestales.

La secuenciación de los genes y la dilucidación de las vías que controlan la floración en los cereales, que difiere en varios aspectos con los correspondientes en *Arabidopsis*, ha sido otro de los logros de la genómica. También se localizaron y caracterizaron, en trigo y cebada,

los genes *VRN1* y *VRN2*, que son los genes centrales en la vía de vernalización en trigo, cebada y otros cereales invernales. Este conocimiento es clave en agricultura, ya que permitiría, potencialmente, controlar la floración en cultivos de gran importancia económica, como los cereales y los forrajes.

### 8 Aportes de la genómica al mejoramiento de cereales

El trigo pan (*Triticum aestivum* L 2n= 6x= 42; AABBDD) es uno de los principales cultivos alimenticios ya que provee cerca del 55% de los carbohidratos consumidos por el hombre. La genética de este cultivo resulta compleja, es de naturaleza hexaploide, con tres genomas homeólogos denominados A, B y D que aportan 7 pares de cromosomas cada uno. Los genes redundantes son una norma, con sets homoealélicos triplicados en la mayoría de ellos. El genoma del trigo hexaploide es el de mayor tamaño (17000 Mbp trigo, 2800 Mbp maíz, 800 Mbp sorgo, 450 Mbp arroz). Más del 80% del mismo está constituido por secuencias de ADN altamente repetitivo. El restante 20% está compuesto por secuencias de bajo número de copias o copia única, donde se encuentran la mayoría de los genes. Si bien las secuencias altamente repetitivas varían de especie en especie, las secuencias génicas, en general, son conservadas. Esta característica permite el uso de **sondas heterólogas** (de otros genomas) en experimentos de hibridación de tipo RFLP (Parte I, Cap. 5) para identificar secuencias conservadas en especies diferentes dentro de una misma familia taxonómica (Devos y Gale, 2000).

En 1989 se publicó el primer mapa genético molecular de trigo, correspondiente a los cromosomas homeólogos del grupo 7, observándose que el orden de los genes y marcadores se conserva en gran parte de los tres genomas de trigo hexaploide. Esta fue la primera prueba de colinearidad en gramíneas y permitió el desarrollo de la genética comparativa en cereales.

En trabajos posteriores se comprobó que largos segmentos de los cromosomas de maíz, sorgo, arroz, trigo y cebada conservan la presencia y el orden de marcadores y genes,

aunque en algunos casos la correspondencia cromosómica fue modificada por duplicaciones, inversiones o translocaciones. De esta manera se logró consensuar un mapa unificado de gramíneas en el que se detallan secuencias y genes conservados en diferentes genomas (**Fig. 2**). Utilizando estas herramientas fue posible transferir información proveniente de plantas de genomas pequeños (arroz, sorgo) a plantas de genomas más complejos como el trigo. Esto ha facilitado la localización más precisa de genes para tolerancia a estrés biótico y abiótico (prebrotado, dormición, vernalización, frío y otros) y el desarrollo de marcadores útiles para el mejoramiento. También ha conducido a la obtención de mapas consenso específicos para trigo pan, candeal, maíz y cebada.

Las nuevas técnicas de genómica funcional, epigenética, mapeo por asociación, "TILLING" y los ARN interferentes (ARNi), entre otras, contribuirán a un mayor conocimiento del funcionamiento del genoma de este cereal. Existen genotecas de BACs/BIBACs para todos los genomas de trigo. Estas representan un importante complemento para saturar los mapas.

En los últimos años fueron confeccionado numerosos mapas genéticos de trigo pan, candeal y cebada a partir de poblaciones de RILs (líneas recombinantes endocriadas) o haploides duplicados, que resultan útiles en estudios de identificación de QTL asociados a calidad y estreses bióticos (enfermedades) y abióticos (sequía, salinidad). Para ello se utilizaron y desarrollaron miles de marcadores moleculares, incluyendo RFLPs, SSRs, AFLPs, SNPs, y marcadores DArT (*Diversity array technology*), (estos últimos basados en matrices de diversidad). La información generada permitió el desarrollo de marcadores perfectos, a partir de la secuencia del gen a seleccionar, en lugar de marcadores genéticamente ligados. Como ejemplo pueden citarse los genes de las enzimas polifenol oxidasas, lipoxigenasas y fitoeno sintasa, entre otros. También permitió la confección de mapas consenso para trigo pan, candeal y cebada.

El acceso a los principales genes que afectan la calidad ha abierto nuevas vías para el desa-

rollo comercial de diferentes productos derivados del trigo. Es posible obtener variedades con diferente calidad de almidón, diferente textura de grano, diferente composición de gluteninas, gliadinas y secalinas en función del uso industrial. El descubrimiento de genes valiosos de otros genomas y la posibilidad de incorporarlos al trigo por transgénesis permitirán resolver problemas del cultivo, como pueden ser limitantes debido a estreses bióticos y abióticos o desarrollar nuevas generaciones de transgénicos en función de las necesidades del consumidor, con mayor contenido de aminoácidos esenciales como la lisina, vitaminas, etc.

La complejidad del genoma del trigo ha hecho aconsejable abordar los estudios genómicos a través del transcriptoma. Para ello se han establecido consorcios internacionales como el International Wheat Genome Sequencing Consortium (IWGSC) o el International Triticeae Mapping Initiative (ITMI), el cual coordina grupos de investigación de distintos países en la construcción de mapas moleculares del genoma de trigo. Existe colaboración internacional para otros cultivos como el arroz, donde el Internacional Rice Genome Sequencing Project (IRGSP) coordina esfuerzos de 10 países y como resultado de esto, en 2002, fueron colocadas en bases de datos públicas secuencias de alta calidad que representaban 366 Mbp del genoma de arroz. En el 2005 se llegó a 371 Mbp, que representan el 95% del mismo. Las bases de datos de EST han crecido exponencialmente en la última década, de manera que en Septiembre de 2008 había 55 millones de ESTs depositadas en el National Center for Biotechnology Information dbEST database (<http://www.ncbi.nlm.nih.gov/dbEST>). Esto incluye cerca de 1.607.934 de ESTs de trigo hexaploide y sus parientes más cercanos de la tribu Triticeae, que son *Hordeum vulgare* L.; especies diploides y tetraploides de *Triticum*, *Secale cereale* L. y *Aegilops speltoides* Tausch. Estas ESTs se utilizan para el desarrollo de marcadores funcionales, la preparación de mapas de transcriptos y la construcción de matrices de ADNc. Actualmente el estudio de ARN interferentes, TILLING y la “genética de expresión” lideran el mapeo de eQTL (*Quantitative Traits Loci* de expresión) y han sido utilizados para identificar funciones de genes individuales.

Toda esta información permite inferir que la biotecnología puede cambiar el escenario de los cereales en dos áreas principales: (1) protegiendo al cultivo de estreses bióticos y abióticos y (2) modificando el concepto de producción de «commodities» por el de producción de «specialities», por ejemplo para derivados de trigo con mayor valor agregado, como noodles (clase de fideo muy utilizado en Asia), galletitas dulces, crackers (un tipo de galletita que se rompe fácilmente), masa congelada, pan de molde, pasta, almidones, plásticos biodegradables, etc.

Se han identificado cinco áreas de investigación a desarrollar en los próximos años para el mejoramiento de trigo, que pueden aplicarse a los cereales en general: (i) mapeo genético, (ii) análisis de QTL, (iii) mejoramiento molecular, (iv) mapeo por asociación, y (v) desarrollo de software.

### 9 Genomas trabajadores

La genómica aplicada a dilucidar los mecanismos por los cuales funcionan los microorganismos puede conducir al aislamiento de sus componentes para desarrollar nanoestructuras que lleven a cabo funciones complejas. Como ejemplos pueden citarse:

***Methanococcus jannaschii***: es una bacteria que produce metano, una importante fuente de energía. Contiene enzimas que soportan temperaturas y presiones elevadas por lo que son potencialmente útiles para fines industriales.

***Deinococcus radiodurans***: soporta niveles de radiación extremadamente elevados por lo cual tiene un alto potencial para limpiar desechos radiactivos.

***Thalassiosira pseudonana***: se trata de una diatomea marina que es la principal participante en el bombeo biológico de carbono hacia las profundidades de los océanos, por lo que posee un elevado potencial para mitigar los cambios climáticos del planeta.

### 10 Proyectos de Genómica en Sudamérica

En Sudamérica se han desarrollado y se encuentran en ejecución algunos proyectos genómicos. Frente al estado del desarrollo de estas áreas en el mundo, existen en la región serias carencias, tanto de recursos humanos

entrenados como de infraestructura y equipamiento. En Brasil se estableció una red de cooperación que secuenció completamente el genoma de *Xylella fastidiosa*, bacteria que ataca a los cítricos causando importantes pérdidas económicas. También incursionó en el estudio de genómica funcional de células cancerosas y de secuenciación del genoma de caña de azúcar, *Saccharum officinalis*. En Argentina se ha trabajado en la secuenciación de *Brucella abortus*, agente causal de la brucelosis bovina.

Investigadores argentinos participan también en proyectos de genómica en cooperaciones internacionales, como el consorcio involucrado en la secuenciación del genoma de *Trypanosoma cruzi*, el proyecto que llevó a la clonación y caracterización de los genes de vernalización en trigo o el proyecto de búsqueda de genes de tolerancia a frío en *Deschampsia antarctica*. En Chile se ha completado recientemente la secuenciación de *Piscirickettsia salmonis*, patógeno intracelular de salmones que causa importantes pérdidas en esta industria. También se trabaja activamente en especies hortícolas, principalmente frutales.

Otros ejemplos son el desarrollo de **marcadores microsatélites** (Parte I, Cap. 5) de diversas especies, como girasol (colaboración de INTA-Argentina con empresas semilleras locales), vides (INIA-Chile como parte de un consorcio internacional), alpacas y otros camélidos sudamericanos (INIA-Chile), pasto llorón (CERZOS, UNR, INTA Castelar), entre otros. El girasol es una especie de importancia económica en la Argentina y es el eje de varios proyectos genómicos. Algunos de los objetivos son la caracterización de la diversidad funcional del girasol y la identificación de SNPs asociados a caracteres agronómicos de importancia como la resistencia a estreses bióticos y abióticos. Por otro lado, el desarrollo de marcadores microsatélites, ESTs y el mapeo genético de las mismas es utilizado para asistir al mejoramiento y realizar identificación varietal.

En Chile se ha desarrollado un programa de genómica funcional coordinado por diversos ministerios, destinado a enfrentar problemas de postcosecha y enfermedades en plantas de interés agrícola. Otro proyecto de genómica, en este mismo país, apunta a desarrollar plata-

formas tecnológicas en genómica forestal con el objeto de mejorar la posición competitiva del sector forestal chileno.

En el marco del Programa Cooperativo para el Desarrollo Tecnológico Agroalimentario y Agroindustrial del Cono Sur (PROCISUR), los Institutos Nacionales de Investigación Agropecuaria de Argentina, Bolivia, Brasil, Chile, Paraguay y Uruguay han conformado una plataforma regional para enfrentar los desafíos que impone la secuenciación del genoma de la papa. Esta iniciativa es desarrollada a nivel mundial por el Consorcio Internacional de Secuenciación del Genoma de la Papa, que lidera la Universidad de Wageningen de Holanda y en la que están comprometidos China, Estados Unidos, Holanda, India, Inglaterra, Irlanda, Nueva Zelanda, Polonia, Rusia y los sudamericanos Chile, Perú, Argentina y Brasil. Estos últimos, en conjunto, secuenciarán el cromosoma 3 del genoma de la papa y en total se deberán secuenciar los 12 cromosomas de la especie.

En varios países se han secuenciado fragmentos de genomas de virus y viroides (Argentina, Chile, Uruguay). En Argentina se está realizando la caracterización biológica y molecular del virus del mal de Río Cuarto (MRCV) con el fin de estimar la diversidad genética de las poblaciones del virus, detectar la existencia de razas y limitar la propagación de esta enfermedad de importancia para el maíz.

En el año 2003 la Agencia Nacional de Promoción Científica y Tecnológica de Argentina (ANPCyT) lanzó una convocatoria para Proyectos de Área de Vacancia (PAV), donde la genómica era una de las prioridades. En respuesta a esta convocatoria se formó una Red de laboratorios dedicados al análisis genómico funcional y comparativo en especies de interés agropecuario, forestal o ambiental (PAV137). Los objetivos fueron generar una infraestructura operativa y comunicacional adecuada y formar recursos humanos necesarios para apoyar el desarrollo de iniciativas e investigaciones en áreas de genómica funcional y bioinformática críticas para la biología agropecuaria, forestal y ambiental. La red pudo capacitar recursos humanos orientados a la búsqueda, prospección y/o análisis funcional de

genes con interés básico o aplicado. Se abordaron cinco áreas de desarrollo e integración: a) transcriptómica, b) proteómica, interactómica y metabolómica, c) genómica comparativa, d) genómica evolutiva para la caracterización de la diversidad genética y e) ecogenómica (aplicada a la evaluación de impacto ambiental) y epidemiología molecular. En los dos primeros subproyectos se encararon trabajos de prospección y caracterización funcional de regiones genómicas de interés en plantas superiores y bacterias zoonóticas mediante el análisis de perfiles de transcripción, interacción de proteínas y perfiles metabólicos (girasol, citrus, soja, solanáceas, *Mycobacterium* y otros). En el c) se caracterizaron molecularmente regiones genómicas involucradas en características reproductivas (apomixis) de especies forrajeras, de resistencia a estreses en especies cultivadas, así como de características productivas en girasol y caprinos. El área d) sirvió para evaluar la diversidad genética de recursos biológicos naturales o cultivados, en particular en especies leguminosas y en forestales. La e) permitió estudiar la dinámica poblacional de variantes alélicas en ecosistemas (impacto del cultivo de maíces transgénicos en potenciales genes de resistencia en insectos plaga, cuantificación de diversidad genética de especies en peligro de extinción en ecosistemas naturales explotados por el hombre) y encarar la epidemiología molecular de la fiebre aftosa y de la peste porcina clásica.

Este proyecto acaba de finalizar (septiembre de 2008) y, si bien aún no se ha realizado la evaluación final del mismo, varios de los objetivos fueron cumplidos exitosamente, ya que se formaron varios doctores en el área en distintas Universidades del país, entrenados para trabajar en equipos multidisciplinarios, además de lograr avances en el conocimiento en las mencionadas áreas.

El INIA (Uruguay) cuenta con proyectos de desarrollo y validación de herramientas bioinformáticas para integrar información genómica en programas de fitomejoramiento y selección de germoplasma. Dentro de los proyectos de genómica de arroz, el EMBRAPA (Brasil), realiza el estudio de genes y sus productos, involucrados en los mecanismos moleculares de

susceptibilidad y resistencia a estreses bióticos, a través de estudios de genómica funcional, utilizando ESTs y unigenes y micromatrices de ADNc.

Los objetivos para programas de genómica sudamericanos deberían enfocarse hacia el aumento de la productividad y la mejora de la calidad de los productos, desarrollando capacidades para identificar genes del germoplasma regional, de manera de disminuir la dependencia de variedades y genes desarrollados y aislados por países del primer mundo y buscar soluciones para problemas propios de la región, que difícilmente pueden ser enfrentados por programas de mejoramiento genético ajenos a la misma.

## 11 Lecturas / sitios recomendados

- Cenci A., Chantret N., Xy K., Gu Y., Anderson O.D., Fahima T., Distelfeld A. Y Dubcovsky J. 2003. Construction and characterization of a half million clones Bacterial Artificial Chromosome (BAC) library of durum wheat. *Theor Appl Genet* 107: 931-939.
- Cervigni G., Paniego N., Díaz M., Selva J.P., Zapacosta D., Zanazzi D., Landerreche I., Felitti S., Pessino S., Spangenberg G. And Echenique V. 2008. Expressed sequence tag analysis and development of gene associated markers in a near-isogenic plant system of *Eragrostis curvula*. *Plant Molecular Biology*. 67: 1-10.
- Devos K.M Y Gale M.D. 2000. Genome relationship: the grass model in current research. *Plant Cell* 12: 637-646
- Echenique V., Stamova B., Wolters P., Lazo G., Carollo V. Y Dubcovsky J. 2002. Frequencies of Ty1-*copia* and Ty3-*gypsy* retroelements within the *Triticeae* EST databases. *Theor. Appl. Genet.* 104: 840-844.
- Gale M. Y Devos K. 1998. Plant comparative genetics after 10 years. *Science*, 282: 656-658.
- Gibson D.G., Benders G.A., Andrews-Pfannkoch C., Denisova E.A., Baden-Tillson H., Zaveri J., Stockwell T.B., Brownley A., Thomas D.W., Algire M.A., Merryman C., Young L., Noskov V.N., Glass J.I., Venter J.C., Hutchison Iii C.A. Y Smith H.O. 2008. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, 319: 1215-1220.
- Glass J.I., Assad-Garcia N., Alperovich N., Yooseph S., Lewis M.R., Maruf M., Hutchison Iii C.A., Smith H.O. Y Venter J.C. 2005. Essential genes of

- a minimal bacterium. Proc. Nat. Acad. Sci. 103: 425-430.
- Goff S.A., Ricke D., Lan T.H., Presting G., Wang R., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H. *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science, 296: 92-100.
- Griffiths A., Miller J., Suzuki D., Lewontin R. Y Gelbart W. 2004. An introduction to genetic analysis. Octava Edición. W.H. Freeman, N.York. 860 pp.
- Gupta P.K., Mir R.R., Mohan A. And Kumar J. 2008. Wheat genomics: Present status and future prospects. International Journal of Plant Genomics. Article ID 896451, 36 pages. doi:10.1155/2008/896451
- Kumar A. Y Bennetzen J.L. 1999. Plant retrotransposon. Annu Rev Genet 33: 479-532.
- Lacadena J.R. 2000. Seréis como dioses. *Crítica (Madrid)*, 874: 12-16.
- Lartigue C., Glass J.I., Alperovich N., Pieper R., Parmar P.P., Hutchison Iii C.A., Smith H.O. Y Venter J.C. 2007. Genome transplantation in bacteria: changing one species to another. Science 317: 632-638.
- Moore G. 2000. Cereal chromosome structure, evolution and pairing. Ann. Rev. Plant Mol. Biol. 51: 195-222
- Sanmiguel P., Ramakrishna W., Bennetzen J.L., Busso C. Y Dubcovsky J. 2002. Transposable elements, genes and recombination in a 215 kb. contig from wheat chromosome 5A. Functional and Integrative Genomics. 2: 70-80.
- Wade C.M., Kulbokas E.J. Iii., Kirby A.W., Zody M.C., Mullikin J.C., Lander E.S., Lindblad-Toh K. Y Daly M.J. 2002. The mosaic structure of variation in the laboratory mouse genome. Nature, 420: 574-578.
- Warren R.L., Varabei D., Platt D., Huang X., Messina D., Yang S., Kronstad J.W., Krzywinski M., Warren W.C., Wallis J.W., Hillier L.W., Chinwalla A.T., Schein, J.E. Siddiqui A.S., Marra M.A., Wilson R.K., Y Jones S.J.M.. 2006. Physical map-assisted whole-genome shotgun sequence assemblies. Genome Res. 16: 768-775.
- Yu J., Hu S., Wang J., Wong G.K., Li S., Liu B., Deng Y., Dai L., Zhou Y., Zhang X. *Et Al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science, 296: 79-92.
- The *Arabidopsis* Genome Initiative: Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000. 408: 796-815.



## I CAPÍTULO 8

### Transcriptómica

Silvina Felitti y Silvina Pessino

El análisis de los niveles de representación de ARN mensajeros proporciona información importante sobre la actividad de un gen, evidenciando si está siendo copiado para posteriormente dirigir la síntesis de la proteína a la cual codifica. En los últimos años los procedimientos para la detección de los niveles de ARN mensajeros han progresado velozmente desde el análisis de genes únicos específicos (como el Northern, slot, y dot blotting, la RT-PCR semicuantitativa y cuantitativa y los ensayos de protección de nucleasas) hacia otros enfocados en la identificación de múltiples transcriptos que difieren en su representación entre las diversas muestras experimentales (como la hibridización substractiva, el display diferencial, el ADNc-AFLPs®, la secuenciación de etiquetas expresadas o ESTs, el análisis serial de la expresión de genes o SAGE y la hibridización de microarreglos). La organización posterior de las secuencias dentro de grupos funcionales basada en los datos de homología proporciona un marco básico para conducir nuevos estudios dirigidos a definir el rol biológico de cada producto génico. Por otra parte, la aplicación de técnicas de PCR en tiempo real y de hibridización *in situ* de tejidos permiten una validación más precisa de los datos de expresión diferencial obtenidos por métodos anteriormente citados.

La capacidad tecnológica creciente permite ahora capturar sectores de tejidos o incluso células aisladas y analizar su contenido de ARNm, lo que provee información específica sobre la representación del transcriptoma para tipos celulares únicos. La asociación de los datos provenientes de la secuenciación genómica con aquellos originados en la transcriptómica y la proteómica facilita la predicción de la existencia de fragmentos codificantes en sectores acotados del genoma, y da información sobre la posible función de los candidatos en tejidos particulares. Además, la asociación de los da-

tos de posición (originados en el mapeo genético) con los datos de expresión (originados en el análisis del transcriptoma) posibilita la selección de candidatos responsables de disparar un determinado carácter de interés.

Este capítulo tiene como objetivo presentar en forma abreviada una serie de metodologías que son utilizadas habitualmente para el análisis del transcriptoma de plantas. Debido al muy rápido cambio y perfeccionamiento en las técnicas empleadas para abordar el estudio de la representación de mensajeros tanto a nivel de mesada como bioinformático, algunas de las metodologías mencionadas aquí ya han sido reemplazadas y se usan actualmente en forma poco frecuente. Sin embargo hemos decidido incluirlas, dada su gran importancia histórica en relación al cambio de abordaje conceptual que va desde el análisis de la actividad de genes únicos al estudio integral de la expresión génica. Comenzaremos describiendo las técnicas en orden cronológico de desarrollo.

#### Hibridización substractiva

Los métodos de hibridización substractiva fueron creados a principios de los años 80 con el propósito de construir bibliotecas de ADNc para obtener sondas de genes expresados diferencialmente. Fueron los primeros que se utilizaron de manera amplia con el propósito de identificar genes regulados positiva o negativamente en una escala global. Sus ventajas incluyen la habilidad de aislar genes de función relacionada sin tener conocimientos previos de su secuencia o identidad y el uso de técnicas comunes de biología molecular que no requieren equipos especializados de detección y análisis. El procedimiento general consiste en la hibridización de un ADNc proveniente de una muestra prueba (tester) con un exceso de ARNm proveniente de una muestra control (driver). Los transcriptos expresados en ambas muestras (tester y driver) forman moléculas de ARNm/ADNc híbridas, mientras que las secuencia de ADNc que están presentes únicamente en la muestra tester permanecen como hebras simples. Las moléculas de hebras simples y dobles se separan usando cromatografía en hidroxilapatita. Los ADNcs expresados diferencialmente pueden entonces ser recu-

perados y clonados o usados directamente como sondas para analizar una biblioteca. Dos limitaciones importantes del protocolo original son: i) el requerimiento de grandes cantidades de ARNm y ii) una tendencia a una menor eficiencia en la identificación de transcritos poco abundantes.

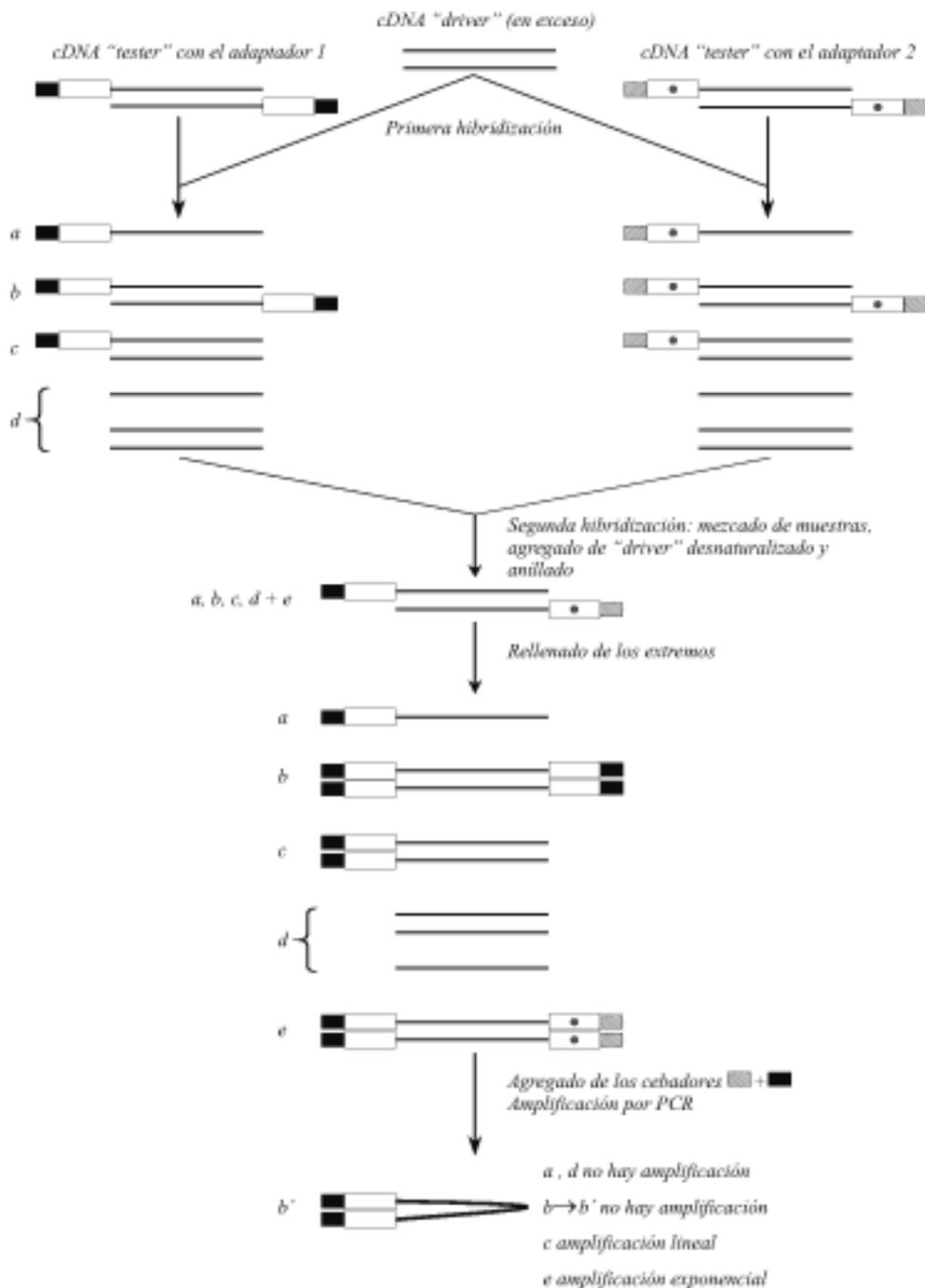
La hibridización substractiva es aplicable sólo a comparaciones de pares de muestras y debe ser realizada por duplicado con el tester y el driver invertidos para detectar tanto aumentos como disminuciones en los niveles de expresión. Además no genera una medida cuantitativa y aunque es eficiente en la identificación de genes que están completamente ausentes en el driver no revela fácilmente a aquellos que están presentes en ambas muestras en distinta proporción. Se realizaron modificaciones para mejorar el protocolo original que incluyen la producción de ADNc con marcas de biotina o oligo(dT)30-látex de manera de refinar la separación de las moléculas de cadena simple y doble. También se incorporó la amplificación de ADNcs selectos por PCR para disminuir la cantidad inicial de ARNm requerido y aumentar la eficiencia de clonado de los transcritos seleccionados. También se ha utilizado un protocolo ingenioso conocido como *hibridización substractiva de supresión* (SSH) que fue diseñado para favorecer la detección de transcritos raros expresados diferencialmente. Este incluye una normalización en la representación de los transcritos diferenciales basada en la inhibición de la amplificación de aquellos que son más abundantes, eliminando también la necesidad de separar moléculas de cadena doble y simple (ver Figura 1).

### **Display diferencial y RAP-PCR**

Las técnicas conocidas en general como "huellas digitales de ARN" (ARN fingerprinting) incluyen al display diferencial (DD) y la PCR de ARN cebada arbitrariamente (RAP-PCR). Ambos métodos están basados en una amplificación por PCR de subgrupos al azar de transcritos a partir de dos o más muestras. La primera etapa de ambos procedimientos es común y consiste en generar ADNcs haciendo una transcripción reversa de una fracción de las moléculas de ARNm de una muestra. En

el display diferencial esto se logra utilizando como cebador de la transcripción reversa a un poliT anclado con una o más bases adicionales al extremo 3' del ARN mensajero (por ejemplo (T)<sub>12</sub>AC). Estos oligonucleótidos se fijan al ARN mensajero a través de la unión de las bases adicionales, impidiendo que el poliT "resbale" sobre distintas zonas del poli A. El único grupo de ARNms que es transcrito en forma reversa es el integrado por las moléculas que llevan las bases complementarias a las bases adicionales del poliT en el extremo del mensajero adyacente al poliA. En contraste, la RAP-PCR utiliza oligonucleótidos de secuencia arbitraria para la transcripción reversa. Estos cebadores tienen normalmente 10 bases de largo y se unen a sus secuencias complementarias permitiendo formar una hebra de ADNc desde este sitio. En este caso el subgrupo de ARNms que sufre transcripción reversa estará integrado por aquellas moléculas que presenten la secuencia complementaria a la del cebador orientada en el sentido adecuado.

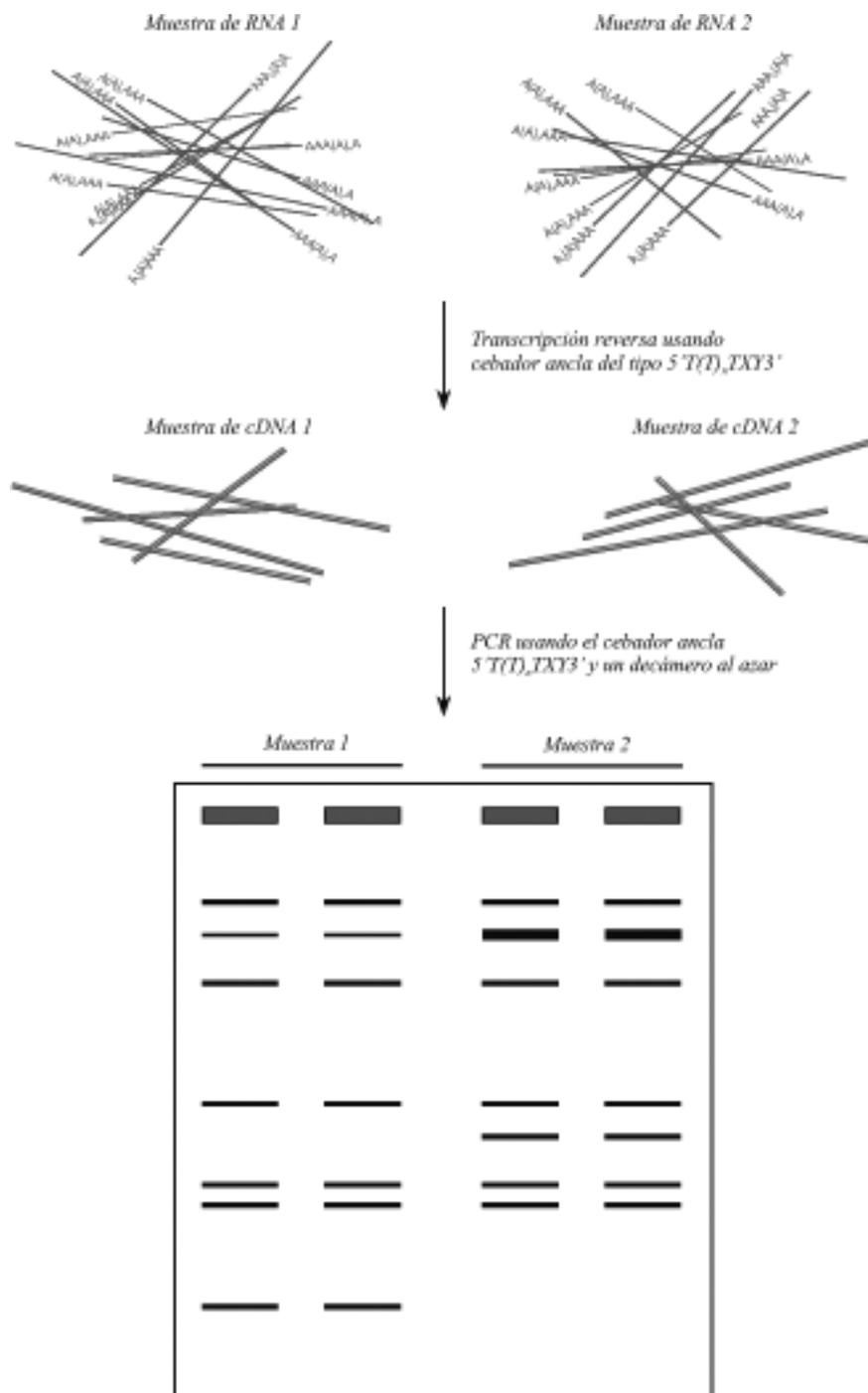
Luego de haberse realizado la síntesis de la primer hebra del ADNc, se amplifican segmentos de los transcritos usando pares múltiples de cebadores de PCR. Tanto para el display diferencial como para el RAP-PCR, el cebador superior es un oligonucleótido arbitrario de 10 pares de bases. El cebador inferior es el mismo oligonucleótido poliT anclado (para el caso del display diferencial) o el decámero arbitrario (para el caso de RAP-PCR) que se usaron alternativamente para hacer la transcripción reversa. Se ha estimado que los productos de PCR de 240 combinaciones particulares de pares de cebadores de display diferencial (por ejemplo todas las combinaciones de 20 oligonucleótidos arbitrarios y 12 oligonucleótidos anclados) representan estadísticamente a todos los ARNm originalmente presentes en la muestra. Los productos de PCR pueden ser marcados por incorporación de un nucleótido radiactivo o de una molécula que pueda ser detectada a través de su interacción con un anticuerpo conjugado a un sistema de detección (por ejemplo digoxigenina). También puede usarse un cebador unido a un fluoróforo u optarse por no marcar los fragmentos amplificados y usar luego una tinción con nitrato de plata para revelar



**Figura 1.**

los geles. La visualización de los productos de PCR se logra luego de la electroforesis en geles de poliacrilamida seguida de la apropiada generación y detección de imágenes, que son evaluadas comparando la intensidad relativa de las bandas producidas a partir de diferentes muestras experimentales. Los fragmentos que están presentes en una muestra y ausentes en

la otra, o aquellos que están presentes con diferentes intensidades relativas en los distintos tratamientos experimentales, representan potenciales transcritos de ARNm de expresión diferencial. Típicamente, las bandas son evaluadas sólo si amplifican en forma consistente en reacciones de PCR duplicadas para cada muestra (ver la Figura 2).



**Figura 2.**

La fase final del display diferencial consiste en la identificación de la secuencia del transcrito representado por el producto de PCR y la confirmación de que éste realmente se expresa en forma contrastante. Estas etapas se logran localizando físicamente y escindiendo la sección del gel de poliacrilamida que contiene al producto de interés. En la mayoría de los

casos debe realizarse un alineamiento de las imágenes de los fragmentos de amplificación con el gel de poliacrilamida seco. En los casos en que se utiliza tinción con nitrato de plata no es necesario realizar este paso, por lo cual la recuperación de la banda es mucho más eficiente. Los productos de PCR son purificados del gel y reamplificados. Para ello el fragmento

de poliacrilamida se procesa en fragmentos pequeños con un bísturí que se transfieren a una solución tampón adecuada para la elución de los fragmentos. Luego de una centrifugación, el sobrenadante se purifica con fenol/cloroformo y el ADN se precipita con etanol, se diluye en agua y se reamplifica por PCR.

Se han utilizado varias estrategias para confirmar la expresión diferencial de los transcritos, que incluyen el uso de los fragmentos como sondas de Northern, la siembra de los fragmentos en membranas para someterlos a análisis de Northern inverso y el clonado y secuenciación de los productos para diseñar cebadores específicos que permitan realizar PCR semicuantitativas, así como también la hibridación *in situ* de tejidos.

Dos ventajas importantes de los métodos de fingerprinting de ARN con respecto a los de hibridación substractiva son la capacidad de comparar muestras experimentales múltiples en forma simultánea y la de identificar genes que están regulados tanto positiva como negativamente en una muestra respecto de las otras. Sin embargo, los métodos de fingerprinting de ARN comparten con el SSH la limitación de que no son cuantitativos.

Uno de los problemas que suelen atribuirse a display diferencial es la amplificación de falsos positivos o sea fragmentos de genes que parecen estar diferencialmente expresados pero son luego considerados artefactos de PCR porque no pueden validarse por Northern blot o PCR en tiempo real. Es necesario dar una mirada más detallada a este punto. Es cierto que es imprescindible realizar las amplificaciones de DD PCR por duplicado o triplicado para asegurarse que la amplificación es reproducible y no se trata de una banda espuria. Pero una vez superado este punto de la reproducibilidad de la amplificación, hay que considerar que no siempre las diferencias detectadas con esta técnica pueden ser validadas por Northern o por PCR en tiempo real. Por ejemplo, cuando existe expresión alélica diferencial entre muestras, o cuando diferentes miembros de una misma familia génica son expresados diferencialmente, es posible que se detecten polimorfismos en los genes de DD, como consecuencia de que los primers se unan a sitios variables

de la secuencia. Sin embargo, es probable que distintos alelos o incluso diferentes miembros de una familia génica hibridicen en conjunto en el experimento de Northern, o que los diseños de cebadores para real time PCR no permitan diferenciarlos. También puede ocurrir que en los experimentos de DD se detecte una hebra antisentido sobreexpresada en una de las muestras. Si la hebra sentido está expresada en la otra, los experimentos de real time no permitirán la detección de la expresión diferencial. Estos son sólo algunos de los muchos casos posibles en los cuales el northern y la PCR en tiempo real no reproducirán los datos de DD, pero no precisamente porque el DD haya generado un falso positivo, sino simplemente porque estas técnicas están basadas en principios de detección diferentes. Muchos de los supuestos "falsos positivos" detectados por DD en el pasado (cuando no se tenía conciencia sobre la frecuencia de la expresión antisentido) pueden englobarse seguramente en estas categorías.

Otro problema es que estas técnicas deben aplicarse en general a muestras provenientes del mismo individuo sometido a distintas condiciones o sobre individuos genéticamente idénticos (por ejemplo isolíneas). Cuando se comparan los perfiles de ARNm de individuos diferentes genéticamente, pueden aparecer falsos positivos que son resultado de una amplificación diferencial debida a una variación en la secuencia de los transcritos más que a diferencias en su representación. Ese problema puede ser evitado utilizando estrategias de análisis de segregantes en grupo (BSA) aplicadas al estudio de perfiles de ARN. Finalmente, la investigación de todos los genes que potencialmente se expresan en forma diferencial requiere el uso de equipamiento de última generación y una inversión extensiva de tiempo y trabajo para detectar y confirmar la expresión diferencial de cada uno de los genes individuales.

#### **Polimorfismos en el largo de los fragmentos de ADNc amplificados (ADNc-AFLP®)**

La tecnología de AFLP®, generalmente utilizada para producir huellas genéticas de DNA genómico, puede ser aplicada también a pre-

paraciones de ADNc de doble hebra para obtener un perfil del transcriptoma. Los patrones obtenidos por esta técnica son una herramienta confiable y eficiente para la identificación de ARNm expresados diferencialmente. La técnica de ADNc-AFLP® detecta fragmentos de restricción de DNA por medio de amplificación por PCR. Comprende las siguientes etapas: i) generación de ADNc, ii) restricción del ADNc con dos endonucleasas específicas, preferiblemente una que reconoce sitios de 6 bases y otra que reconoce sitios de 4 bases, iii) ligación

de adaptadores de cadena doble a los extremos de los fragmentos de restricción, iv) amplificación de un subgrupo de los fragmentos de restricción usando dos cebadores complementarios al adaptador que llevan bases selectivas adicionales en el extremo 3' v) electroforesis de los fragmentos de restricción amplificados en geles de poliacrilamida, vi) visualización de las huellas digitales genéticas mediante el uso de autoradiografías, fosfoimágenes y otros métodos (ver Figura 3).

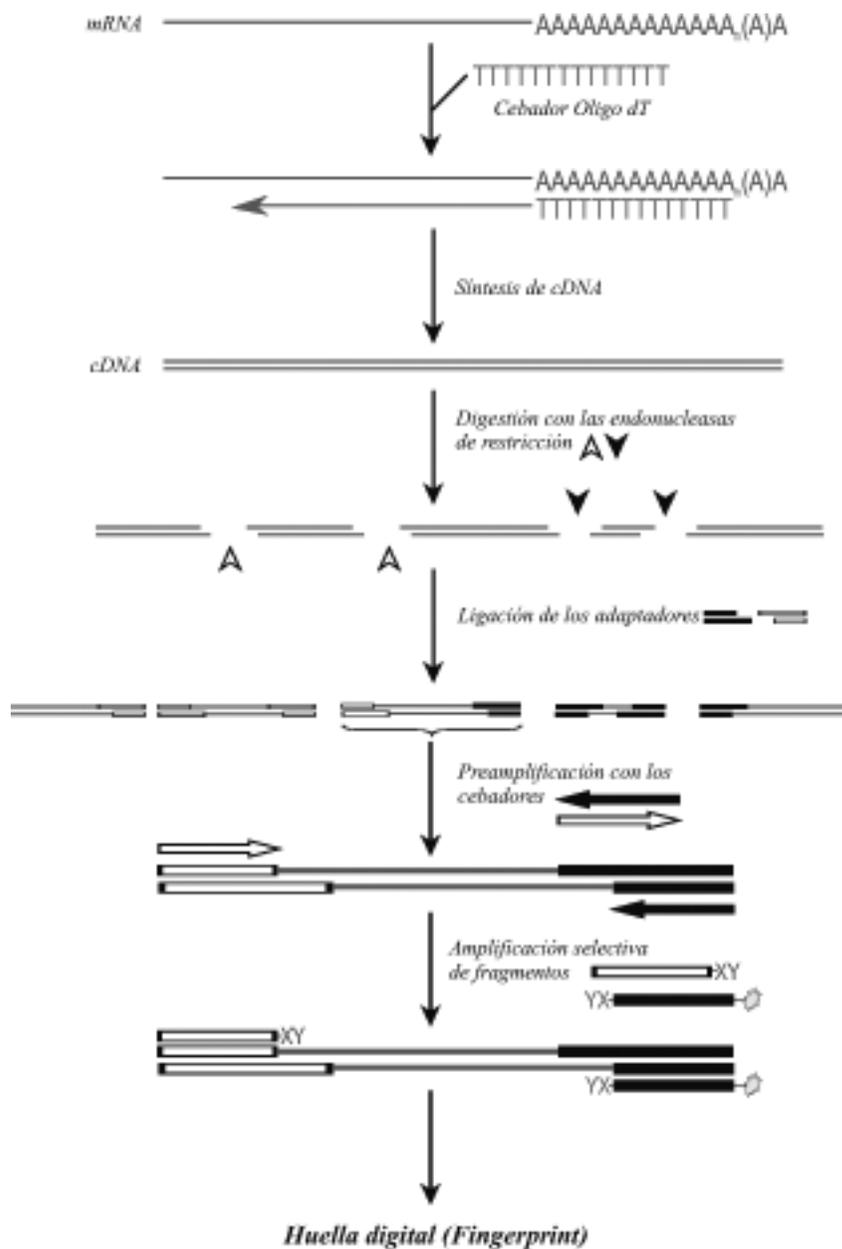


Figura 3.

Las mayores ventajas del uso de la tecnología de ADNc-AFLP® son: i) no se requiere información previa de secuencia; ii) se puede estudiar una fracción muy grande de todos los genes expresados, iii) es una técnica muy sensible que permite la detección de transcritos de baja abundancia, iv) los fragmentos son extraídos fácilmente de los geles y las secuencias correspondientes pueden determinarse sin necesidad de clones previos.

### **Etiquetas de secuencia expresadas (Expressed sequence tags, ESTs)**

La secuenciación exhaustiva de ESTs (Expressed Sequence Tags) es considerado principalmente un método para la caracterización de la expresión génica, a pesar de que la generación de ESTs resulta también importante para deducir la presencia de genes no predichos en ausencia de datos genómicos. Las ESTs son obtenidas a gran escala por secuenciación de una sola hebra de clones de ADNc (aproximadamente 500 pb) provenientes de bibliotecas que representan distintos tejidos o condiciones experimentales. Las ESTs representan descripciones parciales de las regiones que se transcriben de un genoma. Una cantidad creciente de centros de investigación han construido y secuenciado bibliotecas de ADNc en los últimos años. Esta tendencia se ve reflejada en el número de ESTs depositadas en la base de datos del NCBI (ESTdb), el que al 6 de Junio de 2008 es de 2994249 provenientes de 261 especies de plantas.

En teoría, la abundancia de una EST es directamente proporcional a la representación en número de copias de un transcripto en un tejido determinado. El proceso de generación de ESTs es relativamente lento y costoso, lo que hace difícil que se logre la saturación de una biblioteca. Teóricamente, si se secuencia el número suficiente de ESTs, este método permite analizar incluso aquellos genes que presentan niveles de expresión bajos, pero si el número de secuencias obtenidas es limitado conviene recurrir a técnicas de amplificación complementarias (ADNc-AFLP o DD) para detectar transcritos poco abundantes.

Las secuencias ESTs a menudo se generan a partir de bibliotecas de ADNc que han sido

normalizadas para ecualizar la abundancia de clones que corresponden a los diferentes transcritos. Los EST secuenciados a partir de estas últimas pueden ser comparados para identificar transcritos que se expresan en una biblioteca y están completamente ausentes en otra, pero si se pretende obtener datos cuantitativos exactos que describan abundancia relativa se debe recurrir indefectiblemente a bibliotecas de ADNcs no normalizadas. También existen productos comerciales que permiten la construcción de bibliotecas donde los transcritos han sido amplificados pero manteniendo la proporción relativa de unos respecto a otros, lo que facilita la comparación cuantitativa entre muestras.

Las ESTs requieren de varias etapas de procesamiento, ensamblado en grupos (contigs) y anotación para que se pueda extraer información biológica a partir de las mismas. Para esto, resulta muy importante el almacenamiento, la organización y la anotación de estas secuencias utilizando distintas herramientas informáticas (revisadas por Nagaraj et al., 2007).

### **Análisis serial de la expresión de genes (SAGE)**

El análisis serial de la expresión de genes (SAGE) consiste esencialmente en una versión acelerada de la secuenciación de ESTs. Está basada en el concepto de que una etiqueta corta de unas pocas bases es suficiente para identificar inequívocamente a un transcripto, siempre y cuando esté ubicada en una posición definida dentro de la secuencia del mensajero. Una etiqueta SAGE consiste típicamente en 9-14 bases de secuencia ubicadas por debajo del último sitio de reconocimiento de una endonucleasa específica en el transcripto blanco. Múltiples etiquetas SAGE se ligan juntas en un vector de clonado de manera que una reacción de secuenciación de 300 a 500 pb genera las secuencias de 20 a 30 etiquetas al mismo tiempo (ver Figura 4). Como cada etiqueta SAGE representa un único transcripto de ARNm, es posible obtener una visión general de todos los genes expresados en la muestra original. Las diferencias en la expresión de genes entre muestras experimentales pueden entonces ser identificadas comparando la abundancia



los casos se observa que entre el 50-70% de los fragmentos presentan similitud con genes o transcritos de función conocida, mientras que entre 30-50% no son similares a ninguna secuencia conocida. Es esperable que estos porcentajes que son usuales actualmente se modifiquen en el futuro a medida que aumente el número de genes caracterizados y la cantidad de etiquetas promedio secuenciadas por proyecto.

Se ha observado una especificidad relativamente alta de un fragmento de entre 9 y 14 pb para su asociación con transcritos conocidos en el caso de genomas relativamente simples. Sin embargo, la especificidad disminuye cuando se estudian genomas más complejos. Se han hecho intentos para mejorar la especificidad aumentando el largo de los fragmentos. LongSAGE aumenta el largo de los fragmentos a 21 pb y SuperSAGE a 26 pb, utilizando diferentes enzimas de etiquetado. Esta estrategia mejoró la especificidad de un dado fragmento para representar a un transcrito único (aunque no completamente) y también la probabilidad de que dicho fragmento se localice una única vez en el genoma. Sin embargo, el hecho de trabajar con fragmentos más largos aumenta los costos de secuenciación de la técnica. Por estos motivos, el usuario debe considerar cuidadosamente cuál es la longitud de los fragmentos a utilizar para cada caso particular.

Los fragmentos obtenidos mediante SAGE no se pueden utilizar directamente en búsquedas de homología con secuencias contenidas en bases de datos debido a que son muy cortos. Para poder asignar un gen a un dado fragmento es necesario contar con una base de datos de referencia construida previamente. Esta base de datos contiene información de fragmentos obtenidas *in silico* a partir de secuencias conocidas de transcritos o de genes de la especie en estudio. Si un fragmento aislado experimentalmente presenta homología con una secuencia presente en la base de datos SAGE de referencia, entonces se considera que el transcrito que lo originó proviene del gen que había sido asignado a esa secuencia. La confiabilidad de la asignación de genes utilizando estas bases de datos está influenciada por factores biológicos (por ejemplo, isoformas

del transcrito tales como las resultantes de splicing alternativo o SNPs presentes en la etiqueta pueden resultar en distintos fragmentos que en realidad se correspondan con un mismo gen), experimentales (artefactos introducidos en las secuencias como consecuencia de la técnica utilizada) y por redundancia de transcritos. Las bases de datos de referencia se construyen con secuencias que son altamente redundantes. Las mismas son agrupadas en función de la homología de secuencias, lo que puede originar que se encuentren transcritos que corresponden a un mismo gen en grupos distintos (como en el caso de que hubiera splicing alternativo) o que se agrupen transcritos correspondientes a genes distintos porque presentan un alto grado de homología de secuencia. Es importante mencionar que las bases de datos de referencia pueden ser incompletas, debido a que comúnmente se incluyen en las mismas secuencias de alta calidad y anotadas para aumentar la especificidad. Sin embargo, esta estrategia disminuye en gran medida el número de transcritos representados en la base de datos. Esto origina que muchos de los fragmentos obtenidos por SAGE sean clasificados como desconocidos incluso cuando sus transcritos correspondientes hayan sido previamente identificados. Además, un mismo fragmento puede presentar homología con más de un gen, lo que presenta un desafío adicional para la clasificación de los mismos. Por todos estos motivos, es necesario verificar la asignación de un fragmento mediante otras técnicas. Entre las más utilizadas se pueden mencionar el RACE (Rapid Amplification of ADNc Ends) y GLGI (Generation of Long ADNc from SAGE tags for Gene Identification).

Los fragmentos obtenidos por SAGE pueden organizarse en tres grupos: de alta, intermedia y baja abundancia. Sin embargo, no se ha analizado exhaustivamente aún si el número de copias de un fragmento obtenido por SAGE refleja exactamente en todos los casos la cantidad de ARNm mensajero real contenida en la muestra. Dado que el SAGE involucra numerosas amplificaciones por PCR, pueden originarse errores específicos para un dado fragmento. Además se ha observado que las secuencias obtenidas por esta técnica presentan un ma-

yor contenido de G+C, lo que sugiere una preferencia de amplificación que compromete la exactitud de la cuantificación.

Dos ventajas importantes del SAGE sobre la hibridación substractiva y el display diferencial es que los datos obtenidos son cuantitativos y acumulativos. Si se genera la suficiente cantidad de información de secuencia se obtienen perfiles exactos de la expresión de los transcritos, que describen la abundancia de todos los genes expresados en una célula o tejido. Existe una base de datos pública de expresión génica que ha sido creada para el almacenamiento y análisis de secuencias de SAGE cuyo poder continuará creciendo a medida que se contribuya con más información. Otra característica de los datos de SAGE es que pueden complementar otros de ESTs generados a partir de bibliotecas de ADNc normalizadas o substraídas. Los ESTs brindan información de secuencia mientras que el SAGE provee datos cuantitativos describiendo la abundancia de los transcritos. Finalmente, a pesar de que el SAGE requiere capacidad de secuenciación de última generación, la cantidad de datos que aporta puede ser 20 veces mayor que el que posibilita la misma cantidad de secuenciación EST.

### **Hibridación de microarreglos**

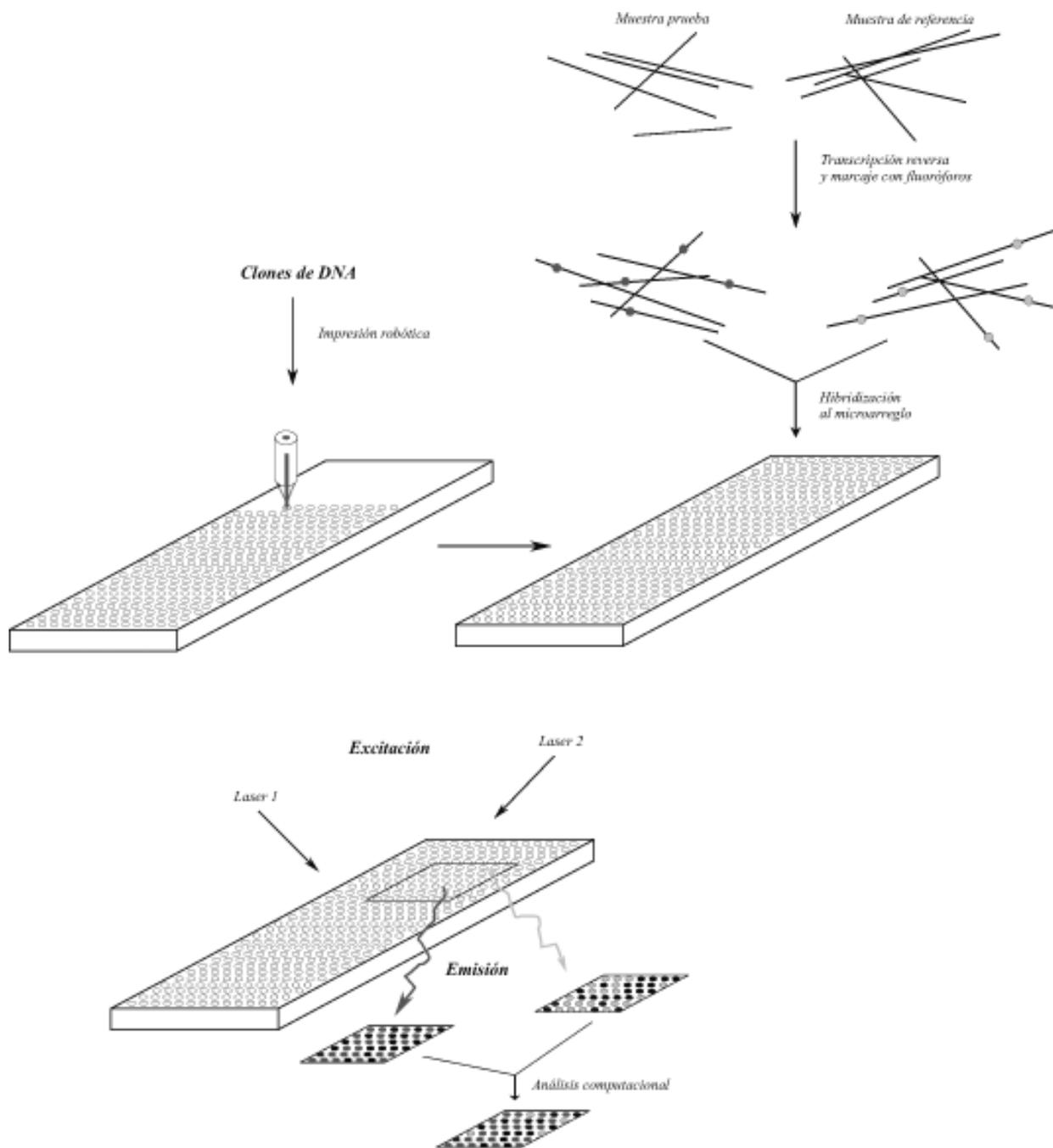
Los microarreglos han revolucionado la biología molecular. Esta técnica utiliza cientos a miles de sondas altamente organizadas sobre una superficie sólida para interrogar de manera simultánea a todas las moléculas de ARN (definidas como blancos) contenidas en una muestra biológica. Las muestras a ser analizadas se marcan fluorescentemente o radiactivamente y se aplican al microarreglo. Los ácidos nucleicos marcados hibridan con las moléculas de ADN complementarias que se encuentran inmovilizadas en el microarreglo. La fuerza de la señal fluorescente o radiactiva capturada por las sondas de ADN en el arreglo es detectada y digitalizada para su cuantificación.

Los microarreglos fueron originalmente diseñados para identificar diferencias de expresión génica en dos muestras distintas basadas en las cantidades relativas de blancos unidos a una sonda determinada en el arreglo. La utili-

dad de la técnica se ve reflejada en su amplia variedad de aplicaciones que abarca desde el análisis de expresión génica hasta estudios de ADN genómico (como Comparative Genomic Hybridization - CGH- y Chromatin Immunopurification microarrays - conocidos como ChIP-chips). Sin embargo, en este capítulo sólo describiremos los microarreglos de ADNc y de oligonucleótidos, ya que son los más utilizados para el análisis del transcriptoma.

Los microarreglos de ADN pueden ser creados de dos maneras: fijando las moléculas de ácidos nucleicos sobre una superficie sólida o por síntesis *in situ* de oligonucleótidos. En el primer caso las moléculas de ADN pueden ser ADNcs amplificados por PCR, oligonucleótidos o fragmentos de ADN genómicos clonados en BACs (Bacterial Artificial Chromosomes).

La utilización de clones de ADNc amplificados por PCR como sondas es el método de elección cuando se interrogan muestras de organismos cuyo genoma no está secuenciado. Estos microarreglos son generalmente fabricados utilizando robots. Los mismos contienen un cabezal de puntas que es sumergido en soluciones de sondas de ADN disueltas en el buffer correspondiente y que luego son depositadas sobre un sustrato de vidrio modificado químicamente. Las puntas se cargan de sonda por capilaridad y la tensión superficial entre el buffer y el vidrio actúa para descargar las sondas. Las perturbaciones en la estructura de las descargas son frecuentes y son una fuente significativa de variación de la señal. Si bien existen otras alternativas, la utilización de robots es la más popular debido a su disponibilidad, alta flexibilidad y bajos costos. Los sustratos inertes sobre los cuales se depositan las sondas son variados. Los portaobjetos de vidrio cubiertos con poli-L-lisina fueron gradualmente reemplazados por sustratos con aminosilano comerciales debido a su mayor consistencia en cuanto a la morfología de las descargas. Para este tipo de químicas, las sondas son inicialmente unidas por atracción electrostática y luego fijadas por ligación cruzada (cross-linking) al sustrato mediante radiación UV o calor. Posteriormente se introdujeron sustratos reactivos y sondas modificadas que permiten una mayor discriminación de secuencia debido



**Figura 5.**

a que hay un único punto de unión y una mayor retención de la sonda al sustrato. También se han desarrollado otros sustratos para producir señales más intensas y que demanden menores cantidades de sondas. Para lograrlo se aumentó la cantidad de grupos reactivos sobre la superficie. Esto se puede lograr cubriendo los portaobjetos con dendrímeros, mezclas de epoxisilano y aminosilano o con polímeros que se auto-absorben.

Originalmente, las sondas de ADN eran disueltas en soluciones con alto contenido de sales. Luego se introdujo la utilización de detergentes para mejorar la morfología de las descargas. Sin embargo, algunos investigadores informaron que la presencia de detergentes resulta en un mayor arrastre de la sonda y una mayor variabilidad entre una deposición y la siguiente. La utilización de soluciones de descarga conteniendo agentes higroscópicos

permitió mejorar la estructura de los puntos y aumentar las intensidades de señal obtenidas. La desventaja de la utilización de aditivos higroscópicos es que producen aumentos en el área ocupada por la sonda, sin embargo, este efecto se puede reducir disminuyendo la velocidad del robot o aplicando micro-vibraciones de las puntas. Disminuir la temperatura de la cámara del robot donde se generan los microarreglos y utilizar soluciones hidrofóbicas contribuye también a reducir el tamaño promedio de las descargas. La variedad de soluciones de descarga que existe sugiere que no hay un tipo de solución que resulte ideal para todas las aplicaciones y que la elección de la misma depende en gran medida del tipo de microarreglo que se quiere construir. Las puntas de los cabezales que se utilizan son normalmente de acero inoxidable o de titanio y contienen un capilar generado mediante descargas eléctricas o la utilización de un láser. Se observa una variación sistemática en la extensión de la descarga, porque ésta es una función del formato de la punta, y existe variabilidad durante la fabricación de las mismas. Se han desarrollado nuevas tecnologías para solucionar este problema como por ejemplo la generación de puntas de cerámica más uniformes.

La evaporación durante el proceso de impresión de los microarreglos es la principal causa de variabilidad. La evaporación reduce el volumen de solvente en las microplacas de 384 pocillos y dentro del reservorio capilar de las puntas, lo que produce un aumento de concentración de las sondas y altera las características de la solución de descarga. Como consecuencia de esto, las propiedades del punto sembrado y la distribución de la sonda de ADN dentro del área del mismo se modifican. Este problema se puede solucionar en parte regulando las condiciones de humedad y temperatura dentro de la cámara del robot donde se realiza la impresión de los microarreglos. Dado que existen numerosas interacciones entre los distintos componentes del proceso, no es sencillo predecir en forma exacta cómo va a resultar la combinación entre una solución de siembra y un sustrato determinados. Sin embargo, se están realizando constantemente estudios para examinar este proceso que, suma-

dos al desarrollo de nuevas tecnologías, van a permitir reducir los niveles de variabilidad sistemáticos resultantes de la fabricación de los microarreglos.

La mayoría de los laboratorios ensayan la calidad de sus microarreglos antes de realizar los experimentos de hibridación con las muestras biológicas de interés. Generalmente se controla la bondad de la impresión y se detectan errores tales como puntos que no están bien formados, aquellos de tamaños mayores a los deseados o faltantes. Los puntos pueden ser visualizados utilizando tinturas fluorescentes de unión a ADN. También se pueden emplear métodos más precisos como son la hibridación con mezclas de oligonucleótidos al azar de 9-mer marcados fluorescentemente o con oligonucleótidos marcados que hibridizan con secuencias marcadoras cortas que están presentes en todas las sondas de ADN del microarreglo. Para controlar la calidad de la producción de microarreglos es recomendable desarrollar, documentar y emplear protocolos estándar (Standard Operating Procedures-SOPs). Se pueden encontrar ejemplos de los mismos en la página <http://www.flychip.org.uk>. Esta práctica no solamente aumenta la consistencia entre experimentos dentro de un laboratorio, sino que facilita la transferencia de información entre laboratorios.

El segundo método posible para la construcción de microarreglos es la síntesis de oligonucleótidos *in situ*, inicialmente desarrollada por Affymetrix (Affymetrix, Santa Clara, EEUU). Este método usa una conjunción de fotolitografía y de química combinatoria que resulta en la síntesis de chips de ADN de alta densidad. También NimbleGen (NimbleGen, Madison, EEUU), Xeotron (Xeotron, Houston, EEUU) y Febit (Febit, Mannheim, Alemania) han desarrollado otros procesos para lograr la síntesis de oligonucleótidos *in situ*. Todos estos procesos se basan en la utilización de radiación UV. En cambio Combimatrix (Combimatrix, Mukilteo, EEUU) usa electroquímica localizada. En el caso de Affymetrix se logra la síntesis *in situ* mediante fotoactivación y desprotección de ácidos nucleicos. Se utilizan fotomáscaras para dirigir la radiación UV a sectores específicos del microarreglo, de modo de lograr

la síntesis en forma selectiva. En el caso de NimbleGen, Xeotron y Febit, todos ellos usan métodos que no requieren la utilización de fotomáscaras. En este caso, la radiación UV es dirigida mediante miles de espejos regulables y el ADN es sintetizado utilizando nucleótidos modificados con fosfoamidita. En el caso de CombiMatrix, numerosos microelectrodos son utilizados para sintetizar distintas moléculas de ADN en paralelo. La reacción química para la síntesis de oligonucleótidos ocurre en el área que rodea a cada microelectrodo a la que se suele denominar “vaso virtual” (virtual flask).

Una estrategia para estudiar la expresión génica de todo un genoma es la detección de la actividad transcripcional completa de los cromosomas usando “arreglos solapados o en tejado” (tiling arrays). En este caso, en lugar de utilizar sondas específicas para cada gen, el genoma completo incluyendo regiones intergénicas está representado en el arreglo. Además de detectar transcritos estos chips permiten realizar hibridaciones comparativas de genomas para detectar deleciones y polimorfismos, estudiar los perfiles de metilación y analizar muestras de inmunoprecipitación de cromatina. Affymetrix desarrolló los primeros arreglos de este tipo para *Arabidopsis*. Los estudios realizados utilizando arreglos de tipo “tejado” permiten la identificación de numerosos sitios transcripcionalmente activos que no habían sido detectados mediante algoritmos predictivos y/o en colecciones de ADNc. Muchos de estos nuevos transcritos son expresados a partir de la hebra complementaria y en antisen-tido con respecto a transcritos anotados previamente. O sea, mientras que en determinados tejidos y/o condiciones se expresa la hebra sentido, en otros se transcribe la antisentido. Sumado a esto, también se demostró que las regiones centroméricas presentan actividad transcripcional en ambos sentidos resultando esta observación sorprendente, ya que hasta hace muy poco tiempo atrás se suponía que en estas regiones no había transcripción activa. Esta metodología permite asimismo el estudio del metiloma, de las regiones de ADN regulatorias, del procesamiento alternativo de ARN y del descubrimiento de polimorfismos de

secuencia. Por lo tanto, los estudios de expresión usando arreglos de tipo “tejado” permiten obtener una visión mucho más completa del transcriptoma y de su relación con el genoma.

La necesidad de descripciones precisas de experimentos de microarreglos para permitir el almacenamiento de los datos experimentales y su análisis posterior ha originado la creación del “Minimum Information About a Microarray Experiment (MIAME) Standard” (MIAME Standard). Al igual que con las secuencias de ADN o con las estructuras de proteínas, es cada vez más común el requerimiento de enviar los resultados a bases de datos de este tipo antes de su publicación. Estos estándares son necesarios para permitir que la información esté públicamente disponible en un formato único, y también para uniformar la nomenclatura empleada lo que hace posible la integración y comparación de resultados obtenidos de distintas fuentes. MIAME administra descripciones técnicas detalladas (plataforma utilizada, métodos de marcación e hibridación, medición de datos y diseño del arreglo), pero no provee información relevante específica del organismo en estudio que es necesaria para comprender el experimento planteado. Por este motivo, se diseñó la base MIAME/Plant, la que incluye detalles biológicos para describir experimentos de microarreglos realizados en plantas.

El análisis de los datos originados en un microarreglo resulta mucho más laborioso que su generación. En el mismo se consideran la hipótesis de trabajo y los objetivos experimentales para lograr identificar a los genes candidatos en las condiciones de estudio. El análisis puede dividirse en dos fases: 1) el procesamiento de datos y 2) la identificación de genes de interés y la interpretación biológica de los resultados. El procesamiento de los datos incluye la determinación del nivel de ruido, la cuantificación de las señales fluorescentes, el examen de la calidad de los datos, la eliminación de los errores técnicos y la normalización de los datos. La identificación de los genes candidatos incluye la reducción del volumen de la información a un nivel manejable y más fácil de visualizar, el reconocimiento de patrones de expresión específicos y la interpretación del significado biológico de los mismos.

## Conclusiones

Los métodos descritos en este capítulo son tecnologías eficaces que expanden los estudios de expresión desde los genes únicos hasta el nivel del transcriptoma completo. Ninguno de estos procedimientos *per se* es óptimo para todas las aplicaciones y la mejor estrategia para situaciones específicas puede ser crear combinaciones de varios de ellos. El continuo refinamiento de las técnicas de transcriptómica y de la bioinformática relacionada proporcionará a los investigadores nuevas herramientas para estudiar la expresión de la información hereditaria y lograr un mejor entendimiento sobre el control genético de los procesos fisiológicos.

Cada plataforma empleada actualmente para estudiar el transcriptoma tiene sus ventajas y desventajas. Por ejemplo, los microarreglos son una técnica de alto rendimiento pero sólo se pueden aplicar a transcritos de secuencia conocida. Los métodos de ADNC- AFLP y DD son abiertos, es decir permiten detectar genes nuevos y son especialmente útiles cuando se trabaja fuera de los modelos, pero son cualitativos o semicuantitativos y requieren de un complemento de PCR en tiempo real para validar y cuantificar las diferencias de expresión.

Las colecciones de ESTs presentan alta especificidad para los transcritos en estudio pero baja sensibilidad para aquellos de escasa abundancia. La técnica de SAGE disminuye el largo secuenciado por transcripto lo que aumenta la velocidad de relevamiento, pero esta simplicidad resulta en una alta sensibilidad pero a costa de una menor especificidad. Una comparación de las características de cada técnica se presenta en la Tabla 1.

La combinación de la información generada por la transcriptómica (nivel y localización espacio/temporal de la expresión génica) y la de los proyectos de secuenciación de genomas (secuencias completas de los genes y sus promotores) permite realizar asociaciones de expresión y homología estructural que en muchos casos facilitan la inferencia de la posible función de los genes identificados. Por supuesto esto constituye una instancia inicial en la definición de la función de cada gen. Para determinar su rol preciso deben hacerse estudios particulares en general dirigidos a bloquear o aumentar su expresión por ingeniería genética y a modificar estructuralmente la molécula de manera de alterar la actividad de los diferentes dominios de la proteína que codifica.

**Tabla 1:** Comparación de los principales métodos de relevamiento del transcriptoma.

Características	SAGE/ESTs	Microarreglos	DD/cDNA-AFLP	Hibridización sustractiva
Detecta transcritos conocidos	Sí	Sí	Sí	Sí
Detecta transcritos desconocidos	Sí	No	Sí	Sí
Detecta transcritos con splicing alternativo	Sí	Sí o no	Sí	No
Detecta transcritos antisentido	Sí	Sí o no	Sí	Sí
Cuantificación	Absoluta	Relativa	Relativa	Relativa
Sensibilidad	Alta	Moderada	Alta	Alta
Especificidad	Moderada	Alta	Moderada	Moderada
Reproducibilidad	Buena para transcritos abundantes	Buena para datos obtenidos con la misma plataforma	Requiere realización de duplicados o triplicados	Buena para transcritos abundantes
Comparable entre sí	Buena para transcritos abundantes	Buena para transcritos abundantes	Buena para transcritos abundantes	Buena para transcritos abundantes
Costo	Más alto que el de microarreglos	5-10 veces menor que el de SAGE	Muy bajo	Muy bajo

Adaptado de Wang 2006 understanding SAGE data, Trends in Genetics.

Para maximizar el potencial de la transcrip-tómica es necesario un cambio de paradigma respecto a la interpretación de los resultados obtenidos en este tipo de experimentos. Este cambio de paradigma implica comprender que los datos de expresión no están libres de errores y aceptar que la variabilidad biológica espacio-temporal es una parte inherente y cuantificable de este tipo de experimentos. A pesar de que los continuos cambios en la expresión de genes en tiempo y espacio colocan al investigador en un escenario de arenas movedizas, es mucha y muy relevante la información que puede obtenerse de este tipo de estudios. Esta transición en la interpretación de los resultados será más fácil si se delinea claramente la diferencia entre proponer una función biológica a partir de simples datos de expresión y postular una hipótesis que debe luego validarse en base a datos experimentales concretos.

### Bibliografía Recomendada

- Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Poly-meropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science (Wash DC)* 252:1651–1666.
- Affymetrix: <http://www.affymetrix.com>
- Combimatrix: <http://www.combimatrix.com>
- Davis, M. M., D. I. Cohen, E. A. Nielsen, M. Steinmetz, W. E. Paul, L. Hood. 1984. Cell-type-specific ADNc probes and the murine 1 region: the localization and orientation of Ad. *Proc. Natl. Acad. Sci. USA* 81:2194–2198.
- Diatchenko, L., Y.-F. C. Lau, A. P. Campbell, A. Chenchik, F. Moqa-dam, B. Huang, S. Lukyanov, K. Pukyanov, N. Gurskaya, E. D. Sverdlov, P. D. Siebert. 1996. Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific ADNc probes and libraries. *Proc. Natl. Acad. Sci. USA* 93:6025–6030.
- Febit: <http://www.febit.com>
- Gregory B.D., Yazaki J. and Ecker J.R. 2008. Utilizing tiling microarrays for whole-genome analysis in plants. *The Plant Journal*, 53, 636-644.
- Gurskaya, N. G., L. Diatchenko, A. Chenchik, P. D. Siebert, G. L. Khaspekov, K. A. Lukyanov, L. L. Vagner, O. D. Ermolaeva, S. A. Lukyanov, E. D. Sverdlov. 1996. Equalizing ADNc subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohem-agglutinin and phorbol 12-myristate 13-acetate. *Anal. Biochem.* 240:90–97.
- Liang, P., A. B. Pardee. 1992. Differential display of eukaryotic messenger ARN by means of the polymerase chain reaction. *Science (Wash DC)* 257:967–971.
- Meyers B.C., Galbraith, D.W., Nelson T. and Agrawal, V. 2004. Methods for transcriptional profiling in plants. Be fruitful and replicate. *Plant Physiol.*, 135, 637-652.
- MIAME Standard: <http://www.mged.org/Workgroups/MIAME/miame.html>
- Moody D. E. 2001. Genomics techniques: an overview of methods for the study of gene expression. *J. Anim. Sci.* 79 (E. Suppl.): E128-E135.
- Nagaraj S.H., Gasser R.B. and Ranganathan S. 2007. A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform.*, 8, 6-21.
- NimbleGen: <http://www.nimblegen.com>
- Okubo, K., N. Hori, R. Matoba, T. Niyama, A. Fukushima, Y. Kojima, K. Matsubara. 1992. Large scale ADNc sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat. Genet.* 2:173–179.
- Rensink W.A. and Buell C.R. 2005. Microarray expression profiling resources for plant genomics. *Trends in Plant Science*, 10, 603-609.
- Sargent T. D., I. B. Dawid. 1983. Differential Gene expression in the gastrula of *Xenopus laevis*. *Science (Wash DC)* 222: 135-139.
- Schena M., D. Shalon, R. W. Davis, P. O. Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (Wash DC)* 270:467–470.
- Velculescu, V. E., L. Zhang, B. Vogelstein, K. W. Kinzler. 1995. Serial analysis of gene expression. *Science (Wash DC)*270:484–487.
- Venkatasubbarao S. 2004. Microarrays – status and prospects. *Trends in Biotechnology*, 22, 630-637.
- Welsh, J., K. Chada, S. S. Dalal, R. Cheng, D. Ralph, M. McClelland. 1992. Arbitrarily primed PCR fingerprinting of ARN. *Nucl. Acids Res.* 20:4965–4970.
- Xeotron: <http://www.xeotron.com>
- Zimmermann P., Schildknecht B., Craigon D., Garcia-Hernandez M., Gruissem W., May S., Mukherjee G., Parkinson H., Rhee S., Wagner U. and Hennig L. 2006. MIAME/Plant – adding value to plant microarray experiments. *Plant methods*, 2:1.

# I CAPÍTULO 9

## Proteómica

Paula Casati y María F. Drincovich

### 1. Introducción

La proteómica, entendida como el estudio del patrón proteico de una célula u órgano, se encuentra ya en su segunda década como campo de estudio. Durante la primera década, la mayor parte de los estudios se realizaron principalmente mediante el uso de geles en dos dimensiones seguidos de técnicas de tinción convencionales. Si bien estos métodos continúan siendo utilizados de manera importante, durante esta segunda década otros métodos más sensibles y que además presentan mayor reproducibilidad han comenzado a ser utilizados. Estas técnicas han sido clasificadas como "proteómica cualitativa". Muchos de estos métodos están siendo aplicados en experimentos de biología vegetal. Por ejemplo, la electroforesis bidimensional DIGE (del inglés, differential gel electrophoresis) y el marcado isotópico de proteínas y péptidos, son técnicas utilizadas en la actualidad para el estudio de diversos proteomas. En el presente capítulo se describen primero los métodos utilizados en proteómica en la actualidad, seguido de una descripción de las aplicaciones de estas técnicas en la biología de plantas.

### 2. Electroforesis en dos dimensiones

El procedimiento más utilizado para los estudios de proteómica es la separación electroforética por isoelectroenfoque (IEF) de las proteínas seguido de una segunda separación en gel de poliacrilamida en presencia de SDS (SDS-PAGE). Esta técnica se denomina comúnmente electroforesis bidimensional (2D), donde las proteínas se separan en base a su carga y masa molecular. Los avances recientes en la resolución del IEF gracias al uso de geles comerciales en tiras han aumentado la reproducibilidad y la resolución de los geles 2D en proteómica. En particular, la disponibilidad de tiras comerciales de diferentes tamaños y rangos de pH permiten un análisis detallado del proteoma.

Luego de la separación electroforética, se lleva a cabo la tinción de los geles y el análisis de las imágenes para visualizar y cuantificar cada proteína. Los métodos de tinción utilizados van desde los más convencionales, como la tinción con azul de Coomassie o nitrato de plata, hasta métodos más sensibles como la tinción fluorescente. Por otro lado han sido desarrollados recientemente varios compuestos fluorescentes que permiten múltiples tinciones de un mismo gel, técnica que permite la tinción y visualización de diferentes grupos de proteínas o diferentes subproteomas (por ejemplo, el fosfoproteoma o el glicoproteoma) de manera secuencial en un mismo gel. La separación en geles bidimensionales puede dar lugar a un mapa proteico complejo, y algunas veces la reproducibilidad de estos geles puede ser problemática debido a la naturaleza diversa de las distintas proteínas, sumado a la tarea de identificar el mismo punto proteico en distintos geles. Esto se debe a que cuando se realizan estudios comparativos es necesario analizar múltiples geles que pueden contener 1000 puntos proteicos cada uno. Así, a pesar del uso de programas avanzados para el análisis de geles 2D, se requiere una validación manual final. Existen distintos paquetes de computación para el análisis de las imágenes que permiten asistir con la substracción del ruido de base, la detección de los puntos proteicos, el ajuste de las imágenes, la detección e identificación de puntos entre geles, y la cuantificación mediante un análisis estadístico. Los geles bidimensionales son una herramienta para la resolución y visualización de isoformas proteicas, productos de degradación y modificaciones postraduccionales; además de permitir el cálculo empírico de la masa molecular, punto isoelectrónico y niveles de expresión. Las limitaciones de la técnica son la no automatización, la dificultad para resolver proteínas de membrana y proteínas muy básicas o ácidas.

Una técnica que aumenta la sensibilidad y disminuye la variabilidad que existe entre distintos geles es la llamada electroforesis diferencial en gel o DIGE. En esta técnica, las muestras proteicas se preincuban con compuestos fluorescentes activos que se unen a residuos de Lys o Cys que pueden ser utilizados para

cuantificar la abundancia de cada proteína en una población. La mayor ventaja de esta técnica es la posibilidad de correr dos muestras distintas en un mismo gel, siempre y cuando las muestras se hayan marcado con marcadores fluorescentes con distintos espectros de emisión. Esto disminuye los problemas que se generan durante la identificación de los puntos y la cuantificación. En teoría, la cantidad máxima de muestras que se pueden separar en un mismo gel se limita al número de compuestos fluorescentes que se encuentren disponibles y que no muestren espectros de emisión que se superpongan, aunque en general se utilizan sólo tres: dos para el marcado de las muestras a comparar, y el tercero se utiliza para marcar un control interno. La imagen de los geles es generada utilizando un escáner de fluorescencia capaz de adquirir las imágenes provenientes de cada uno de los fluorescentes utilizados. Esta técnica es muy sensible, ya que permite resolver alrededor de 1000 puntos proteicos con solo 50 µg de proteínas sembradas, por lo que el límite de detección es del orden de pica a femtomolar de proteína en cada punto.

Una ventaja de realizar estudios de proteómica mediante separación en geles bidimensionales es que estos experimentos pueden ser realizados utilizando muestras de cualquier especie vegetal, sin necesidad de que el genoma de la especie en estudio se encuentre secuenciada. Además, esta técnica es ideal para experimentos en donde se observen pocos cambios en los patrones proteicos.

Una alternativa a las separaciones 2D por IEF/SDS-PAGE es la técnica de separación en electroforesis nativa en presencia de azul de Coomassie (BN-PAGE, del inglés "blue-native"). Esta técnica es utilizada para la separación de proteínas hidrofóbicas y proteínas en su estado nativo, ya que éstas no se separan correctamente en los geles IEF/SDS-PAGE. Esta técnica se basa en la integración de una carga neta negativa en las proteínas y complejos proteicos por adición del colorante azul de Coomassie, y permite la separación en condiciones nativas en la primera dimensión. En la segunda dimensión se realiza un SDS-PAGE, donde los complejos proteicos que migraron juntos en la primera dimensión se separan en

sus subunidades, que se observan en filas verticales de puntos proteicos en los geles 2D. Esta técnica complementa a los geles 2D IEF/SDS-PAGE.

### **3. Cuantificación por espectrometría de masa (MS)**

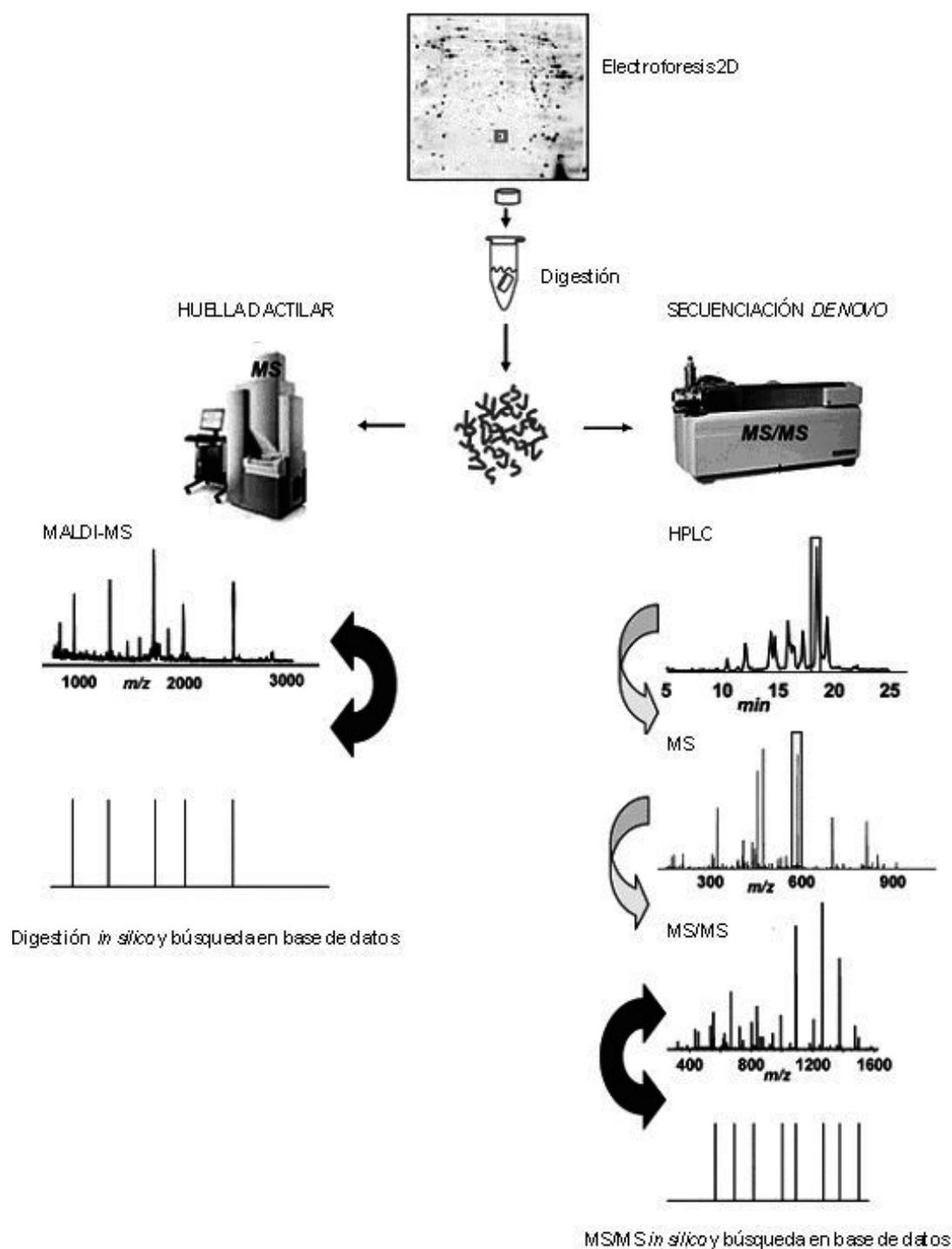
Luego de la selección de las proteínas de interés en un gel 2D, éstas pueden ser identificadas por espectrometría de masas. Básicamente, luego de la digestión de las proteínas cortadas del gel con una proteasa (típicamente tripsina), la mezcla compleja de péptidos es analizada. Existen distintos tipos de espectrometría de masas, que utilizan distintos métodos para la ionización de las muestras. Por un lado, la espectrometría de masa por desorción por ionización mediante láser asistida por una matriz (MALDI-TOF, del inglés matrix-assisted laser desorption ionisation (MALDI))-tiempo de vuelo (time-of-flight (TOF)), se utiliza principalmente para realizar huellas dactilares de péptidos y proteínas. Por otro lado, la espectrometría en tandem por electropulverizado (ESI), usualmente se acopla a una separación previa en un sistema de cromatografía de alta presión (HPLC) y se utiliza para la elucidación de la secuencia e identificación de la proteína correspondiente (Figura 1).

En estos momentos, existe una actualización constante en los distintos tipos de espectrómetros de masa que se encuentran disponibles en el mercado, y la sensibilidad, selectividad y precisión de los equipos son cada vez mayores. Los principios de esta técnica pueden ser resumidos en cuatro funciones del equipo:

1-Ionización: la molécula de interés sufre un cambio en la carga neta, formándose un anión o un protón. Dependiendo del método de ionización utilizado, es posible que se generen iones fragmentados. En el caso de usos en proteómica, los métodos de ionización utilizados son MALDI y electrospray/microspray/nanospray.

2-Separación: las especies iónicas son separadas de acuerdo a su relación masa/carga ( $m/z$ ),

3-Medida de la masa: las masas son asignadas según las medidas de algún parámetro físico.



**Figura 1. Técnicas de MS para identificación de proteínas a partir de geles bidimensionales.** Los puntos de interés del gel 2D son cortados y colocados en tubos. La proteína del bloque de acrilamida es digerida con una proteasa, y los productos obtenidos son analizados por distintas técnicas para su identificación. Adaptado de Nesatyy and Suter (2007).

4-Medida de la abundancia: la medida de la abundancia de cada ion se realiza en base a la altura o área del pico en el espectro, permitiendo una semicuantificación o cuantificación de cada ion.

### 3.1. MALDI-TOF y huella dactilar de péptidos:

En esta técnica, los péptidos generados luego de la digestión con tripsina son mezclados a una matriz en solución, que está formada por una solución concentrada o saturada de un compuesto de bajo peso molecular ácido y que absorbe el UV. En el caso de péptidos, el compuesto más utilizado es el ácido  $\alpha$ -ciano-4-hidroxicinámico. Aproximadamente 1  $\mu$ L de la mezcla muestra/matriz es colocada en una placa para MALDI y se permite co-cristalizar en un único punto. Como la matriz se coloca en exceso con respecto a la muestra y absorbe UV, cuando se irradia con un láser de UV esta energía es absorbida por la matriz y pasada a la muestra de manera que se ioniza pero no se fragmenta. Estos iones son luego acelerados por el espectrómetro de masa: los iones más livianos viajan más rápido que los pesados, por lo que el tiempo en el que llegan al detector puede ser utilizado para determinar la relación masa/carga de cada ion, obteniéndose un espectro que contiene todos los iones acelerados, y se obtiene una lista de las masas de los picos que presentan mayor intensidad. La lista de picos es enviado a una base de datos que compara las masas de los picos de los espectros con los teóricos de las proteínas digeridas, y produce una lista de las posibles identidades utilizando un sistema de puntaje.

### 3.2. Espectrometría de masas en tandem por electropulverizado.

Este proceso analiza los péptidos generados luego de la digestión con tripsina. En este caso, el espectrómetro de masas está acoplado a un equipo de HPLC. Durante la ionización por electropulverizado, los péptidos se separan previamente por HPLC (usualmente son utilizadas columnas de C18) y luego entran a una fuente de ionización. Un voltaje alto es aplicado a una aguja por donde pasan las muestras que llegan como gotas, produciendo una carga en la superficie. En la gota se genera una repul-

sión electrostática que hace que se liberen iones gaseosos que son acelerados en el espectrómetro de masas. Cuando los péptidos entran al espectrómetro de masa, algunos iones peptídicos son seleccionados y fragmentados por colisión con un gas para producir un número de fragmentos iónicos. Las relaciones m/z de estos fragmentos son registrados para dar el espectro de fragmentación de ese péptido. Un péptido se puede considerar como una cadena aminoacídica que se puede fragmentar en cualquier punto; sin embargo la fragmentación ocurre preferencialmente entre la unión de N y C entre aminoácidos, por lo que la secuencia de un péptido puede determinarse por el patrón de picos que se obtienen del espectro de fragmentación. El espectrómetro de masa realiza el registro de todos los picos de fragmentación, y luego de completada la separación y espectrometría de masa de todos los péptidos, los espectros son enviados a una base de datos para la identificación de la proteína en estudio por comparación al patrón de ionización teórico de los péptidos de la proteína.

### 3.3. Bases de datos para la identificación de proteínas:

Existen numerosas bases de datos para la identificación y caracterización de proteínas luego de la espectrometría de masas, búsqueda de semejanzas, patrones de fragmentación y perfiles, predicción de sitios de modificación postraducciona, etc. Entre ellos se encuentran MASCOT, ProteinProspector, ExPASy Proteomics. Muchos de estos sitios son accesibles luego de una suscripción, mientras que otros están disponibles en la web a todos los usuarios interesados. La tabla 1 muestra las páginas correspondientes de algunos de los sitios.

## 4. Alternativas a los geles bidimensionales

Algunas alternativas al uso de geles en dos dimensiones son los métodos basados en espectrometría de masa para comparar la abundancia de péptidos. Luego de la digestión de las proteínas (sin que se realice ninguna separación previa), la mezcla compleja de péptidos se separa por cromatografía previo al análisis por MS. La espectrometría de masas

**Tabla 1.**

Páginas disponibles correspondientes a bases de datos para identificación de proteínas para el análisis posterior a la espectrometría de masas

ExPASy Proteomics	<a href="http://us.expasy.org/tools">http://us.expasy.org/tools</a>
MASCOT	<a href="http://www.matrixscience.com">http://www.matrixscience.com</a>
ProSight PTM	<a href="http://proSightptm.scs.uiuc.edu">http://proSightptm.scs.uiuc.edu</a>
Protein Prospector	<a href="http://prospector.ucsf.edu">http://prospector.ucsf.edu</a>
GPM	<a href="http://thegpm.org">http://thegpm.org</a>
PROWL	<a href="http://prowl.rockefeller.edu">http://prowl.rockefeller.edu</a>

puede ser corrida de manera independiente o bien acoplada a la separación por HPLC. Los péptidos producidos por la digestión de las proteínas totales son ionizados para adquirir el espectro de MS de todos los iones presentes en la mezcla. Algunos péptidos son seleccionados y son fragmentados individualmente para obtener un segundo espectro MS (MS/MS) que permite obtener la secuencia de aminoácidos del péptido. La señal o integración de los picos de los iones puede ser utilizada como técnica de cuantificación, ya que en teoría la intensidad de los picos es proporcional a la cantidad de proteína de la cual proviene el ion peptídico. Sin embargo, la variabilidad técnica, tanto a nivel de la cromatografía líquida como a nivel de los pasos de ionización, hacen que la comparación de las intensidades no sea confiable. Por ello, se han desarrollado diferentes técnicas de marcado de las proteínas que permiten una directa comparación de los picos en un mismo espectro MS o MS/MS. La base de todas estas técnicas es la misma: se introducen marcas inertes, estables e isotópicas a los péptidos de manera que los patrones de ionización y las propiedades cromatográficas de los péptidos marcados sean similares a los de las muestras de origen, pero pueden ser diferenciados luego del análisis por un leve corrimiento en el espectro de MS. De esta manera, las dos muestras a comparar pueden ser corridas de manera simultánea y analizadas a partir de un mismo experimento. Por ello, la separación de los péptidos seguida de la cuantificación de los iones refleja la abundancia relativa de cada proteína intacta de la cual deriva el péptido cuantificado. Los distintos métodos de marcado se diferencian en el momento en

el que la marca es introducida a la proteína o a los péptidos, y cómo son extraídos los datos de cuantificación. A continuación se describen brevemente los distintos tipos de marcado.

#### 4.1. Marcado isotópico *in vivo*.

Este es un método comúnmente utilizado en proteómica comparativa. Para estos experimentos, una de las muestras se crece en una atmósfera de nitrógeno natural, mientras que la muestra a comparar se crece en presencia del isótopo pesado. Esto puede hacerse por la introducción de algún aminoácido marcado en cultivos celulares o utilizando  $^{15}\text{N}$  como única fuente de nitrógeno, típicamente en la forma de  $\text{K}^{15}\text{NO}_3$ . Como resultado, las masas de los péptidos de las poblaciones a comparar serán diferentes, permitiendo una comparación directa de las intensidades de los iones de las dos mezclas. En principio, las dos muestras a comparar pueden ser mezcladas al comienzo del experimento, por lo que se eliminan virtualmente todas las potenciales variaciones técnicas. Debido a que la diferencia isotópica de cada péptido es la misma entre los espectros de las dos muestras, el estudio del proteoma completo puede ser problemático debido a la complejidad de los espectros, por lo que esta técnica es recomendable para sub-proteomas, como por ejemplo el fosfoproteoma.

#### 4.2. Marcado isotópico *in vitro*.

Una alternativa es introducir la marca en los péptidos de manera química durante o luego de la digestión de las proteínas. Una ventaja de esta técnica es que cualquier muestra proteica puede ser marcada, eliminando la limitación que existe al querer introducir la marca en una célula viva. La mayor desventaja es el cuidado que hay que tener para no introducir variaciones técnicas durante la obtención de la muestra proteica (en especial si existe algún paso de fracción subcelular).

##### 4.2.1. Marcado con $^{18}\text{O}$ durante la digestión con tripsina

Durante la hidrólisis de las proteínas con tripsina, el oxígeno del agua se incorpora al C terminal del péptido triptico. Por lo tanto, si la digestión de una muestra se realiza en presencia de  $\text{H}_2^{18}\text{O}$  mientras que la de la muestra

a comparar se realiza en un medio con  $H_2^{16}O$ , en teoría se generan péptidos que difieren en cuatro unidades de masa en el espectro. Experimentalmente, sin embargo, la incorporación es siempre incompleta, generándose espectros que se solapan. Por esta razón este método no es tan elegido como otros, aunque presenta bajo costo comparado a los que se describen a continuación.

#### 4.2.2. Marcado isotópico con sondas por afinidad (ICAT).

Otro método de marcado se basa en la modificación covalente de Cys con compuestos químicos biotinilados que sólo difieren en la inclusión de isótopos pesados y livianos luego de la digestión. El uso de biotina permite el rápido enriquecimiento de los péptidos marcados, y debido a que la mayoría de las proteínas contienen pocas Cys, sólo algunos péptidos de cada proteína son analizados. Por ello, este método permite el análisis de muestras complejas. Sin embargo, dado que una de cada siete proteínas no contiene residuos de Cys, existen limitaciones en el análisis.

#### 4.2.3. Marcas isobáricas para cuantificación relativa y absoluta (iTRAQ).

Un método alternativo involucra la modificación de aminas primarias con marcas isobáricas. Esta técnica es similar a ICAT en cuanto a que cada muestra es marcada con una marca diferente, pero con los reactivos iTRAQ todos los péptidos tripticos son marcados. Además, esta técnica permite el análisis de hasta 8 muestras de manera simultánea si se utilizan distintos reactivos iTRAQ para marcar las distintas muestras.

#### 4.3. Cuantificación libre de marca.

A pesar de los problemas de variabilidad descritos anteriormente en la cuantificación directa de la intensidad de los picos de MS luego de la cromatografía líquida, debido a la aparición de espectrómetros de masa de última generación y al aumento de la reproducibilidad de las separaciones en las cromatografías gaseosas, esta técnica ha comenzado a ser utilizada nuevamente. Para ello, varios programas de computación han sido desarrollados para comparar

los péptidos de múltiples espectros de LC-MS/MS. Una vez identificado el mismo péptido en distintos espectros, éstos se comparan y se genera una relación de expresión luego de un análisis estadístico.

#### 4.4. Cuantificación absoluta utilizando péptidos AQUA

Los métodos descritos anteriormente son capaces de determinar cantidades relativas de proteínas o péptidos. Para convertir datos relativos a absolutos se requiere la inclusión de estándares internos de concentraciones conocidas, que deben estar marcados para distinguirse de la proteína nativa. Para ello se utiliza la técnica de marcado isotópico estable, en la que péptidos sintéticos son marcados con un aminoácido pesado en una o más posiciones (llamado péptido AQUA). Debido a que las propiedades cromatográficas e ionizantes de los péptidos AQUA son idénticas a las de los péptidos nativos, al analizar e integrar los cromatogramas de los iones del par isotópico, se obtiene una relación entre el péptido nativo y el sintético de concentración conocida. Estos péptidos pueden ser sintetizados para múltiples proteínas de interés y usados para determinar concentraciones absolutas en una muestra. Además, pueden ser utilizados para cuantificar modificaciones postraduccionales.

### 5. Aplicaciones de la proteómica

El análisis del proteoma de órganos o tejidos de plantas se ha llevado a cabo para analizar, por ejemplo, la respuesta a hormonas o moléculas señal, la respuesta a estímulos del ambiente, cambios durante el desarrollo o también para analizar los patrones proteicos de líneas con diferente base genética. Proteomas subcelulares también han sido analizados en diversas especies y, más recientemente, se ha utilizado este tipo de estudio para analizar interacción de proteínas o modificaciones postraduccionales. Además, se han analizado por esta técnica la equivalencia y seguridad de alimentos derivados de plantas transgénicas, así como el estudio de la presencia de alérgenos, entre otros. Así, los campos de aplicación de la proteómica en plantas ha ido en aumento en los últimos años (ver revisiones: Cánovas et al.

2004; Rossignol et al. 2006; Jarrín et al. 2007) y a continuación se detallarán algunos de los aspectos más estudiados. La Tabla 2 resume las áreas más relevantes en las que se ha aplicado la proteómica en plantas.

#### 5.1. Análisis de proteomas subcelulares de plantas

El mayor avance en proteomas subcelulares se ha llevado a cabo en plastidios, mitocondrias y membranas. Estos estudios no sólo han contribuido a la localización de proteínas específicas, sino también que ha brindado información relevante sobre transporte, metabolismo, vías de secreción, diferenciación y respuesta a estrés. A pesar de que la purificación de plastidios es una técnica fácil y que brinda fracciones de alta pureza, el aislamiento de otras organelas o fracciones subcelulares es técnicamente más dificultosa. En todos los casos, siempre se debe estimar el nivel de contaminación de la fracción purificada utilizando marca-

dores específicos. Cabe destacar que muchas de las proteínas identificadas en los diversos compartimentos provienen de anotaciones de genes con función desconocida, por lo cual el conocimiento de la localización de la proteína que codifican permite comenzar a identificar la función de dichos genes.

El análisis de proteomas de plastidios, especialmente de cloroplastos, ha brindado información sobre nuevas proteínas localizadas en membrana externa y en tilacoides. Estos estudios, llevados a cabo fundamentalmente en *Arabidopsis* y espinaca, han evidenciado diversos procesos de procesamiento y translocación y modificaciones postraduccionales de proteínas. Gracias al uso de extracciones con diversos solventes, se han identificados nuevas proteínas candidatas a ser transportadores de membrana. La técnica de electroforesis nativa en presencia de azul de Coomassie (BN-PAGE), seguida por segunda dimensión e

**Tabla 2 Aplicaciones de la proteómica en plantas:** Especies analizadas, tipos de proteomas estudiados, procesos biológicos y aspectos prácticos donde se ha analizado el proteoma.

Especies analizadas	Proteomas	Procesos Biológicos	Aspectos prácticos
<b>Sistemas modelos</b> <i>Arabidopsis thaliana</i> <i>Medicago truncatula</i>	<b>Individuales</b> Genotipos Mutantes Trangénicas	<b>Desarrollo</b> Germinación Maduración polen Embriogénesis Respuesta a hormonas Muerte celular programada	Identidad de marcadores moleculares  Caracterización del genotipo
<b>Cereales</b> Maíz Arroz Trigo	<b>Fracciones subcelulares</b> Cloroplasto Mitocondria Plastidio		Equivalencia de OGM
<b>Leguminosas</b> Alfalfa Arveja Soja Lenteja	Citosol Pared celular Membrana Vacuola	<b>Estrés abiótico</b> Sequía Osmótico Temperatura Anoxia Deficiencia nutricional Metales pesados UV-B	Identificación alérgenos
<b>Otros</b> Papa Tabaco Tomate Vid Espinaca	<b>Órganos y Tejidos</b> Raíces Semillas Coleoptiles Haces vasculares Hojas Tricomas Polen	<b>Estrés biótico</b> Patógenos Herbívoros Estrés oxidativo	
	<b>Estados de desarrollo</b> Germinación Plantas maduras Plantas jóvenes		

identificación de las proteínas por MS, técnica ampliamente utilizada para el estudio de complejos mitocondriales, también ha sido utilizada para el estudio de la composición de los complejos proteicos de la membrana tilacoidal.

El estudio del proteoma mitocondrial se ha dividido en la identificación de proteínas solubles, proteínas periféricas y proteínas integrales de membrana. En este caso, el uso de BN-PAGE seguido por 2-DE e identificación de las proteínas, se ha utilizado para analizar la composición de los diversos complejos respiratorios, identificándose nuevos componentes y diferencias con complejos respiratorios de otros organismos.

El análisis de proteínas de membrana es uno de los desafíos más difíciles, debido a problemas de metodología por la baja abundancia de las proteínas y sus características hidrofóbicas, siendo muchas de ellas proteínas transmembrana. Para el análisis de este tipo de proteínas, se han empleado partición de proteínas en distintas fases de solventes seguido de fraccionamiento teniendo en cuenta las propiedades fisicoquímicas de las proteínas. Por ejemplo, el proteoma de la membrana plasmática de *Arabidopsis* fue analizado, siendo uno de los primeros que, además, se sometió a análisis de proteínas fosforiladas, identificándose más de 300 sitios de fosforilación. El proteoma vacuolar de *Arabidopsis*, tanto de la membrana como de las proteínas solubles, también ha sido analizado, confirmándose dicha localización subcelular para algunas de las proteínas identificadas por otras técnicas.

Además, se han llevado a cabo estudios para identificar proteínas de pared celular y extracelulares –identificándose la presencia de varias proteasas-, proteínas nucleares –muy difícil debido a su baja cantidad y a sus asociación con el Retículo Endoplasmático-, y proteomas mixtos del aparato de Golgi y del Retículo endoplasmático.

## 5.2. Proteomas durante el desarrollo de órganos

Uno de los primeros objetivos de la proteómica en plantas fue generar mapas de referencia de proteínas solubles de un órgano en un determinado estado de desarrollo. En

los últimos años un gran número de trabajos se han realizado con el objetivo de identificar los cambios proteicos que ocurren asociados al crecimiento y desarrollo, con la finalidad de identificar las proteínas claves implicadas en dichos procesos. Así, se han descrito los cambios en el proteoma asociados al crecimiento o desarrollo del grano de polen, semillas, raíces, haces vasculares, entre otros.

Muchos son los estudios que demuestran que los análisis de proteómica pueden generar información relevante para dilucidar los mecanismos moleculares involucrados en procesos de desarrollo. Por ejemplo, estudios llevados a cabo en raíces de maíz han indicado que, de las proteínas identificadas luego de 5-días de germinación, sólo el 28% permanece en las últimas etapas del desarrollo, identificándose, además, proteínas asociadas a la formación de raíces laterales. Procesos como la embriogénesis somática también han sido analizados en diversas especies por proteómica, identificándose proteínas de membrana, nucleares, así como proteínas involucradas en metabolismo y producción de energía, en diversas etapas de este proceso. En muchos casos, se han identificado proteínas cuyos ARNm no fueron identificados por estudios de transcriptoma, como es el caso del análisis del grano de polen en *Arabidopsis*, entre otros.

## 5.3. Proteomas en respuesta a estrés abiótico

El estudio de la interacción de las plantas con el ambiente ha sido clásicamente abordado mediante estudios bioquímicos, genéticos y más recientemente, a nivel de transcriptoma, mediante PCR en tiempo real (qRT-PCR), microarreglos, entre otros. Sin embargo, la proteómica, desde su aplicación en este tipo de situaciones, ha comenzado a brindar contribuciones relevantes en cuanto a los elementos involucrados tanto en la percepción de la situación ambiental como en la traducción de dicha señal en la planta. Los tipos de estrés analizados mediante proteómica han sido: sequía, estrés salino, alta temperatura, alta luz, radiación UV-B, deficiencia nutricional, niveles elevados de CO<sub>2</sub>, anoxia, metales pesados y herbicidas.

Este tipo de aplicación de la proteómica incluye típicamente análisis de 2-DE/MS para estudiar variaciones proteicas entre especies sensibles y resistentes frente a una determinada situación de estrés. Se analizan las diferencias cuantitativas y cualitativas de proteínas con la finalidad de correlacionar estos cambios con las diferencias fenotípicas. Se estudian también plantas mutantes o transgénicas, donde una línea es resistente y otra sensible. Así, las proteínas diferenciales se seleccionan contrastando los proteomas entre líneas resistentes *versus* susceptibles, o entre líneas crecidas en condiciones óptimas *versus* estresadas.

#### 5.4. Proteomas en respuesta a estrés biótico

En los últimos años, muchos estudios se han llevado a cabo para analizar la respuesta de los proteomas de diversos órganos frente al ataque por hongos, virus o bacterias. Típicamente, extractos crudos de hojas, semillas, raíces o cultivos celulares se analizan a distintos tiempos luego del tratamiento y se separan organelas o subfracciones para el análisis de los cambios en el proteoma. Además, se analiza el efecto de elicitores o moléculas claves de la señalización (como ácido salicílico, peróxido de hidrógeno u óxido nítrico) sobre el proteoma. Por otro lado, se estudian los proteomas de especies tolerantes o resistentes *versus* el de sensibles. Estos estudios, junto con el estudio del proteoma de diversos compartimentos celulares (apoplasto, tricoma, sap de floema o xilema, entre otros), han evidenciado que varias proteínas relacionadas a la defensa frente al ataque de patógenos están presentes incluso en ausencia del mismo y que las mismas proteínas responden a tipos variados de estrés. Además, muchas proteínas derivadas de genes de función desconocida, han sido encontradas en estos estudios.

En general, la mayoría de las proteínas identificadas en los estudios de proteomas de genotipos resistentes comparadas con los de sensibles, han mostrado dos tipos generales de proteínas: proteínas relacionadas a la defensa y enzimas asociadas al metabolismo del carbono o nitrógeno y al metabolismo secundario. Dentro del primer grupo, se han encontrado proteínas PR (como glucanasas,

quitinasas, proteasa, inhibidores de proteasas), antioxidantes (catalasas, SODs, peroxidases), chaperonas, HSPs, proteínas LEA. En el segundo grupo, varias enzimas asociadas a metabolismo como la glucólisis, fotosíntesis, ciclo de Krebs, asimilación del nitrógeno o implicadas en metabolismos secundarios han sido encontradas. Es notable que los estudios basados en análisis del transcriptoma han analizado en mucha mayor medida el primer grupo que el segundo. Sin embargo, la presencia de mayor concentración de proteínas implicadas en metabolismo primario en genotipos resistentes *versus* sensibles podría indicar una ventaja genética para defenderse frente al ataque.

#### 5.5. Proteómica para el estudio de simbiosis en plantas

La proteómica se ha aplicado también para el estudio de las proteínas involucradas en la formación de nódulos en especies leguminosas. Mapas de proteínas de raíces de especies hospedadoras, nódulos, cultivos de bacterioides fueron generados y comparados, con el objetivo de identificar proteínas involucradas en la simbiosis. Además, se analizaron las proteínas involucradas en distintas etapas de la nodulación.

#### 5.6. Modificación postraducciona e interacción de proteínas

Una de las modificaciones postraduccionales de proteínas estudiadas mediante proteómica es la fosforilación. En estos estudios, se han identificado los cambios cuantitativos y cualitativos de las proteínas fosforiladas en respuesta a diversas señales. Además, se han analizado las proteínas modificadas en su estado redox en diferentes sistemas, identificándose muchos candidatos putativos de ser blanco de reducción por tioredoxinas. Recientemente, se ha comenzado a analizar el patrón de proteínas sujetas a ubiquitinación en plantas, mediante estudios de proteómica.

En cuanto al estudio de interacción de proteínas mediante proteómica, los estudios llevados a cabo en plantas son todavía muy pocos comparados con los llevados a cabo en otros sistemas, como mamíferos y levaduras. Algunos ejemplos incluyen estudios de com-

plejos formados por virus de plantas y proteínas del huésped, análisis de inmunoprecipitados de proteínas que interaccionan con fitocromos, así como proteínas que interaccionan con cobre.

## 6. Perspectivas

Los resultados obtenidos mediante estudios de proteómica en sus distintas aplicaciones han indicado que los estudios de genómica y transcriptoma no son suficientes, dado que la proteómica generalmente evidencia procesos no detectados por los otros estudios mencionados.

Esto se debe fundamentalmente a que las proteínas son física y químicamente más diversas que los ácidos nucleicos. Además, debido al procesamiento del ARN y a las modificaciones postraduccionales, las proteínas de un dado organismo exceden en número la cantidad de genes. Otro nivel de complejidad se presenta si se tienen en cuenta la interacción de proteína-proteína, lo cual modifica las propiedades de las proteínas individuales.

De esta manera, el estudio del proteoma, en paralelo con el del transcriptoma, llevado a cabo sobre el mismo sistema, podrá dilucidar los mecanismos de control de expresión de proteínas y evidenciará los procesos biológicos llevados a cabo en una determinada situación.

## 7. Bibliografía

- Cánovas F. M., Dumas-Gaudot E., Recorbet G., Jorrin J., Mock H-P and Rossignol M. 2005. Plant Proteome Analysis. *Proteomics*, **4**, 285-298
- Chen S., and Harmon A. C. 2006. Advances in plant proteomics. *Proteomics*, **6**, 5504–5516.
- Jorrin J. V., Maldonado A. M. and Castillejo M. A. 2007. Plant proteome analysis: A 2006 update. *Proteomics*, **7**, 2947-2962
- Nesatyy V. J. and Suter M. J-F. 2007. *Environmental Science & Technology*, **41**, 6891-6900.
- Newton R. P., Brenton, A. G., Smith, C. J. and Dudley E. 2004. Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. *Phytochemistry*, **65**, 1449-1485.
- Thelen J.J. and Peck S.C. 2007. Quantitative Proteomics in Plants: Choices in Abundance. *Plant Cell*, **19**, 3339–3346.

Rossignol M., Peltier J-B., Mock H-P., Matros A., Maldonado A. M. and Jarrin J. V. 2006. Plant proteome analysis: A 2004-2006 update. *Proteomics*, **6**, 5529-5548.

# I CAPÍTULO 10

## Metabolómica

Fernando Carrari, Telma E. Scarpeci,  
Luciano A. Abriata, Alejandro J. Vila y  
Estela M. Valle.

### Definición de metaboloma

Se conoce como metabolismo al conjunto de reacciones químicas que ocurren en las células vivas, tejidos u organismos, mediante el cual estos participan del intercambio de materia y energía con su entorno. El origen de la palabra metabolismo procede del griego *metabolé* que significa cambio, transformación. En los sistemas biológicos los metabolismos se encuentran interconectados formando redes bioquímicas. En los últimos años y en esta era post-genómica, se ha acuñado el término "metaboloma", para indicar la totalidad de los metabolitos que actúan en las redes metabólicas de un sistema biológico. Es decir, el metaboloma implica la identificación y cuantificación simultáneas de todos los metabolitos del material investigado.

El metaboloma está conformado por dos tipos de compuestos: los metabolitos primarios y los secundarios. Los metabolitos primarios son compuestos que están involucrados en funciones básicas de la célula, como la respiración o la biosíntesis de aminoácidos. Todos los organismos comparten básicamente los mismos tipos de metabolitos primarios y en términos generales se puede decir que se trata de moléculas simples y de bajo peso molecular. En cambio, los metabolitos secundarios son específicos de la especie y muchos de ellos están implicados en la defensa de la planta contra estrés biótico (siendo los terpenos los más abundantes), en procesos de desarrollo específicos de la especie (hormonas) y en estrategias de atracción de polinizadores (pigmentos de las flores). En las plantas, se estima que el número total de metabolitos supera los 200.000. El análisis del metaboloma de un sistema biológico dado es conocido como metabolómica. Este término no siempre tiene el mismo significado, ya que la metabolómica es multifuncional y depende del diseño experimental y del objeto de estudio. En

general, la metabolómica indica la totalidad de los patrones metabólicos de los sistemas biológicos, también conocidos como perfiles metabólicos. Como los metabolitos son el producto de reacciones enzimáticas, ellos brindarán información importante sobre las propiedades regulatorias o catalíticas de un producto génico dado. Los metabolitos presentes en un sistema biológico bajo condiciones específicas (estadios de desarrollo, condiciones ambientales, modificaciones genéticas) se pueden analizar mediante el uso de técnicas cromatográficas de separación en distintas fases (gaseosa, líquida, de alta presión, etc) y de identificación mediante espectrometría de alta resolución que han comenzado a establecerse recientemente en Argentina, como la espectrometría de masa (MS) gaseosa o líquida y técnicas espectroscópicas como la resonancia magnética nuclear (RMN). A su vez, el acoplamiento de los sistemas de separación con los de identificación mencionados más arriba permite la obtención de perfiles metabólicos de muestras complejas como por ejemplo, extractos crudos obtenidos de órganos fotosintéticos de plantas.

### Diseño experimental, técnicas, entrenamiento. Métodos cromatográficos (fase gaseosa y líquida).

El análisis de los metabolitos presentes en muestras complejas (como las que pueden obtenerse a partir de extractos crudos de plantas) puede realizarse en primer lugar a partir de la separación de las moléculas que las componen. Existen diversos métodos de separación con distintos grados de resolución y sensibilidad. La cromatografía en fase gaseosa (GC) ha sido extensamente usada justamente por su alta resolución y sensibilidad. Consiste en la separación de las moléculas por su capacidad móvil en una fase gaseosa impulsada por un gas usualmente inerte (por ejemplo Helio -He-) de modo de evitar reacciones con la matriz a analizar. Luego de su obtención y acondicionamiento, los extractos son inyectados en una cámara de vaporización a alta presión y temperatura, donde la muestra es volatilizada para luego ser transportada a una columna bajo un flujo constante de gas. Existen diferentes protocolos en cuanto a las condiciones y tiempos

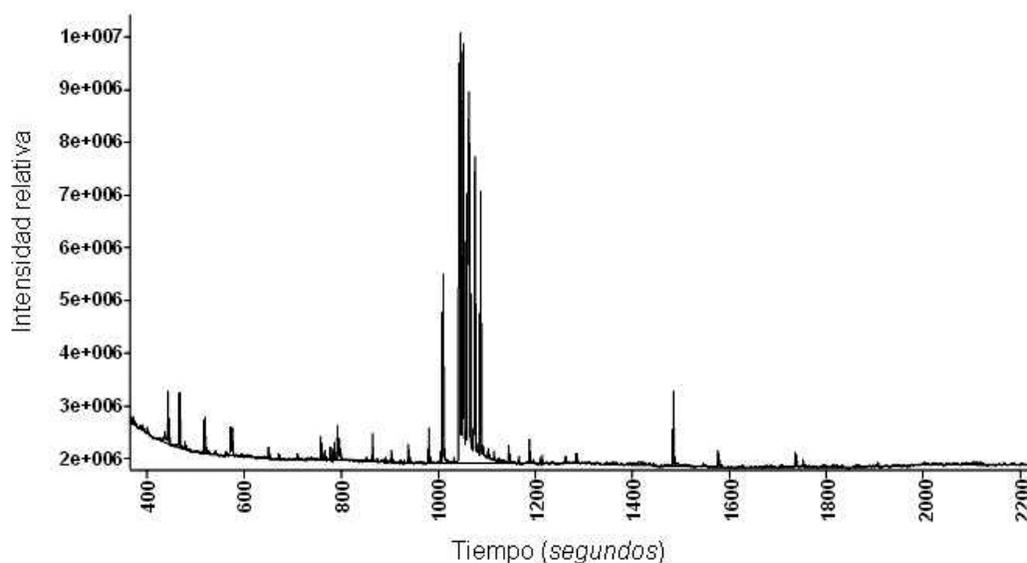
de corrida y se utilizan columnas diferentes en función del tipo de muestras y moléculas a analizar. Las moléculas se separan de acuerdo a sus propiedades físicas y químicas y al final de la columna, un detector emite una señal que es registrada como una serie de picos los que constituyen el cromatograma. En la figura 1 se observa un cromatograma obtenido luego de aproximadamente 40 minutos de corrida de una muestra de extractos de frutos de tomate. El ápice de un pico dado representa el momento de mayor intensidad registrado y se lo conoce como el tiempo de retención (RT) de ese grupo de moléculas. El RT es una propiedad de las moléculas que permite la identificación de las mismas en el cromatograma.

Sin embargo, dada la necesidad de volatilizar los extractos, es necesario un tratamiento previo de los mismos a fin de transformar las moléculas en compuestos menos polares y más volátiles. Este paso, denominado derivatización, consiste en la adición de grupos químicos (tales como sililo  $-\text{Si}(\text{CH}_3)_3-$  y **metoxi**) a las moléculas que componen la muestra com-

pleja. Si bien se han desarrollado diferentes métodos de derivatización, esta etapa constituye la principal desventaja de la cromatografía gaseosa dada la diversidad de propiedades físico-químicas de las miles de moléculas que componen un extracto crudo. La principal dificultad radica en la exacta identificación de los fragmentos obtenidos luego de la ionización de las distintas moléculas (ver siguiente sección: “Espectrometría de masas. Principios básicos”).

En este sentido, la aplicación de la cromatografía en fase líquida (LC) constituye un complemento importante a la gaseosa y contribuye a expandir el rango de aplicaciones de las técnicas cromatográficas en metabolómica. Si bien la resolución de ésta es menor que la gaseosa, permite la separación de compuestos de mayor peso molecular y polaridad sin la necesidad del paso de derivatización previamente mencionado.

No obstante los avances en el conocimiento en cuanto a la separación e identificación de compuestos orgánicos, la cromatografía no ha



**Figura 1.** Cromatograma obtenido luego de aproximadamente 40 minutos de corrida de un extracto crudo de fruto de tomate. Sobre las abscisas se observan los tiempos de retención para grupo de moléculas en una columna en fase gaseosa. Sobre las ordenadas se grafican las intensidades relativas de cada grupo de moléculas separado. Bajo condiciones experimentales definidas estas intensidades son proporcionales a la cantidad de la molécula en análisis presente en una muestra dada.

resuelto aún la exacta identificación de moléculas individuales presentes en una muestra compleja. El empleo de sistemas de cromatografía acoplados a espectrometría de masas resulta de gran utilidad en la identificación de moléculas que se encuentran en trazas en una muestra compleja con un grado de exactitud de hasta 3-4 decimales.

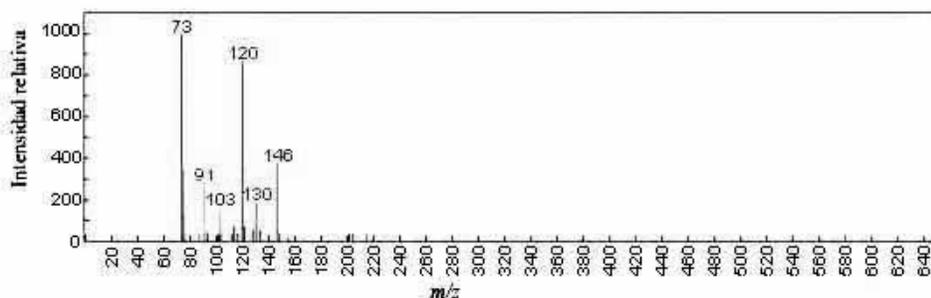
### **Espectrometría de masas. Principios básicos.**

La espectrometría de masas es un método que permite la identificación de moléculas a partir del patrón de fragmentos que se generan luego de la ionización de las mismas. A este patrón se lo denomina espectro de masas y constituye la “huella digital” de una molécula dada. Esta es una de las funciones primarias de los espectrómetros de masas. Existen varias métodos para la ionización de moléculas: ionización química (CI), por impacto de electrones (EI), bombardeo rápido de átomos (FAB), por campo eléctrico (FI), por láser (LIMS), desorción asistida por láser (MALDI) e ionización por electrospray (ESI). Este último método es uno de los más utilizados para la generación de los espectros de masas de moléculas orgánicas el cual consiste en un impacto con electrones de alta energía aplicado sobre una corriente de moléculas en fase líquida. Esto produce la nebulización y fragmentación de las moléculas y mediante cambios rápidos de temperatura se

produce luego una corriente gaseosa que conduce a los iones obtenidos hacia los detectores del espectrómetro. Durante esta corriente los iones generados son separados de acuerdo a su masa y carga que al impactar sobre los detectores generan una señal eléctrica que es registrada de acuerdo a su intensidad. La manera gráfica de representar estas señales es lo que se conoce como espectro de masas de una molécula. En la figura 2 se observa un espectro de masas del aminoácido fenilalanina.

### **Preparación de muestras. Extracción y derivatización.**

La obtención de muestras biológicas para el análisis de sus perfiles metabólicos, como cualquier otra técnica que involucre la identificación y cuantificación de metabolitos, requiere la inmediata inactivación del metabolismo de modo de evitar efectos de su manipulación sobre el mismo. La inhibición rápida del metabolismo se logra mediante la disminución brusca de temperatura ( $\leq -60^{\circ}\text{C}$ ) por inmersión de las muestras en  $\text{N}_2$  líquido. A su vez, los procedimientos posteriores de extracción requieren el uso de materiales (tales como tubos plásticos o de vidrio) de alta calidad y pureza de modo de prevenir fuentes de eventuales contaminaciones. A su vez, dado que una de las aplicaciones comunes de estas técnicas es la comparación del estado metabólico de organismos frente a perturbaciones ambientales (por ejemplo de



**Figura 2.** Espectro de masas del aminoácido fenilalanina. Sobre las abscisas se observan las relaciones masa/carga ( $m/z$ ) de los diferentes fragmentos obtenidos luego de la ionización de la molécula. Sobre las ordenadas se grafican las intensidades relativas de cada fragmento obtenido. Bajo condiciones experimentales definidas estas intensidades son proporcionales a la cantidad de la molécula en análisis presente en una muestra dada. Fuente: adaptado de “*The Golm Metabolome databases*”.

plantas de cultivo en ambientes extremos) y/o a cambios en la constitución genética de los mismos (por ejemplo de plantas transgénicas que expresen diferencialmente una proteína involucrada en señalización por azúcares) una necesidad básica es las consideraciones de controles tanto de los materiales “*per se*” como de las condiciones en las que las muestras son extraídas y procesadas. A su vez, la aplicación de algoritmos estadísticos a los datos obtenidos requiere la determinación del grado de variación dentro de la población de muestras a ser analizadas.

A efectos explicativos, en este capítulo nos concentraremos en un ejemplo de muestreo, acondicionamiento y extracción de muestras de hojas de plantas (de tomate, papa, tabaco o *Arabidopsis*) en condiciones ambientales controladas (22-24°C, 250  $\mu\text{mol fotones}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$  y 16 h de luz/8 h de oscuridad). El momento del día para el muestreo debe ser cuidadosamente controlado. Como regla general, la mitad del periodo de luz es el más indicado debido a que numerosos reportes en la literatura dan cuenta de los ritmos diurnos en los cambios en la composición metabólica de plantas. El otro punto importante a tener en cuenta es el estado fenológico de las plantas y los órganos a muestrear. La posición de las hojas en la planta y el área de las mismas también debe ser tenido en cuenta. Además, una vez que las porciones de hojas fueron cortadas de la planta, se debe registrar la masa exacta de la muestra y congelarlas inmediatamente. Las muestras deben mantenerse a la temperatura indicada durante todo el proceso de extracción. Para ello es conveniente enfriar todos los instrumentos necesarios para la toma de muestras, tales como tijeras, pinzas, sacabocados, tubos, etc. Para el ejemplo mencionado, la cantidad de tejido necesario puede variar entre 100-150 mg (peso fresco). La muestra debe ser homogeneizada en  $\text{N}_2$  líquido, las enzimas inactivadas mediante el agregado de metanol y al mismo tiempo deben agregarse estándares internos (en concentraciones conocidas) no presentes en la muestra a analizar. La extracción se realiza luego calentando la muestra a 70°C con agitación constante. Si bien este protocolo puede aplicarse a la obtención tanto de la fase polar

como apolar, a los efectos de este ejemplo, los pasos siguientes se concentran en la obtención de la fase polar. Mediante un paso de extracción con cloroformo-agua, las alícuotas de la fase polar son concentradas y liofilizadas para su posterior derivatización.

Tal como se mencionó previamente, la derivatización consiste en la modificación de las moléculas a fin de disminuir su polaridad y permitir la volatilización de las mismas. Existen varios métodos y reactivos derivatizantes. Entre los más usados para esta aplicación, el MSTFA (N-metil-N-trimetilsililtrifluoroacetamida) ha demostrado ser uno de los más reactivos en reemplazar los hidrógenos lábiles de un amplio rango de compuestos polares por grupos sililos y el hidroxiclورو de metoxiamina es usado en un segundo paso de derivatización a fin de proteger grupos aldehídos y cetonas de las moléculas presentes en la matriz.

Los extractos así preparados son inmediatamente inyectados en el cromatógrafo acoplado al espectrómetro de masas. Tanto las condiciones de corrida (tiempo, rampa de temperaturas, frecuencia de captura de datos, etc) como las especificaciones de flujo de gas y columnas cromatográficas deben ser puestas a punto para cada aplicación particular. En este sentido, en los últimos años se dispone de programas de computación especialmente diseñados por los fabricantes de espectrómetros que controlan toda la operación y posterior procesamiento de datos. A su vez, la disponibilidad de colecciones de espectros de masas en bases de datos de acceso público (ver sitios recomendados) permite la rápida evaluación de los datos adquiridos.

### **Resonancia magnética nuclear. Principios básicos**

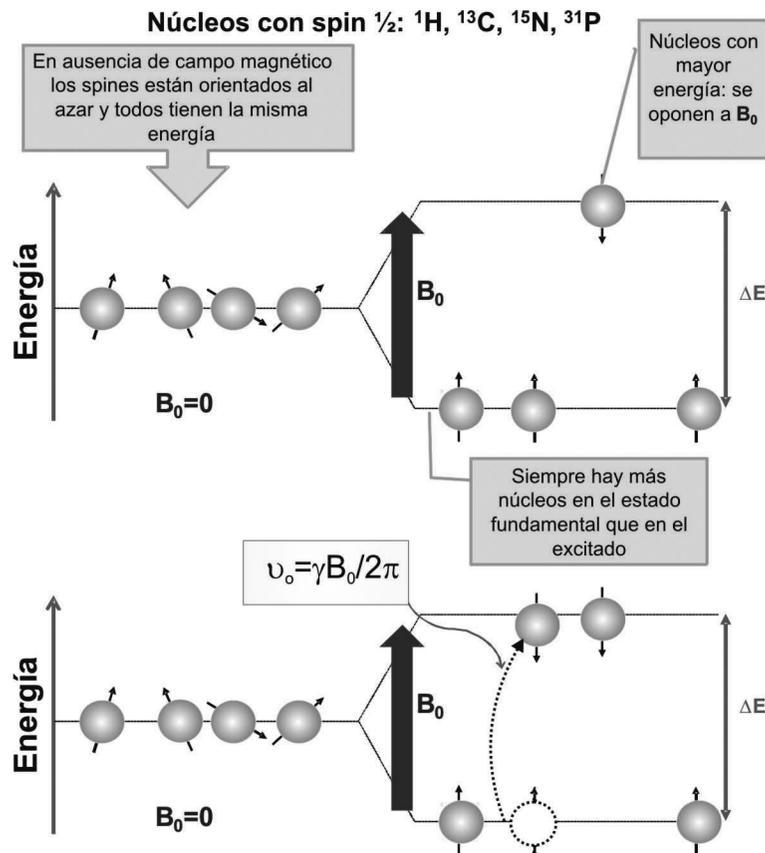
La resonancia magnética nuclear es un método de análisis espectroscópico no destructivo, que se basa en la absorción de energía en la zona de radiofrecuencia por parte de los núcleos de algunos átomos, en presencia de un campo magnético. La MS es la técnica más utilizada en metabolómica. Sin embargo, la técnica de RMN está ganando terreno en este sector, como una técnica que brinda información complementaria, ya que amplía el rango

de metabolitos detectados y provee información estructural detallada de las moléculas presentes. Más aún, la RMN permite identificar sustancias previamente no halladas en muestras similares.

La RMN explota una propiedad de los núcleos atómicos, el spin nuclear. El spin nuclear posee un momento magnético, es decir que se comporta como un pequeño imán. Normalmente, estos momentos magnéticos están orientados al azar y tienen la misma energía (figura 3). Si se aplica un campo magnético fijo ( $B_0$ ) con una dada dirección, estos momentos magnéticos quedarán alineados a favor o en contra del campo magnético externo (figura 3). Según su orientación con respecto al campo  $B_0$ , los núcleos tendrán ahora distinta energía. La diferencia de energía ( $\Delta E$ ) entre estos dos niveles está en el orden de las ondas de radio, es decir con frecuencias de MHz. Si ahora aplicamos al sistema orientado la frecuencia correspondiente a esa  $\Delta E$  ( $\nu_0$ ), se inducirá una transición entre

estos niveles, que es el fenómeno de Resonancia (figura 3).

La frecuencia de absorción  $\nu_0$  depende del campo aplicado ( $B_0$ ), el cual está fijo para cada equipo, y la constante giromagnética nuclear ( $\gamma$ ) que es característica de cada núcleo (figura 3). Es decir que cada núcleo tendrá una frecuencia de absorción característica. Es importante tener en cuenta que, como se trata de una propiedad del núcleo atómico, dado un elemento químico, los distintos isótopos del mismo presentarán distintas propiedades magnéticas. Algunos isótopos (dependiendo de su composición) pueden carecer de spin, y por lo tanto no serán activos en RMN y este es el caso de  $^{12}\text{C}$ ,  $^{16}\text{O}$  y  $^{32}\text{S}$ . Por simplicidad, en este capítulo se considerarán sólo los núcleos con spin  $\frac{1}{2}$ , que son los de interés biológico. Dentro de los núcleos de spin  $\frac{1}{2}$  se encuentran los núcleos de  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  y  $^{31}\text{P}$ , que pueden ser estudiados por RMN. Cada núcleo posee una sensibilidad determinada, que está determinada por su constante giromagnética. El or-



**Figura 3.** Esquema en donde se muestra los niveles de energía de los núcleos en presencia o ausencia de un campo magnético externo ( $B_0$ ) y las transiciones de los núcleos entre un nivel de energía y otro superior al absorber una frecuencia ( $\nu_0$ ) correspondiente a la diferencia de energía ( $\Delta E$ ).

den de sensibilidad de estos núcleos es:  $^1\text{H} > ^{31}\text{P} > ^{13}\text{C} > ^{15}\text{N}$ . La detección de  $^1\text{H}$  o  $^{31}\text{P}$ , además de su alta sensibilidad, está favorecida por la alta abundancia natural de estos isótopos (99,9% y 100%, respectivamente). En cambio, los isótopos de  $^{13}\text{C}$  y  $^{15}\text{N}$  poseen una abundancia de 1,1% y 0,4%, lo cual dificulta más aún su detección. Este inconveniente se puede sortear enriqueciendo la muestra con el isótopo activo de interés. Cabe destacar que ninguno de estos isótopos es radioactivo, lo que facilita la manipulación de las muestras.

Todos los elementos con isótopos magnéticos detectables por la técnica RMN pueden ser usados para identificar metabolitos en tejidos de plantas y sus extractos. El  $^1\text{H}$  es el isótopo magnético más utilizado debido a su alta sensibilidad intrínseca y su abundancia natural. La concentración límite para poder detectar un metabolito por RMN de  $^1\text{H}$  en un extracto es alrededor de 10  $\mu\text{M}$ . Para una muestra dada, la sensibilidad aumenta con el campo magnético del instrumento utilizado.

Los metabolitos fosforilados pueden ser identificados y cuantificados por RMN utilizando el isótopo magnético  $^{31}\text{P}$ . De esta manera se pueden estudiar un número pequeño de moléculas abundantes y de gran importancia metabólica

como es el fosfato inorgánico, nucleósidos trifosfatos, fosfomonoésteres, etc.

### Desplazamiento químico

La frecuencia de absorción de cada núcleo depende de su constante giromagnética y a un campo dado. Por ejemplo, a un campo de 9,4 T, los  $^1\text{H}$  absorben a 400 MHz, los  $^{13}\text{C}$  a 100,6 MHz y los  $^{15}\text{N}$  a 40 MHz. Sin embargo, no todos los  $^1\text{H}$  ni todos los  $^{13}\text{C}$  absorben a la misma frecuencia. La frecuencia de absorción de cada protón depende de la ubicación del núcleo en las moléculas, es decir, de su ambiente químico. Las distintas posiciones de los núcleos en un espectro se cuantifican con un parámetro llamado *desplazamiento químico* ( $\delta$ ), cuyo valor no sólo informa la frecuencia precisa de absorción, sino que nos permite distinguir el tipo de protón (por ejemplo alifático, aromático, amídico, etc.). La intensidad de una señal de resonancia, o sea el área bajo cada pico, es proporcional al número de núcleos de ese tipo (figura 4).

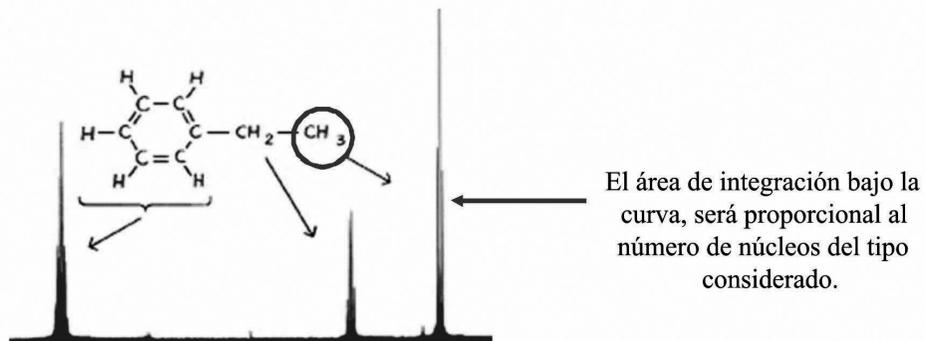
### Espectrómetro de RMN

Cada equipo de RMN trabaja a un campo magnético fijo. En general es conveniente

**Para el caso hipotético de un núcleo aislado:**



**En una molécula, los núcleos van a absorber a distintas frecuencias debido a diferencias locales en el campo, lo cual generará distintas señales en el espectro de RMN**



**Figura 4.** Dependencia ubicación del núcleo en las moléculas y su frecuencia de absorción.

trabajar con un campo magnético alto porque esto trae aparejado una mayor sensibilidad y resolución. Cada equipo se suele denominar no por la magnitud del campo (por ejemplo, 16 Tesla), sino por la frecuencia de absorción del  $^1\text{H}$  a ese campo (para el caso mencionado, 600 MHz). La mayoría de los análisis metabólicos se realizan con instrumentos que operan en el rango de 300 a 900 MHz para  $^1\text{H}$ .

El campo magnético está generado por Helio líquido a bajas temperaturas (4 K), que es superconductor, y se encuentra en un termo que rodea la muestra, que constituye el imán o magneto. Los pulsos de excitación de radiofrecuencia se generan mediante bobinas que están alrededor del tubo donde se encuentra la muestra, complementadas con un conjunto de bobinas que permiten la detección de la señal resultante.

### **Preparación de las muestras**

Se pueden obtener espectros de RMN a partir de una amplia variedad de materiales: extractos vegetales (por ejemplo: jugo de tomate), suspensiones celulares e inclusive plántulas intactas.

En el caso de realizar RMN de  $^1\text{H}$ , es conveniente liofilizar las muestras para eliminar el agua. Esto se realiza con el fin de reducir la señal correspondiente al solvente, que (por su alta intensidad) dificulta la detección de señales de los metabolitos. Luego se resuspende la muestra liofilizada en agua deuterada ( $\text{D}_2\text{O}$ ). La adición de  $\text{D}_2\text{O}$  a la muestra hará que no se puedan detectar las señales de protones intercambiables con el solvente, como los de grupos amida o alcohol. En caso de que sea necesario poder identificar los mismos, deben realizarse experimentos que permitan la eliminación selectiva de la señal del solvente. En caso de que se deseen extraer algunos componentes específicos, también pueden utilizarse solventes orgánicos, de preferencia deuterados (que se venden comercialmente). Conviene utilizar solventes químicamente inertes y volátiles, lo que permite eliminar el solvente en caso de que la muestra deba ser analizada por otra técnica. Esto es posible debido a que la técnica de RMN es no destructiva. Las muestras se colocan en tubos de vidrio borosilicato puro y el

volumen de extracto utilizado es dependiente del ensayo pero varía entre 250 y 400  $\mu\text{l}$ .

### **Utilización de la técnica RMN en metabolómica**

#### **1-Identificación y cuantificación de metabolitos en tejidos vegetales: ventajas y desventajas**

Existen diversas ventajas de la técnica de RMN para el análisis de perfiles metabólicos. Las mayores ventajas con respecto a otras técnicas son: (1) no se requiere tratamiento previo de la muestras (pudiéndose utilizar extractos crudos); (2) es un método no destructivo, lo cual permite el análisis subsiguiente de la muestra con otros métodos; (3) se puede obtener información sobre una amplia variedad de metabolitos en un único experimento; (4) no existen restricciones en cuanto a la volatilidad, polaridad de los compuestos, o la presencia de determinados cromóforos que interfieran; y (5) la técnica puede ser aplicada con escaso conocimiento previo de la composición de la muestra.

Otra gran ventaja de ésta técnica para el análisis cualitativo de metabolitos es que no requiere del conocimiento de ningún parámetro equivalente a los coeficientes de extinción necesarios para cuantificar muestras mediante técnicas espectrofotométricas. Esto se debe a que la magnitud de la señal de resonancia de un isótopo es directamente proporcional al número de núcleos que contribuyen a dicha señal. Lo anterior implica que, por una parte, no se requieren curvas de calibración para efectuar el análisis cuantitativo, y por otro lado no es necesario conocer el espectro de un compuesto puro ya que por lo general basta con identificar señales debidas a una porción dada de la molécula para obtener la información deseada. La información cuantitativa se obtiene a partir del área integrada bajo la curva.

#### **Ejemplo 1: caracterización de extractos de frutos de tomate**

La Figura 5A muestra un espectro de RMN de  $^1\text{H}$  en una muestra de jugo de tomate. En la región de campo alto (3,1-0,3 ppm) del espectro de protones contiene la resonancia de los grupos alifáticos de los aminoácidos y áci-

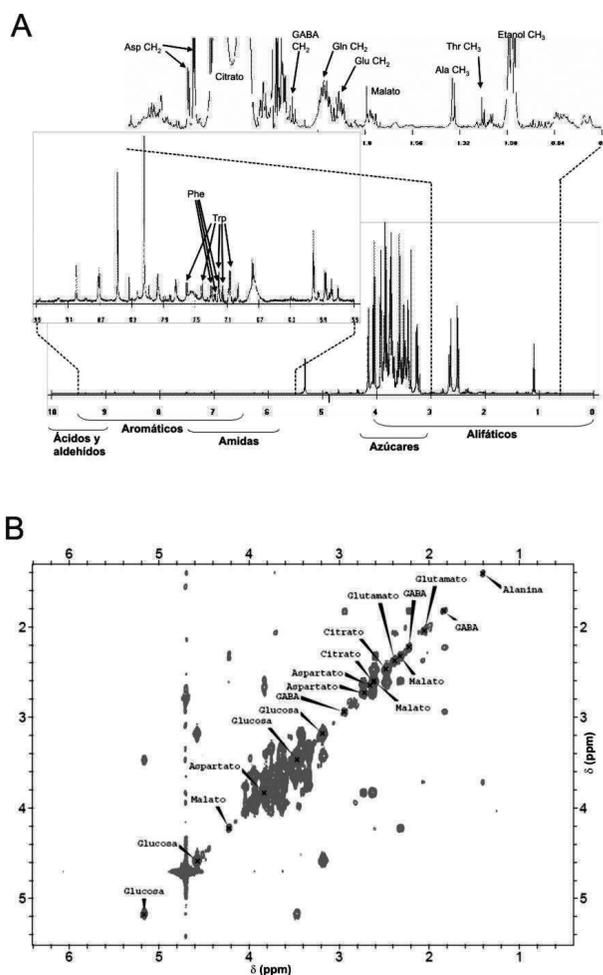
dos orgánicos pertenecientes al metabolismo de los ácidos tricarbónicos. En particular, las señales provenientes de treonina, alanina, glutamato, glutamina, GABA, aspartato, citrato y malato fueron identificadas claramente. Las resonancias encontradas en la región del espectro de campo intermedio (desde 3,2 a 5,3 ppm) corresponden en su mayoría a sacáridos, de los cuales D-glucosa y D-fructosa son los componentes principales. En la región de campo bajo (5,5-9,4 ppm) del espectro se encuentran las señales correspondientes a aminoácidos aromáticos y compuestos fenólicos. Debido a la baja intensidad de la señal, las asignaciones se realizaron parcialmente. Estos resultados muestran que la técnica RMN puede ser una herramienta útil para ser usada en la caracterización de tomates.

La principal desventaja de la técnica de RMN es su baja sensibilidad. El advenimiento de instrumentos de RMN de alto campo, de técnicas especiales de detección y el diseño de criosondas (de mucha mayor sensibilidad) ha mejorado este aspecto. De todas formas, rara vez se usa RMN para el análisis de un componente que se encuentra en trazas. Generalmente, el problema de la baja relación señal-a-ruido en los espectros de RMN (que sucede principalmente con isótopos poco abundantes) puede ser superado haciendo varios espectros de manera automática, y sumándolos. Esto redundará en mayor tiempo de adquisición del experimento. Sin embargo, si se tiene en cuenta que cada experimento permite seguir muchos metabolitos a la vez, este aspecto negativo se ve compensado.

La técnica de RMN aplicada a heteronúcleos (tales como  $^{13}\text{C}$ ,  $^{15}\text{N}$  o  $^{31}\text{P}$ ) es útil debido al hecho de que estos núcleos presentan una gran dispersión en sus señales, dado que el rango de desplazamientos químicos de los mismos es mayor que el de  $^1\text{H}$ . En el caso particular de  $^{13}\text{C}$  o  $^{15}\text{N}$ , como ya se mencionó, los núcleos son menos sensibles y abundantes, por lo cual es conveniente realizar una marca isotópica. Generalmente, se utiliza la estrategia de marcado isotópico selectivo (no uniforme) para el análisis de flujos metabólicos, más que para realizar perfiles metabólicos.

Otra de las desventajas de la técnica es el solapamiento de las señales obtenidas en  $^1\text{H}$

RMN. A fin de minimizar esto se puede combinar esta técnica con una de separación cromatográfica, de manera que el extracto se fraccione, disminuyendo la complejidad de la muestra y por ende del espectro de RMN. Esto es análogo a lo que se realiza cuando se acopla la cromatografía gaseosa a un espectrómetro de masa. La segunda opción para incrementar la resolución espectral es adquirir espectros bidimensionales, distribuyendo las señales a lo largo de dos ejes de frecuencia. Estos métodos bidimensionales pueden además utilizarse para incrementar la sensibilidad de isótopos menos abundantes como  $^{13}\text{C}$ , y establecen conexiones entre las señales de  $^1\text{H}$  y  $^{13}\text{C}$  lo cual es útil para la identificación de los metabolitos.



**Figura 5.** Perfil metabólico correspondiente a una muestra de jugo de tomate (A) espectro de  $^1\text{H}$  RMN en una dimensión y (B) en dos dimensiones (COSY) de la zona de azúcares y alifáticos

La gran cantidad de información que se puede extraer de estos espectros bidimensionales compensa el hecho del mayor tiempo que lleva realizarlos. Estos experimentos explotan las interacciones entre los distintos isótopos detectables por RMN en una molécula y resultan en lo que se denomina correlación homonuclear, donde los dos ejes de frecuencia corresponden al mismo núcleo, por lo general  $^1\text{H}$ , o correlación heteronuclear en donde uno de los ejes de frecuencia corresponde al  $^1\text{H}$  y el otro a  $^{13}\text{C}$ ,  $^{15}\text{N}$  y menos frecuentemente  $^{31}\text{P}$ . Los experimentos de correlación homonuclear son particularmente útiles para realizar perfiles metabólicos, permitiendo que subgrupos de picos que se encuentran juntos en el espectro de una dimensión puedan ser identificados y asignados a compuestos particulares. La correlación heteronuclear también es útil cuando los extractos derivan de tejidos que han sido marcados con  $^{13}\text{C}$  o  $^{15}\text{N}$ .

Con el objeto de realizar una asignación completa de las señales resonantes a compuestos específicos presentes en el jugo de tomate, se realizaron diversos experimentos de RMN bidimensionales. La Figura 5B corresponde a un COSY de la zona de azúcares y alifáticos. Usando la información de conectividad y aprovechando la mayor separación de las señales en la dimensión adicional, muchas de las señales fueron asignadas sin ambigüedad. En los casos en donde la asignación fue dudosa, se adicionaron los compuestos puros al extracto para corroborar así la identidad del metabolito analizado. Se utilizaron las tablas de desplazamientos químicos existentes en la literatura como guía para asignar correctamente las señales.

### **Ejemplo 2: caracterización de plantas transgénicas**

El desarrollo de técnicas para la generación de plantas transgénicas ha revolucionado el área de la biología de plantas. Estas plantas transgénicas pueden tener perturbaciones en las rutas metabólicas debido a la introducción del transgen en el genoma de la planta. La técnica de RMN permite tener un pantallazo general del estado metabólico de las plantas lo cual es de suma utilidad para generar líneas

de plantas que posean específicamente alguna ruta metabólica alterada. Un ejemplo de esto es el análisis de plantas de tabaco que producen constitutivamente el ácido salicílico, las cuales son más resistentes a infecciones. Los perfiles metabólicos que se obtuvieron por RMN mostraron que las plantas transgénicas tenían alterados los niveles de distintos compuestos como flavonoides, azúcares, aminoácidos, etc. Este ejemplo muestra claramente la potencialidad de utilizar la técnica de RMN para estudiar cambios metabólicos.

### **2-Mapas de distribución espacial de metabolitos en tejidos vegetales**

La técnica  $^1\text{H}$  RMN puede ser utilizada *in vivo* para obtener información acerca de la distribución espacial de un número limitado de metabolitos. Este método se utiliza generalmente para el  $\text{H}_2\text{O}$  ya que es la señal más fuerte detectada *in vivo*. Las imágenes brindan información acerca de la anatomía del tejido y el movimiento del  $\text{H}_2\text{O}$  dentro del mismo. También se puede obtener información acerca de la distribución espacial de ciertos metabolitos menos abundantes que el  $\text{H}_2\text{O}$  como ser aminoácidos y carbohidratos. Debido a que es una técnica no destructiva y no invasiva, se pueden monitorear cambios temporales en ciertos metabolitos. Es posible a partir de la misma muestra grabar una serie de espectros *in vivo* y luego analizar en el tiempo cómo es la respuesta metabólica frente a cambios en el estado fisiológico del tejido vegetal.

### **RMN versus MS**

La espectrometría de masa acoplada a cromatografía gaseosa (GC-MS) es la técnica más utilizada en metabolómica. Sin embargo, la técnica de RMN está siendo utilizada como una técnica que brinda información complementaria a la que se obtiene con GC-MS.

La sensibilidad es tal vez el requisito más importante que debe cumplir una técnica analítica para ser utilizada en metabolómica, ya que una alta sensibilidad favorece el análisis rápido de una fracción mayor del metaboloma. La técnica  $^1\text{H}$  RMN, que tiene un límite de detección de 5 nmoles, es varios órdenes de magnitud menos sensible que la técnica GC-MS, cuyo límite

de detección es de  $10^{-12}$  mol. Esta diferencia se incrementa aún más si se tiene en cuenta la técnica de RMN con otros isótopos menos abundantes.

Se estima que con la técnica de GC-MS se puede identificar menos del 5 % de los metabolitos de una célula vegetal si bien todo el metaboloma es potencialmente detectable. Este argumento ignora cualquier diferencia que pudiese existir en la eficiencia de extracción y derivatización de metabolitos cuyas características son muy diversas. Otro obstáculo que impide un análisis completo tanto en RMN como en GC-MS es la dificultad de detectar componentes minoritarios en presencia de señales más intensas.

Existen otros factores que influyen en la elección de una técnica o la otra. Uno de ellos es el procesamiento de las muestras. Como se ejemplifico mas arriba, en el caso de GC-MS, es necesario realizar una extracción con solventes y además derivatizar los compuestos por lo que la preparación de las muestras es más laboriosa y además es más probable que en la muestra no estén representados todos los metabolitos que existen en el tejido de partida. Este no sería el caso de las muestras que son analizadas por RMN, en donde se pueden obtener espectros a partir de una amplia variedad de materiales sin necesidad de realizar extracciones con solventes.

La técnica de RMN tiene una ventaja para el análisis cuantitativo y es el hecho de que los espectrómetros modernos son muy estables en comparación con los GC-MS en donde es necesario realizar calibraciones frecuentes y la variabilidad en los tiempos de retención pueden complicar significativamente el análisis cuantitativo. Ambas técnicas generan usualmente múltiples señales lo cual es una ventaja para la identificación de metabolitos, pero a la vez representa una desventaja en términos de complejidad espectral. Sin embargo, en MS, algunas de estas multiplicidades se deben al fraccionamiento de la molécula ionizada, complicando el análisis cuantitativo. En cambio, para el caso de RMN, las señales múltiples provienen de la misma molécula, lo cual permite una verificación cruzada de los datos de cuantificación de los metabolitos.

Resumiendo, una comparación entre RMN y GC-MS muestra que esta última técnica es más sensible aunque el procesamiento de las muestras es más laborioso y la eficiencia de extracción y derivatización no sería homogénea para todos los metabolitos. Por otro lado, la facilidad con la que se generan los espectros y se cuantifican los metabolitos y la información complementaria (como ser estructural) que suministra hacen de la técnica RMN una herramienta de gran utilidad para los estudios metabolómicos.

### Lectura y sitios recomendados

- Fiehn O. Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology* 48:155-171
- Köckenberger W. Nuclear magnetic resonance micro-imaging in the investigation of plant cell metabolism. *Journal of Experimental Botany*. 2001, 52(356): 641-656.
- Krishnan P, Kruger NJ y Ratcliffe RG. Metabolite fingerprinting and profiling in plants using NMR. *Journal of Experimental Botany*. 2005, 56(410): 255-265.
- Lisec J, Schauer N, Kopka J, Willmitzer L y Fernie AR. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nature protocol*. 2006, 1(1): 387-396.
- Ratcliffe RG y Hill YS. Probing plant metabolism with NMR. *Annu Rev Plant Physiol Plant Mol Biol*. 2001, 52: 499-526.
- Silverstein RM y Webster FX. *Proton magnetic resonance spectrometry*. En: *Spectrometric identification of organics compounds*, 1998, 6° edición, Wiley & Sons.
- Sobolev AP, Segre A, Lamanna R. Proton high-field NMR study of tomato juice. *Magnetic Resonance in Chemistry*, 2003, 41: 237-245.
- Verpoorte R, Choi YH, Kim HK. NMR-based metabolomics at work in phytochemistry. *Phytochem Rev*. 2007, 6:3-14.
- Weckwerth W. *Metabolomics. Methods and Protocols*. *Methods in Molecular Biology Series 358*. Humana Press 2007.
- [www.metabolomicssociety.org](http://www.metabolomicssociety.org) (Sociedad de Metabolómica).
- [www.springeronline.com/sgw/cda/frontpage/0,11855,5-40109-70-34409863-0,00.html](http://www.springeronline.com/sgw/cda/frontpage/0,11855,5-40109-70-34409863-0,00.html) (Sitio web de la revista científica *Metabolomics*, revista oficial de la la Sociedad de Metabolómica).
- [www.metabolomics-nrp.org.uk](http://www.metabolomics-nrp.org.uk) (Norwich Research Park – Metabolomics and plants).

[www.metabolomics.bbsrc.ac.uk](http://www.metabolomics.bbsrc.ac.uk) (Rothamsted: The National Centre for Plant and Microbial Metabolomics).

<http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html> (Max Planck Institute of Molecular Plant Physiology / The Golm Metabolome databases consortium).

[www.noble.org](http://www.noble.org) (Noble Foundation / Plant Biology / The Sumner group).

[www.infometrix.com](http://www.infometrix.com) (Diversos paquetes informáticos para alineamiento de datos cromatográficos y espectroscópicos y análisis multivariados: PCA, SIMCA, etc).

[www.metalign.nl](http://www.metalign.nl) (Software para pretratamiento suavizado, estimación del ruido, corrección de la línea base, alineamiento, etc de datos de GC y LC)

<http://metacyc.org> (La base de datos MetaCyc incluye información sobre las rutas metabólicas, sustratos, reacciones, enzimas, etc de más de 150 organismos diferentes).

# I CAPÍTULO 11

## Metagenómica

O. Mario Aguilar<sup>1</sup> y Daniel H. Grasso<sup>2</sup>

### Introducción

Encaramos este capítulo reconociendo el protagonismo de los microorganismos en la vida de nuestro planeta, desde su origen más remoto. Tenemos el caso de nuestra atmósfera rica en oxígeno resultante de la actividad fotosintética de los primeros organismos microbianos, y aún hoy los microorganismos aportan la mayor capacidad fotosintética del planeta. Cada proceso en la biósfera hace uso de la casi ilimitada potencialidad de los microorganismos para transformar el medio que los rodea. El resto de los organismos vivos dependemos en gran parte de ellos gracias a su capacidad de transformar compuestos en nutrientes asimilables, accesibles y requeridos para nuestra forma de vida. En este sentido mencionemos los siguientes ejemplos: La capacidad de transformar el nitrógeno molecular atmosférico en una forma asimilable por las diferentes formas de vida se encuentra restringida a los microorganismos. Miles de millones de microbios benignos que habitan el intestino nos ayudan a digerir los alimentos, degradar potenciales toxinas y combatir otros microorganismos patógenos. También, nos beneficiamos de aquellos microorganismos que son capaces de mantener el medio ambiente libre de contaminantes y xenobióticos.

La capacidad de adaptación de los procariotas, que les permite prosperar y poblar cada ambiente, aún aquellos de condiciones extremas tales como los respiraderos hidrotermales del fondo del océano y los drenajes ácidos de las minas y efluentes industriales, es consecuencia de su gran diversidad metabólica y fisiológica. Además de su importancia debido a su rol esencial en la biósfera, los microorganismos han sido objeto del desarrollo de numerosas tecnologías que han contribuido a mejorar la calidad de vida. Son empleados industrialmente para producir la mayoría de los antibióticos y muchas otras drogas de uso clínico, también como agentes de remediación de

suelos y agua, mejoramiento de la producción agrícola, producción de biocombustibles, fermentar alimentos para el hombre, etc.

Alrededor de la década de 1950, los microbiólogos disfrutaban la percepción de haber logrado un conocimiento satisfactorio de los microorganismos del suelo. Así, en el año 1931 el microbiólogo Waksman, quien contribuyó significativamente al conocimiento de microorganismos del suelo productores de antibióticos, concluía "... disponemos de una buena información que nos da un panorama claro del mundo microscópico del suelo...". Sin embargo, otros investigadores con iniciativas aisladas y dispersas, persistían en su afán por demostrar la existencia de otras poblaciones microbianas ignoradas, recogiendo evidencias experimentales que cuestionaban esa aparente tranquilidad científica. En 1985, Staley y Konopka a partir de la revisión de los datos relacionados a la capacidad de cultivo de microorganismos de muestras ambientales, concluyen en la llamada "gran anomalía del conteo en placa". Esta anomalía es la discrepancia entre los valores de recuento de células obtenidos mediante microscopía y los valores de recuentos en placa. La conclusión latente, de un conocimiento limitado de los microorganismos del suelo, restringida esencialmente a aquellos que los microbiólogos habían sido capaces de cultivar *in vitro* en sus laboratorios, ha sido tanto abundante como sólidamente documentada durante la última década. El concepto de organismo viable pero no cultivable se hizo evidente con *Vibrio cholerae*, el cual es viable y virulento cuando se aísla de un medio acuático pero crece en cultivo hasta después de su pasaje por el intestino humano o de ratón. Este tipo de evidencias comenzaron a redirigir la atención hacia el mundo de los organismos no cultivables, entre las cuales destacamos dos descubrimientos que han tenido una gran relevancia en esta área. En el primero de ellos se describió mediante curvas de reasociación DNA-DNA la diversidad de las bacterias del suelo, encontrándose que era al menos 100 veces mayor que la que puede ser estimada mediante técnicas dependientes de cultivo. El segundo descubrimiento fue la demostración de *Helicobacter pylori* como agente etiológico de úlceras y cáncer gástrico. Pese a

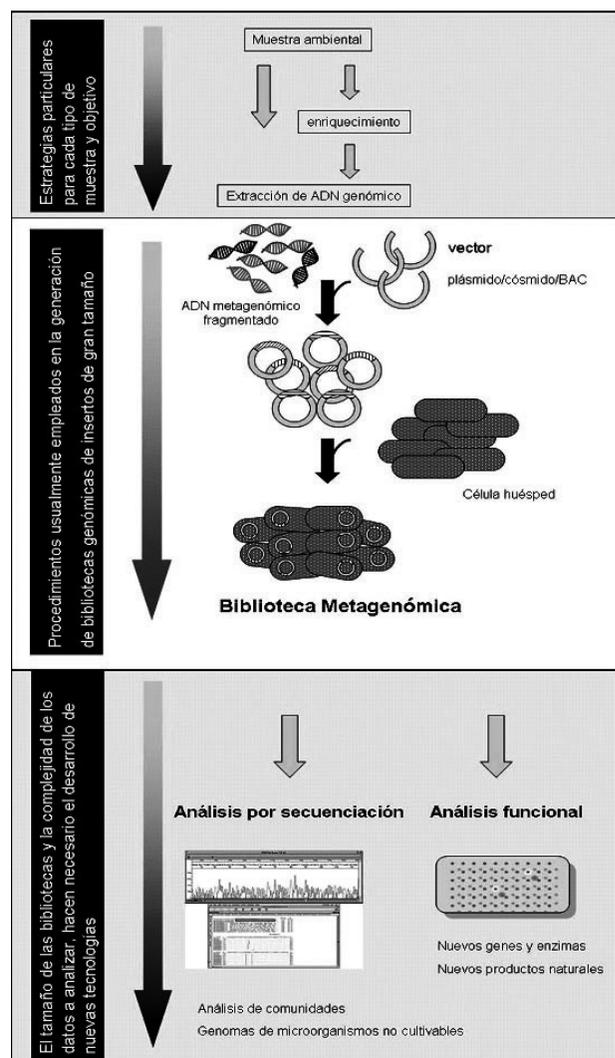
que hace más de un siglo que esta bacteria ha sido observada en la mucosa gástrica, hasta que no fue cultivada no se aceptó su rol en la enfermedad.

La ecología microbiana ha experimentado una gran transformación en los últimos 25 años debido a la introducción de la filogenética, revolucionando la forma de ver a los microorganismos. El primer hito fue el trabajo de Carl Woese quien propuso a los genes rARN como cronómetros evolutivos (Woese, 1987). Posteriormente, Pace y colaboradores utilizaron el análisis de las secuencias de 5S y 16S rARN para describir la diversidad de los microorganismos (cultivables y no cultivables) presentes en muestras ambientales (Pace, 1997). Estos primeros trabajos se realizaron secuenciando directamente el ARN o sus respectivas copias de cADN. El desarrollo de la tecnología de PCR y el diseño de los cebadores que permiten amplificar el gen completo, permitió un acceso mayor a este procedimiento por una comunidad científica más amplia y a una mayor diversidad de nichos ecológicos. La aplicación de la amplificación por PCR de los genes 16S rARN a partir del ADN total de una comunidad, seguido del clonado y secuenciación, ha generado una gran cantidad de datos y ha redefinido la diversidad procariota. Si la muestra ambiental contiene varios tipos diferentes de organismos, se espera entonces encontrar diferentes secuencias de rARN, cuya diversidad será una medida de la complejidad de la comunidad y en el contexto de un árbol filogenético nos dirá quienes son sus miembros. Se ha empleado para caracterizar el perfil poblacional de diversos ambientes naturales. Así, mediante este tipo de análisis se ha demostrado que el suelo es el hábitat de la Tierra más rico en diversidad procariota, aún cuando los métodos basados en el cultivo *in vitro* no ponen de manifiesto esa alta diversidad.

El advenimiento de nuevas tecnologías y procedimientos de análisis de muestras ambientales, metodologías independientes del cultivo de los microorganismos, con potencialidades distintas para revelar componentes de la comunidad microbiana, impulsó un nuevo enfoque de estudio ampliando los horizontes del conocimiento. Se ha denominado con el

término **metagenómica** al análisis del genoma de comunidades independientemente de las tareas de aislamiento y cultivación. El conjunto de métodos desarrollados para acceder al conocimiento fisiológico y genético de comunidades, comprende la extracción directa de ADN de muestras ambientales, clonado de fragmentos de ADN en general o resultante de la amplificación de secuencias del gen 16S rRNA o genes asociados a funciones tales como metabolismo nitrogenado, antibiosis, etc. (Figura 1).

El avance y los logros ya alcanzados por la metagenómica están asociados al desarrollo de capacidades técnicas de secuenciación de



**Figura 1.** Bibliotecas metagenómicas. Esquema del proceso de obtención y análisis de bibliotecas construidas con ADN extraído directamente de muestras ambientales.

ADN con alta eficiencia de generación de datos primarios y a la bioinformática que ha desarrollado herramientas útiles para el procesamiento y análisis de datos.

Los microbiólogos encontraron la oportunidad para dirigir sus esfuerzos a una descripción amplia de la diversidad filogenética de ambientes comunes y exóticos tales como aguas oceánicas de superficie y profundidad, rumen de animales, superficie de redes y cañerías de agua, aguas termales surgentes y suelo. Más allá de permitirles realizar una descripción de la comunidad microbiana, con un nivel de resolución taxonómico preciso, promovió la aceptación y difusión del concepto ecológico de consorcio microbiano para lo cual el estudio de microorganismos aislados y cultivados *in vitro* resulta insuficiente para explicar la funcionalidad y dinámica poblacional.

Mirar más allá de los microorganismos que se pueden aislar, alcanzando al metagenoma permitirá responder no sólo la pregunta **cuáles** son los componentes de una determinada comunidad sino aún más importante **qué hacen** o serían capaces de hacer.

En el desarrollo de este capítulo se describirán las potencialidades de la metagenómica en general, haciendo énfasis en el suelo.

### **El suelo: Un hábitat complejo**

El suelo constituye un soporte para la vida formado por partículas minerales de diferentes formas, tamaños y composición química, compuestos orgánicos en distintas etapas de degradación, e inmersos en este medio encontramos a la biota. A nivel macroscópico se detectan complejos de arcilla, materia orgánica, partículas de arena, todo en su conjunto formando agregados que constituyen una matriz. Los procariotes son los componentes predominantes de la biomasa del suelo, los cuales se encuentran adheridos o adsorbidos a las partículas del suelo. Dada la heterogeneidad de la matriz del suelo, los microorganismos pueden constituir micro-ambientes sobre la superficie y en los poros que se forman. La persistencia y el metabolismo de los microorganismos del suelo dependen de la disponibilidad de agua y nutrientes. El suelo es un ambiente que experimenta cambios drásticos en sus niveles

de humedad, pasando desde una saturación por persistentes lluvias o defectos de drenaje hasta períodos de aridez. Esto sugiere que la composición de la comunidad microbiana experimenta ajustes periódicos en respuesta a esos cambios ambientales, aunque se ignora la manera cuali y cuantitativa del impacto de esos efectores.

El suelo es un reservorio muy importante de carbono, y son los microorganismos quienes desempeñan roles protagónicos en el reciclado del carbono. Asimismo, funciones importantes tales como la captación del nitrógeno molecular, la movilización de fósforo inorgánico, y la formación de nitratos residen en actividades microbianas, los cuales impactan en la fertilidad de los suelos (Daniel, 2005).

Teniendo en cuenta la complejidad fisicoquímica del suelo y la diversidad de funciones que transcurren en el mismo, es posible imaginar al suelo como un cuerpo orgánico con capacidades metabólicas asociadas a poblaciones microbianas diversas, que su vez se complementan entre sí en forma semejante a lo encontrado entre los órganos de un cuerpo. Coincidentemente con este paralelismo, se ha renovado el enfoque del estudio de los genomas dirigido a la consideración global de las interacciones mutuas y la coordinación de la complejidad del sistema, que se lo conoce como biología de sistemas ("system biology"). Actualmente, el estudio del tamaño y la diversidad de las comunidades del suelo, es un desafío para los microbiólogos de suelo, que se proyecta en aspectos aplicados tales como la fertilidad de los suelos y su sustentabilidad.

Se ha estimado que un gramo de suelo contiene aproximadamente  $4 \times 10^7$  células procariontas. El estudio de esta diversidad puede encararse aplicando métodos de cultivo directo o métodos moleculares de metagenómica. Los métodos tradicionales de aislamiento y cultivos en medios nutritivos y condiciones de laboratorio son apropiados pero insuficientes. Se ha estimado que solamente entre 0.1% y 1.0% de las bacterias de suelo son cultivables, lo cual remarca la magnitud de la comunidad aún no examinada. Estos números dan una idea de la magnitud de la complejidad, y del tamaño de la "caja negra" que representa descubrir los

elementos de esa biodiversidad presentes en una “pizca” de suelo. Es importante tener en cuenta que la metagenómica es aplicable no sólo a examinar la diversidad microbiana de un ambiente natural con el objetivo último de lograr el ensamble de genomas a partir de datos fragmentados (“meta análisis”), sino también revelar un espectro amplio de productos génicos con potencialidades biotecnológicas.

### **Métodos de análisis**

El análisis metagenómico implica el aislamiento del ADN de la muestra, clonado de fragmentos de ADN usando vectores plasmídicos y transformación en alguna cepa apropiada de la bacteria *E. coli* (Figura 1). Aparentemente, el procedimiento es experimentalmente muy directo sin embargo el metagenoma comprende a numerosos genomas individuales que contribuyen, con distintos grados de representatividad, al tamaño y la complejidad del mismo. Esto impacta en el tamaño de la muestra a examinar.

La etapa siguiente al clonado molecular constituye la oportunidad de aplicar estrategias alternativas y en general exigen la creatividad del investigador a los efectos de maximizar las posibilidades de alcanzar resultados deseables e interesantes. En términos generales, los procedimientos de análisis de clones interesantes se pueden agrupar en aquellos basados en secuenciación masiva de ADN y en aquellos otros basados en la expresión heteróloga de funciones buscadas. Adoptaremos en este texto el anglicismo *screening* para referirnos a la etapa experimental de análisis de clones e identificación de genes y secuencias de interés.

Para lograr una idea de la envergadura de la tarea de screening hagamos la siguiente referencia a algunas estimaciones numéricas sobre el tamaño del ADN metagenómico.

- Es muy probable que el número de genomas distintos presentes en el suelo refleje el tipo y características del mismo; aquellos dedicados a la agricultura con un buen nivel de materia orgánica y humedad presentarán una diversidad más amplia que un suelo oligotrófico y características extremas de pH, humedad y/o salinidad. A partir de 1 gramo de suelo se puede obtener entre 1-500 ug de ADN, según

el protocolo de extracción aplicado y la naturaleza del suelo. Por otro lado y aplicando un procedimiento basado en cinética de reasociación de ADN, se ha estimado que 1 gramo de suelo comprendería entre 2.000 y 18.000 genomas.

- Según el vector usado para el clonado del ADN metagenómico de suelo, se estima recomendable examinar más de  $10^7$  clones plasmídicos o  $10^6$  clones BACs, con insertos de un tamaño de 5 kb y 100 kb, respectivamente. Para lograr representación significativa de los genomas de miembros raros de la comunidad (menos del 1%) se ha calculado necesario secuenciar 10.000 Gb del ADN extraído de suelo. Por otro lado, se ha estimado en  $10^{13}$  genes en un gramo de suelo provenientes de al menos  $10^3$  especies, lo que indicaría que hay al menos  $10^6$  genes nuevos en 1 gramo de suelo (Schloss y Handelsman, 2006). El objetivo de establecer el metagenoma teniendo en cuenta estos números adquiere un valor relativo cuando se los pone en el contexto de los recursos tecnológicos y financieros necesarios para su abordaje. Actualmente, con el desarrollo de tecnologías de secuenciación de alta generación de datos, basadas en el denominado procedimiento de pirosecuenciación que no requiere de la construcción de bibliotecas metagenómicas, el tiempo requerido para la resolución de un metagenoma adquiere una dimensión real. Sin embargo, la limitación de estos procedimientos consistente en generar lecturas relativamente cortas –aproximadamente entre 100 y 200 nucleótidos- complica la tarea posterior de ensamble de secuencias parciales en un genoma completo.

No obstante, y aún haciendo uso de la tecnología precedente basada en el procedimiento de Sanger, el costo de secuenciación masiva es caro en general, y más aún para nuestro país. Sin embargo, y como se trata en las secciones siguientes de este capítulo, existen estrategias alternativas a la secuenciación de bibliotecas metagenómicas, que “asisten” la búsqueda y resultan útiles para revelar diversidad y descubrir nuevos genes.

### **Screening basado en el análisis del gen 16S rARN**

Este método está basado en el análisis de la

secuencia del gen que codifica para el gen que codifica la subunidad 16S rARN de los ribosomas procarióticos. Todos los microorganismos poseen este ARN que presenta una alta similitud entre ellos pero con diferencias suficientes para su uso como medida de distancias evolutivas. Existe una amplia base de datos de secuencias del 16S rARN de organismos diversos que ha permitido elaborar un árbol filogenético llamado árbol de la vida. Esto permite usar los datos metagenómicos de 16S rARN generado a partir de una determinada muestra ambiental, para posicionar filogenéticamente a las distintas secuencias, y eventualmente inferir la biología y ecología de los mismos. El análisis del (los) gen (es) 16S rARN es económicamente accesible a los laboratorios y es útil para evaluar diversidad microbiana. En particular, ha sido efectivo para revelar la abundancia de especies, y la estructura poblacional de muestras de suelo (Aguilar et al., 2006).

La amplificación mediante la PCR, usando oligonucleótidos cebadores que hibridan las regiones conservadas de los genes 16S rARN de bacterias y arqueas, genera fragmentos que pueden ser clonados y secuenciados. El tamaño del ADN resultante de la amplificación, de aproximadamente 1,4 Kb se inserta en un plásmido vector tipo TA, seguido de transformación de *E. coli*. El ADN usado como molde puede ser el ADN metagenómico o también el proveniente de una biblioteca metagenómica (Furlong et al., 2002). Los datos primarios pueden ser examinados con programas computacionales apropiados que permiten evaluar la predominancia de especies (DOTOUR), y el grado de similitud entre los miembros de dos comunidades (SONS, LIBSHUFF, <http://www.libshuff.mib.uga.edu>).

El método tiene sus limitaciones. Por ejemplo, a- provee un marco filogenético de la comunidad pero da información escasa sobre la funcionalidad de la misma; b- como todo procedimiento basado en PCR, no todos los genes rARN amplifican igualmente, resultando en la sub-estimación de especies, y c- los genes 16S rARN pueden existir en múltiples copias no idénticas.

Es claro entonces que el análisis genético empleando 16S rARN provee valiosa informa-

ción acerca de la diversidad y de la evolución de poblaciones microbiana, sin embargo el gen 16S representa aproximadamente el 0.05% de un genoma procariota. Se ha demostrado que microorganismos que poseen secuencias de 16S rADN idénticas pueden poseer genomas muy distintos y presentar diferentes fisiologías y temperaturas óptimas de crecimiento. Los miembros de una comunidad microbiana pueden diferir enormemente en sus actividades bioquímicas e interacciones, no sólo entre especies sino dentro de una misma especie. Por ende la filogenética nos dice en el mejor de los casos quiénes son los miembros de la comunidad pero muy poco acerca de qué es lo que hace cada miembro. Uno de los principales objetivos de la ecología microbiana es el poder asociar la identidad de los diferentes microorganismos dentro de un hábitat con los procesos que ellos llevan a cabo en ese ambiente. Los métodos metagenómicos, que serán discutidos más adelante, comienzan a dar respuestas a esta segunda cuestión.

### **Construcción de bibliotecas genómicas ambientales**

Muchos de los métodos corrientemente utilizados en la construcción de bibliotecas ambientales, son adaptaciones de las técnicas empleadas para la construcción de bibliotecas de insertos grandes de organismos eucariotas, tales como clonado en cromosomas artificiales (BACs) y lisis celular en tacos de agarosa. Sin embargo, en la construcción de genotecas ambientales se encuentran obstáculos que no se presentan en la construcción de bibliotecas de organismos aislados. La diversidad de nichos ecológicos plantea desafíos interesantes a la hora de diseñar una estrategia adecuada para la obtención de un ADN de calidad apropiada y que el mismo represente la población de microorganismos presentes en ese ambiente. En la actualidad se han preparado diversas bibliotecas metagenómicas de variados ambientes, en esta sección discutiremos algunos de los aspectos críticos a tener en cuenta en su preparación.

La primera biblioteca ambiental se construyó a partir de ADN de microorganismos oceánicos, los que fueron concentrados a partir de

agua de mar antes de la extracción de ADN. Sin embargo en el caso de otros hábitats tales como el suelo, existen procedimientos alternativos, cada uno con sus ventajas y desventajas.

La construcción de una biblioteca metagenómica de suelo se inicia con la toma de muestra. Como las muestras de suelo son heterogéneas, los datos fisicoquímicos tales como tamaño de partícula, tipo de suelo, contenido de agua, pH, temperatura y las plantas que lo cubren son útiles para la evaluación y comparación de los resultados obtenidos en estos estudios. Dado que las poblaciones microbianas son grandes, los volúmenes de muestras pueden ser pequeños. Durante la toma de muestra necesariamente se generan perturbaciones que pueden modificar la composición de las comunidades microbianas del suelo, por lo que es importante disminuir al mínimo el tiempo de transporte y almacenamiento de la misma.

Los métodos de extracción de ADN se pueden dividir en dos categorías: a) lisis directa de las células contenidas en la matriz de la muestra seguido de la separación del ADN de la matriz y los desechos celulares, o b) la separación de las células de la matriz del suelo seguido por lisis celular. El ADN crudo recuperado por ambos métodos se purifica por los procedimientos habituales. La recuperación de ADN aislado de diferentes tipos de suelo utilizando ambos tipos de protocolos van desde menos de aproximadamente 1 µg a 500 µg de ADN por gramo de suelo, sin embargo el rendimiento es entre 10 y 100 veces mayor por lisis directa cuando se compara para una misma muestra. Para lograr la lisis celular directa, se utilizan una combinación de tratamientos enzimáticos, detergente y altas temperaturas. Además del ADN de las células procariontas lisadas, también se recupera ADN extracelular y eucariota.

Los métodos de extracción de ADN basados en la separación de células como una etapa previa a la lisis de las mismas, aunque menos eficientes en términos de cantidad de ADN recuperado son menos dañinos para la integridad del mismo. La separación de los microorganismos de la matriz del suelo se logra mediante suaves fuerzas mecánicas como los procedimientos de mezcla, la rotación del pilón en mortero o químicos, tales como la adición

de resinas de intercambio catiónico, seguido de gradiente de densidad o centrifugación diferencial. En este caso el ADN obtenido es casi exclusivamente procariótico. Además, el ADN recuperado por este método parece ser menos contaminado con compuestos de la matriz compuestos, entre ellos sustancias húmicas. Además, el tamaño medio de los ADN aislados es mayor que el que suele obtenerse mediante lisis directa y, por lo tanto, es más adecuado para la generación de bibliotecas de insertos grandes.

Como los diferentes microorganismos del suelo tienen diferentes susceptibilidades a los métodos de lisis celular, las secuencias presentes en el ADN aislado y en las bibliotecas dependen del método de extracción. No se ha estudiado cuál es el sesgo que se introduce en las bibliotecas debido al método de extracción, sin embargo es de presumir que el ADN aislado por lisis directa represente mejor la diversidad microbiana de una muestra de suelo, ya que este no incluye el paso de separación de las células, por lo tanto serán lisados aún los microorganismos que se adhieren fuertemente a las partículas.

El análisis metagenómico puede requerir de bibliotecas con insertos de tamaño grande o pequeño. Las bibliotecas de insertos pequeños son suficientes en el caso en que el análisis involucre el estudio de un único gen o un pequeño grupo de genes. Por el contrario, se requieren insertos mayores para el estudio de rutas metabólicas completas, procesos complejos u organización genómica. Independientemente del tamaño de inserto requerido, la preparación de ADN es crítica para el éxito del análisis. En el caso de emplear digestiones parciales de ADN para la construcción de la biblioteca se requiere partir de un ADN de un tamaño promedio 3 veces mayor al de los insertos deseados. Por ende si se desea un inserto de 100 kb se deberá partir de un ADN de tamaño promedio > 300 kb. Un ADN de este tamaño es muy vulnerable a las fuerzas de cizalla que se producen en el manipuleo de las soluciones, centrifugaciones, congelado-descongelado, etc. Por ello en estos casos la lisis de las células para la liberación del ADN y su posterior digestión parcial con enzimas de restricción, se realiza en tapones de agarosa.

Dependiendo del tamaño promedio de los insertos, las bibliotecas se construyen empleando diferentes tipos de vectores. Así las de insertos pequeños (menores a 15 kb) emplean vectores plasmídicos, en tanto que las de insertos de hasta 40 kb hacen uso de cósmidos o fósmidos, y BACs para más de 40 kb. Los vectores tipo BACs con número de copia inducible son particularmente útiles, dado que permiten un mantenimiento en la célula huésped a bajo número de copia, pero permiten el incremento del número de copias para facilitar el estudio de expresión.

Si bien el huésped de elección para la construcción y mantenimiento de todas las bibliotecas publicadas es *E.coli*, es fácil de notar que la búsqueda de fenotipos interesantes se verá restringida a aquellos que nos permita expresar este huésped. Se han construido vectores cosmídicos y BACs que permiten la transferencia de bibliotecas producidas en *E. coli* a otras especies huéspedes tales como *Streptomyces* o *Pseudomonas* (Martínez et al., 2004; Wexler et al., 2005).

### **Screening basado en la expresión funcional**

Otra forma de acceder al metagenoma, particularmente cuando se trata de ambientes complejos en cuanto a su diversidad poblacional y a su alta densidad de microorganismos diferentes, es la búsqueda de genes que generen productos definidos e identificables, o asociados a determinadas actividades que se pueden evaluar *in vitro* aplicando bio-ensayos. La ventaja más importante de este abordaje es el desarrollo de un proyecto acotado de menor envergadura que el análisis alternativo dirigido a la secuenciación masiva del metagenoma. Este criterio de screening comprende varias estrategias agrupadas dentro de lo llamado metagenómica funcional. A continuación se describirán algunas estrategias que ilustran formas ingeniosas de abordar el conocimiento del metagenoma con un propósito directo de descubrimiento y uso de propiedades específicas codificadas por el metagenoma.

El usufructo exitoso de actividades microbianas o vías metabólicas requiere del logro de varias etapas críticas. En primer lugar, el

ADN ambiental debe contener el/los gen(es) que codifican las funciones buscadas, y es deseable que las secuencias correspondientes se encuentren representadas en una frecuencia alta en el ADN ambiental; segundo, los procedimientos de preparación del ADN metagenómico y de clonado molecular deben permitir la recuperación e integridad de genes y operones; y finalmente, los genes deben ser detectables genéticamente o fenotípicamente. Resulta obvio al encarar un proyecto de este tipo, que la correcta selección del ambiente apropiado, constituya una de las etapas fundamentales para darle una base probabilística razonable de éxito a nuestro diseño experimental. Por ejemplo, en un trabajo publicado por Rhee et al., (2005), cuyo objetivo fue el aislamiento de genes nuevos codificantes de enzimas termoestables con actividad esterasa, se realizó la búsqueda sobre el metagenoma de muestras de suelo de sitios con aguas termales cuyas temperaturas fueron entre 65 y 90 °C (Rhee et al., 2005). Esta consideración da cierta certeza sobre la presencia de las funciones buscadas en la muestra pero, no asegura que la representación de los genes buscados en el metagenoma, sea suficientemente alta para lograr su recuperación. Por ejemplo, la búsqueda de clones con actividad lipolítica en una biblioteca proveniente de ADN de suelo, resultó en la identificación de un clon entre 730.000 clones examinados. Con el propósito de aumentar la probabilidad de detección de genes de interés se ha recurrido a la inclusión en el diseño experimental de etapas de enriquecimiento previas a la construcción de las bibliotecas génicas. En la mayoría de estos estudios, las fuentes de carbono y/o nitrógeno han sido los criterios selectivos de especies microbianas con genes deseados que son requeridos para su desarrollo en esas condiciones de incubación. El análisis metagenómico de cultivos enriquecidos ha demostrado su potencialidad para el aislamiento de genes determinantes de funciones catalíticas, degradativas, y nuevos antibióticos a partir de suelo y otros ambientes (Entcheva et al., 2001; Gupta et al., 2002; Knietzsch et al., 2003; Daniel et al., 2004; Mori et al., 2008). Sin embargo, se deben considerar sus limitaciones: Las poblaciones mi-

crobianas que contienen los genes deseados no responden eficientemente a las condiciones de enriquecimiento, su crecimiento lento determina baja representación en la muestra previa a la extracción de ADN.

Los métodos de screening funcional son potencialmente útiles para descubrir nuevas variantes de funciones interesantes. Su eficiencia, usando bibliotecas metagenómicas, depende a su vez de la eficiencia y sensibilidad del ensayo *in vitro* y también de la compatibilidad del sistema de transcripción y traducción de la célula hospedadora para transcribir y traducir el/los gen(es) contenidos en el clon transgénico. Aunque la bibliografía demuestra que el hospedador más usado es *Escherichia coli*, el rango de hospedadores se ha ampliado en los últimos años a otros microorganismos tales como *Streptomyces lividans*, *Pseudomonas putida*, y *Rhizobium leguminosarum* (Martínez et al., 2004; Li et al., 2005) e incluso eucariotes (Al Hasani et al., 2003). Esto permitirá un mayor acceso hacia la expresión de un amplio rango de actividades génicas del medio ambiente.

El análisis metagenómico basado en identificación de clones que expresan una determinada función ha sido exitoso cuando se lo ha combinado al uso de mutantes de *E. coli*. Por ejemplo, Daniel y su grupo usaron mutantes de *E. coli* deficientes en sus sistemas antiporter  $\text{Na}^+$  ( $\text{Li}^+$ )/ $\text{H}^+$  para descubrir dos antiporters nuevos a partir del screening de una biblioteca de aproximadamente  $1,5 \times 10^6$  clones. El procedimiento experimental consistió en transferir los clones a *E. coli*, y evaluar el crecimiento de los transformantes en cajas de petri conteniendo un medio de cultivo con 7,5 mM LiCl (Majernik et al., 2001). Aplicando un procedimiento similar se identificaron genes participantes en la biosíntesis de biotina, a partir de ADN de suelo (Entcheva et al., 2001).

Por otro lado, la selección por resistencia a antibióticos ha conducido al aislamiento de nuevas formas de resistencia a tetraciclina y a aminoglicósidos a partir de muestras de la biota de la boca humana, y de suelo, respectivamente. El descubrimiento de resistencia a aminoglicósidos es un buen ejemplo del uso del screening funcional. Se identificaron 9 clones, de los cuales 6 codifican para 6'-acetiltransfe-

rasas que forman un grupo nuevo. Estos genes resultaron del análisis de una biblioteca consistente en 4 Gb de ADN, equivalente a  $10^6$  genes, sugiriendo la obvia y baja representación de dichos genes en la biblioteca. El uso del procedimiento experimental de selección positiva hizo posible su detección que de otra manera –por ejemplo mediante secuenciación– hubiera resultado una tarea casi imposible. No obstante, el número de clones/muestras individuales de una biblioteca que son necesarios examinar, sigue siendo experimentalmente un número grande. Con el objetivo de encontrar resultados en tiempo real y examinar un universo grande, se han diseñado estrategias de alta productividad (en Inglés son referidas como “*high-throughput screen*”), las cuales incorporan sistemas automatizados y robotizados con alta tecnología instrumental.

Otro enfoque es el estudio de antibióticos nuevos en metagenoma de suelo el cual ha expandido nuestra visión de las comunicaciones entre comunidades microbianas. Se ha encontrado que concentraciones sub-inhedoras de crecimiento inducen respuesta del tipo quórum sensing, aún cuando los compuestos no muestran similitud con las conocidas homoserina lactonas que han sido descritas como inductores naturales en bacterias. Esto sugiere la existencia de un diálogo comunicativo entre microorganismos que activan funciones en el entorno. Además, este hallazgo ha sido adoptado para el screening de moléculas que funcionan como inductoras de quórum sensing y eventualmente también como antibióticos. Esta oportunidad condujo al diseño de un procedimiento de screening con alta eficiencia cuantitativa de análisis de clones e identificación de compuestos que funcionan como inductores de genes que se encuentran bajo el control de un promotor sensible a quórum sensing. Este sensor consiste en el promotor del gen *luxR* fusionado al gen reportero *gfp*, residente en un plásmido replicativo en una variante de *E. coli* que no induce quórum sensing. En el caso que el ADN metagenómico exprese un inductor del promotor *luxR*, se sintetizará la proteína GFP revelándose fluorescencia. Este procedimiento acoplado a *cell sorting* activada por fluorescencia, permite capturar aquellos clones potencial-

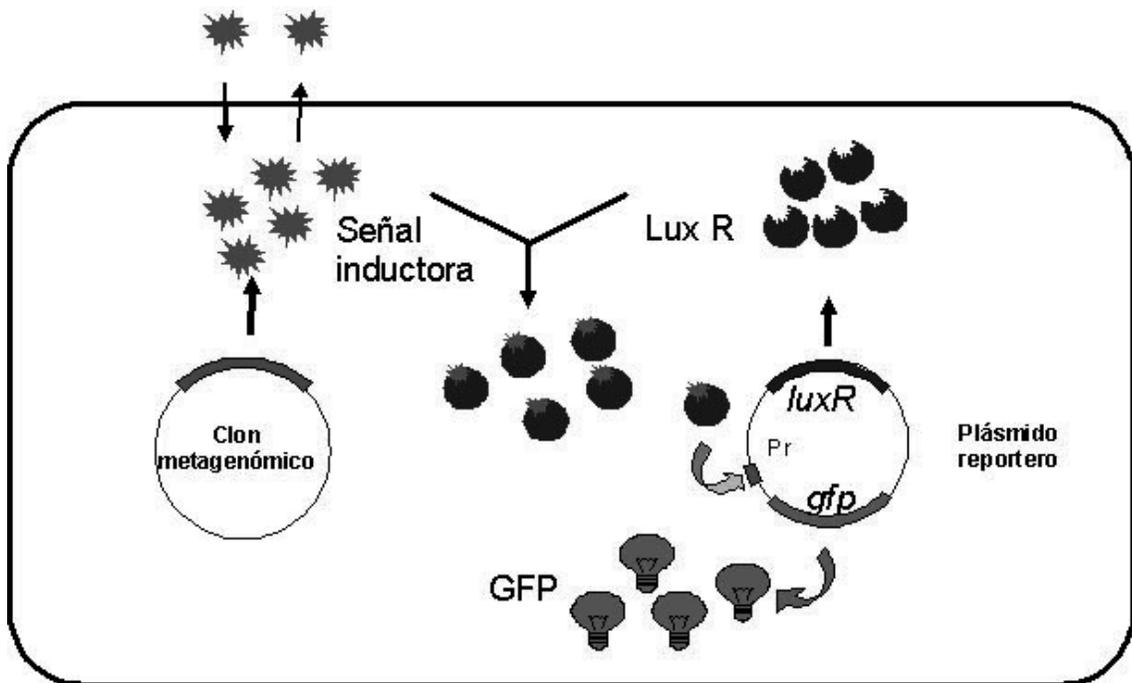
mente productores de compuestos de *quorum sensing* (Figura 2) (Lynn et al., 2005; Uchiyama et al., 2005).

Otro método de enriquecimiento se basa en el uso de isótopos estables, en el cual se administra -como única fuente de carbono- un sustrato marcado con  $^{13}\text{C}$  a una comunidad de microorganismos provenientes de suelo. Aquella comunidad bacteriana capaz de utilizar ese sustrato incorporará  $^{13}\text{C}$  en sus macromoléculas, particularmente en ADN, determinando que el ADN será más denso que el ADN normal de la bacteria. El procedimiento continúa con la ultracentrifugación del ADN en gradientes de densidad de CsCl para separar el ADN marcado del ADN normal o no marcado. Este ADN puede separarse de la columna de gradiente, someterse a diálisis y posteriormente concentrarse por precipitación etanólica. El ADN se fragmenta y clona en vectores plasmídicos. La biblioteca resultante puede analizarse mediante un screening funcional, o haciendo uso de sondas dirigidas a revelar un determinado gen. El método tiene potencialidad para fraccionar

las poblaciones provenientes del suelo concentrando aquellas asociadas a una función (por ejemplo degradación de un compuesto hidrocarburo que es asimilado). Sin embargo, el método tiene sus limitaciones principalmente generadas por el llamado "cross feeding", en el cual bacterias carentes de esa propiedad pueden desarrollar en el medio a expensas de metabolitos provistos por otras bacterias (Radajewski et al., 2000)

### Lo no cultivable se vuelve cultivable

El estudio metagenómico del biofilm que se forma sobre las aguas que drenan de una mina de hierro abandonada, ubicada en la Iron Mountain en Richmond, California ha permitido establecer la estructura de la comunidad y las funciones metabólicas y biogeoquímicas. Esa agua de drenaje constituye un medio extremadamente adverso, su acidez es de pH 1.0 y posee un alto contenido en metales. También es rica en pirita ( $\text{FeS}_2$ ). No hay fuentes asimilables de carbono o nitrógeno más allá de las que pueden derivar del aire. Tyson et al. (2004)



**Figura 2.** Aislamiento de clones provenientes de un biblioteca metagenómica, que expresan productos génicos inductores/activadores de quórum *sensing*. LuxR es un producto génico que interacciona con moléculas efectoras producidas por microorganismos, activador de un promotor sensible fusionado al gen reportero para GFP (green fluorescent protein) facilitando su detección.

secuenciaron una biblioteca construida con ADN extraído del biofilm, y lograron ensamblar a partir de datos metagenómicos los genomas de miembros del grupo *Leptospirillum* y *Ferroplasma* tipo II, también una sustanciosa información de otros miembros de la comunidad. El análisis de estos resultados fue muy sugestivo. Los autores notaron que no habían detectado la presencia de *Leptospirillum ferrooxidans*, la especie fijadora de N<sub>2</sub>. Por el contrario, el genoma de la comunidad reveló la presencia de miembros pertenecientes a los grupos II y III de *Leptospirillum*. El género *Leptospirillum* ha sido sub-dividido en 3 grupos, de los cuales aún no se había logrado cultivar a representantes del grupo III. El estudio también reveló la presencia de genes *nif* (genes de fijación biológica de nitrógeno) que los autores asociaron al genoma de *Leptospirillum* grupo III. Estos datos permitieron concluir que la contribución de esta población al consorcio del biofilm de la mina ácida, es el aporte de nitrógeno asimilable, y además, el conocimiento de las capacidades metabólicas inferidas de los datos metagenómicos guió los pasos siguientes dirigidos a lograr la cultivación en el laboratorio de *Leptospirillum* grupo III y a la definición de la especie *L. ferrodiazotrophum* (Tyson et al., 2005). De la misma manera, estudiando los datos resultantes, se ha especulado que *Ferroplasma* y *Leptospirillum* sp derivarían la energía a partir de la oxidación del hierro. Además, se encontró que todos los genomas muestran genes asociados con la remoción de elementos potencialmente tóxicos, tales como sistemas de flujo de protones y bombas de exclusión de metales (Tyson et al., 2004).

El estudio de las comunidades en el drenaje de aguas de la mina ácida de Richmond, es un modelo de complementariedad de funciones asignables a miembros del consorcio microbiano. Es posible imaginar que la proyección de este modelo de estudio a otros sistemas ambientales no resulte fácil. La extrema adversidad de la mina ácida limita la diversidad de genomas en la comunidad y facilitó el análisis

***El mensaje intrínseco de este descubrimiento de componentes importantes de las comunidades del suelo, aplicando un análisis metagenómico, es su uso como guía en***

***la formulación del medio de cultivo y logro de su cultivación in vitro.***

### **Potencialidades de la metagenómica**

Las posibilidades que este nuevo campo ofrece son alucinantes. Descifrar la enorme gama de procesos e interacciones que caracterizan las comunidades microbianas en la Tierra puede conducir a avances en la salud humana, la mejora de la comprensión a gran escala de los cambios climáticos y atmosféricos, los métodos para obtener cultivos más fuertes y nutritivos, nuevos enfoques para la limpieza de la contaminación ambiental, y el desarrollo de nuevas fuentes de energía renovable. Estos son sólo algunos ejemplos de las muchas posibles aplicaciones prácticas de la metagenómica.

El ser humano siempre vivió en un mundo dominado por los microbios, y la estrecha relación entre los microbios y los seres humanos es un tema antiguo. Las células microbianas que viven en el cuerpo humano adulto son diez veces más numerosas que las células humanas. Los genomas microbianos de las comunidades que viven dentro y en el cuerpo humano (el microbioma humano) contiene muchos más genes que el genoma humano. Estudiar el microbioma humano puede dar lugar a valiosas nuevas herramientas en nutrición humana y animal, el descubrimiento de medicamentos, y en medicina preventiva. Este estudio también puede ampliar mucho la profundidad de nuestra comprensión de enfermedades complejas tales como obesidad, el cáncer y ciertas enfermedades inmunológicas, como el asma (Turnbaugh et al., 2006).

La inmensa mayoría de nuestros microorganismos asociados viven en el intestino. Estos 10 a 100 billones microorganismos desempeñan funciones tales como la extracción de nutrientes y calorías de componentes de nuestras dietas y la síntesis de vitaminas esenciales y aminoácidos. Las bacterias intestinales también ayudan a detoxificar productos químicos potencialmente nocivos que pueden estar presentes en lo que comemos. Algunos de los microbios que viven en y sobre el cuerpo humano desempeñan un papel fundamental en la defensa contra agentes patógenos. Esta relación mutuamente beneficiosa nos ayuda a prote-

germos de enfermedades al tiempo que se da a los microbios un lugar para vivir. El uso de metagenómica para obtener un conocimiento más profundo de las comunidades microbianas en el cuerpo humano podría ser inmensamente valioso en la comprensión tanto de microbios dañinos como de beneficiosos y puede conducir a formas más efectivas de diagnosticar, tratar y prevenir enfermedades.

El enfoque metagenómico ofrece la posibilidad no sólo de analizar la diversidad filogenética de diversos ambientes y biofilms, sino también de localizar los genes y operones que codifican propiedades de interés biotecnológico. La metagenómica ha dado lugar al descubrimiento y caracterización de una amplia gama de biocatalizadores, que revela mucho acerca de la diversidad natural de las enzimas y los factores que influyen en sus funciones. Uno de los principales enfoques para la detección de nuevos biocatalizadores implica la detección funcional lo que requiere la expresión de genes heterólogos generalmente en *Escherichia coli*. Mediante el enfoque metagenómico se estudian hábitats tan diversos como intestinos de termitas, el rumen de rumiantes como ciervos o vacas, las muestras de suelo de los desiertos y glaciares, así como hábitats extremos como géiseres, mar abierto de agua o chimeneas hidrotermales de aguas profundas. Es de esperar que la composición genómica de las poblaciones microbianas en estos hábitats se diferencien unos de otros, y con ella los biocatalizadores y biomoléculas que componen las respectivas bibliotecas metagenómicas. La metagenómica industrial se centra principalmente en procariotes, ya que sus genomas puede ser objeto fácilmente de las herramientas de selección funcional disponibles en la actualidad, y porque se supone que la mayor diversidad se encuentra entre estos microorganismos. Las enzimas se utilizan en una amplia gama de aplicaciones e industrias. Su versatilidad permite su uso tanto en los procesos para degradar polímeros naturales entre ellos almidón, celulosa y proteínas, así como para las síntesis enantioselectiva asimétrica de productos químicos, aunque hay muchas más aplicaciones.

En el año 2000 Rondon y col. publican las primeras bibliotecas de suelos clonados en

vectores de tipo BAC. Se generaron dos bibliotecas metagenómicas una con un tamaño de inserto promedio de 27 kb y otra de 44.5 kb. Pese a que el tamaño de inserto no es mayor al de una biblioteca convencional construida con vectores de tipo cósmido o fago lambda, es la primera publicación en la que se informa el clonado de fragmentos de ese tamaño a partir de las poblaciones microbianas de suelo. A partir de estas bibliotecas logran identificar clones que expresan fenotipos tales como, lipasas y amilasas. De igual modo Brady y col, recuperaron a partir de una biblioteca cósmica de ADN de suelo el conjunto de genes necesarios para la síntesis de violaceína, un antibiótico de amplio espectro,.

Aunque las bibliotecas metagenómicas constituyen en la actualidad la herramienta más poderosa para evaluar la diversidad funcional de comunidades microbianas naturales, no cubren los genomas de baja abundancia en ambientes complejos como los suelos. Como resultado, la frecuencia de clones con un fenotipo deseado en una biblioteca puede ser muy baja. Esto implica la búsqueda y selección en miles de clones, lo que significa una tarea larga y tediosa. En la actualidad se dispone de equipos con tecnologías que facilitan la recolección de colonias, y la inoculación en placas de numerosos clones al mismo tiempo.

Mientras que la industria de alimentos y de detergentes se concentran en un número limitado de reacciones enzimáticas y sustratos, las industrias química y farmacéutica trabajan con miles de moléculas química y estructuralmente diversas, y la producción de cada uno de estos requiere soluciones enzimáticas individuales. En consecuencia, debido a la riqueza de biocatalizadores potencialmente útiles, el uso de recursos microbianos es cada vez más difundido en las industrias químicas y son considerados indispensables para la química orgánica moderna. Es obvio que existe una amplia demanda de nuevos enzimas y biocatalizadores, y la metagenómica aparece como una de las tecnologías con mayor potencialidad para proveer de las moléculas necesarias.

Es creciente la conciencia y la percepción global de que la disponibilidad de combusti-

bles fósiles no renovables no es sostenible en el futuro próximo, y se deberá superar esa dependencia energética. Además, las emisiones de gases de invernadero como resultado de la quema de combustibles fósiles son una de las causas principales del calentamiento global. Estos factores determinan la búsqueda de combustibles renovables, y amigables para el medio ambiente sea una de las principales prioridades para el mundo. Una fuente de energía emergente es el etanol (alcohol de grano)-un biocombustible de alto octanaje de derivado de maíz, caña de azúcar, o de otro tipo de fuentes agrícolas. El etanol celulósico se fabrica a partir de la celulosa que se encuentra en el común de los desechos agrícolas tales como la fibra de maíz, tallos de maíz, paja de trigo, y otras como la biomasa derivada del mijo y el miscanthus. Pero el proceso que convierte la celulosa a partir de residuos agrícolas a etanol utilizable depende de un ingrediente esencial: las comunidades microbianas. En primer lugar, varios tipos de microorganismos deben trabajar en conjunto para transformar la celulosa a partir de desechos agrícolas en azúcares. Luego, los azúcares son fermentados-también por los microbios- para producir etanol.

Las vacas, con la ayuda de las bacterias, son capaces de convertir las fibras vegetales (celulosa) en energía, pero este es un proceso complejo para mimetizarlo para la producción de biocombustibles.

La enzima que permite que una vaca digiera los pastos y otras fibras vegetales puede ser usada para convertir otras fibras vegetales en azúcares simples. Transformar las fibras vegetales en azúcar requiere de tres enzimas. Estas tres enzimas han sido recientemente introducidas en una variedad de maíz denominada Spartan III, que deriva de dos versiones anteriores. La primera versión contiene una enzima que procede de un microbio que vive en agua manantial caliente, y que es capaz de transformar el gran polímero de celulosa en fragmentos de menor tamaño. La segunda versión contiene un gen de un hongo natural que codifica para una enzima capaz de clivar estos fragmentos de celulosa en disacáridos. En la última versión se introdujo el gen codificante de una enzima capaz de catalizar la hidrólisis

de los disacáridos en azúcares simples. Este gen fue aislado a partir de microbios presentes en el rumen de vaca. Mediante ingeniería genética este gen ha sido modificado de forma tal de dirigir el producto de su expresión hacia vacuolas, de esta forma las enzimas digestivas permanecen almacenadas hasta su cosecha (Sticken, 2008). Estos son algunos ejemplos que ponen de manifiesto la potencialidad de los enfoques metagenómicos tanto en su propósito biotecnológico como en el análisis de las comunidades ambientales. En definitiva, la metagenómica surge hoy como una disciplina nueva que nos permitirá ampliar nuestro conocimiento de la microbiología hasta dimensiones aún inimaginables.

### Lecturas recomendadas

- Aguilar O.M., M. V. López, M. Donato, B. Morón, M. E. Soria-Díaz, C. Mateos, A. Gil-Serrano, C. Sousa, and M. Megías. 2006. Phylogeny and nodulation signal molecule of rhizobial populations able to nodulate common beans -other than the predominant species *Rhizobium etli*- present in soils from the Northwest of Argentina. *Soil Biology and Biochemistry* 38:573-586.
- Al-Hasani, K., Simpfendorfer, K., Warden, H., Vado-las, J., Zaibak, F., Villain, R., and Ioannou, P.A. 2003. Development of a novel bacterial artificial chromosome cloning system for functional studies. *Plasmid* 49:184-187.
- Brady SF, Chao CJ, Handelsman J, Clardy J. 2001. Cloning and heterologous expression of a natural product biosynthetic gene cluster from cDNA. *Org Lett* 3:1981-1984.
- Daniel, R. 2005. The metagenomics of soil. *Nature* 3:470-478.
- Entcheva P., Liebl, W., Johann, A., Hartsch, T., and Streit, W.R. 2001. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl. Environ. Microbiol.* 67:89-99.
- Environmental Microbiology* [www.blackwell-synergy.com/toc/emi/7/12](http://www.blackwell-synergy.com/toc/emi/7/12)
- Furlong, M.A., Singleton, D.R., Coleman, D.C. and Whitman, W.B. 2002. Molecular and Culture-Based Analyses of Prokaryotic Communities from an Agricultural Soil and the Burrows and Casts of the Earthworm *Lumbricus rubellus*. *Appl. Environ. Microbiol.* 68:1265-1279.
- Knietsch, A., Bowien, S., Whited, G., Gottschalk, G. and Daniel, R. 2003. Identification and Charac-

- terization of Coenzyme B12-Dependent Glycerol Dehydratase- and Diol Dehydratase-Encoding Genes from Metagenomic DNA Libraries Derived from Enrichment Cultures. *Appl. Environ. Microbiol.* 69:3048-3060.
- Lynn, L., Williamson B.R., Borlee, R. Schloss P.D., Guan C., Allen, H.K., and Handelsman, J. 2005. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum sensing biosensor. *Appl. Environm. Microbiol.* 71:6335-6344.
- Majernik, A., Gottschalk, G., Daniel, R. 2001. Screening of Environmental DNA Libraries for the Presence of Genes Conferring Na super(+)(Li super(+))/H super(+) Antiporter Activity on *Escherichia coli*: Characterization of the Recovered Genes and the Corresponding Gene Products . *J. Bacteriol.* 183:6645-6653.
- Martínez, A. Kolvek, S.J., Yip, C.L.Y., Hopke, J., Brown, K.A., MacNeil, I.A., and Osbourne, M.S. 2004. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expresión hosts. *Appl. Environm. Microbiol.* 70:2452-2463.
- Mori, T., Mizuta, S., Suenaga, H. And Miyazaki, K. 2008. Metagenomic Screening for Bleomycin Resistance Genes. *Appl. Environm. Microbiol.* 74:6803-6805.
- Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276:734-740.
- Radajewski, S., Ineson, P., Parekh, N.R., and Murrell, J.C. 2000. Stable isotope probing as a tool in microbial ecology. *Nature* 403:646-649.
- Rhee, J.K., Ahn, D-G, Kim, Y-G and Oh, J-W. 2005 New Thermophilic and Thermostable Esterase with Sequence Similarity to the Hormone-Sensitive Lipase Family, Cloned from a Metagenomic Library. *Appl Environm Microbiol* 71:817-825.
- Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C *et al.*. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* 66:2541-2547.
- Staley J.T., and Konopka, A. 1985. Measurement of in situ activity of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.* 39:321-346.
- Sticklen, M.B. 2008. Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol, *Nature Reviews Genetics* 9, 433-443.
- Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444:1027-1031.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428:399-428.
- Tyson, G.W., Lo, I., Baker, B.J., Allen, E.E., Hugenholtz, P., and Banfield, J.F. 2005. Genome-Directed Isolation of the Key Nitrogen Fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an Acidophilic Microbial Community. *Appl Environm. Microbiol.* 71:6319-6324.
- Uchiyama, T., Abe, T., Ikemura, T. and Watanabe, K. 2005. Substrate-induced gene-expression of environmental metagenome libraries for isolation of catabolic genes. *Nature* 23:88-93.
- Waksman, S.A., and Starkey, R.L. 1931. The soil and the microbe. John Wiley, New York, N.Y.
- Wexler, M., Bond, P.L., Richardson, D.J., and Andrew W. B. Johnston. 2005. A wide host-range metagenomic library from a waste. *Environmental Microbiology* 7:1917–1926
- Woese, C.R. 1987. Bacterial evolution. *Microbiol. Rev.* 51:221–271.

## I. CAPÍTULO 12

### Bioinformática aplicada a la biotecnología vegetal

N Paniego; R Heinz; P Fernández; V Lia;  
C Fusari

#### Introducción

La Bioinformática es una disciplina científica que ha evolucionado rápidamente en los últimos años respondiendo al avance y a las necesidades de procesamiento, almacenamiento y análisis de datos biológicos derivados de las tecnologías asociadas a las *ómicas*, para generar nueva información y conocimientos en el área de la biología. El campo de la bioinformática es multidisciplinario abarcando el desarrollo de bases de datos, el alineamiento de secuencias, la predicción de estructuras proteicas, la construcción de árboles filogenéticos, entre otros. En este capítulo nos dedicaremos específicamente a la bioinformática aplicada a la biotecnología vegetal, particularmente a cultivos agronómicos y plantas modelo, con el propósito de guiar al estudiante de un curso inicial de biotecnología vegetal o de biología molecular de plantas en la aplicación de los conceptos y las herramientas básicas de la bioinformática, complementando los conceptos introducidos en la primera edición de este libro. Este material no pretende cubrir la totalidad de los recursos existentes para la materia, el objetivo es motivar el interés de los estudiantes en la exploración y el uso de recursos y métodos computacionales básicos para el desarrollo de su actividad científica o profesional futura. Para ello, describiremos en primer lugar algunas bases de datos específicas de iniciativas genómicas de plantas e introduciremos el concepto de anotación de datos genómicos. La segunda parte del capítulo esta dedicada a la bioinformática aplicada a estudios de expresión, la búsqueda de marcadores moleculares funcionales y el análisis de su variabilidad genética.

#### Bases de datos generales y específicas de especie

El uso de tecnologías de bases de datos ha

sido adoptado por la comunidad científica desde el inicio de las iniciativas genómicas para facilitar la organización y la difusión de los datos. Esto hace que se distingan distintas categorías de bases de datos, entre las principales están aquellas que son repositorios públicos a gran escala y bases de datos específicas de iniciativas comunitarias

Los repositorios públicos a gran escala son bases de datos estables, generalmente mantenidas por agencias gubernamentales, que archivan principalmente información estática, un ejemplo típico es Genbank. La información disponible en esa base de datos en particular puede ser accedida a través de ENTREZ, un buscador que combina la interrogación simultánea de 35 bases de datos disponibles en el sitio NCBI. Algunas de éstas son bases de datos secundarias que agregan información a los datos primarios (secuencias nucleotídicas), entre las cuales podemos mencionar Gene, UniGene, HomoloGen o RefSeq.

Dentro de las bases de datos comunitarias están aquellas que almacenan datos derivados de estudios sobre especies modelo, generalmente focalizando en una especie o en un grupo de especies relacionadas. La primera base de datos de esta categoría establecida para especies vegetales es "The Arabidopsis Information Resource" (TAIR), que reúne la información asociada a esta especie modelo en relación a secuencias, genes y proteínas, marcadores, alelos, mapas, vías metabólicas, literatura, protocolos, microarreglos, ontología de genes, anotaciones, disponibilidad de germoplasma/semillas y herramientas de análisis. En los últimos años, ha surgido una nueva generación de bases de datos que integran la información de muchas especies para permitir la comparación de genomas, entre ellas está Gramene que integra recursos para la comparación de mapas genéticos y físicos de arroz y otras especies de gramíneas; LIS permite la comparación genómica y de transcritos de distintas especies de leguminosas; finalmente Phytozome (<http://www.phytozome.net>: Tool for green plant comparative genomics) reúne la información competa de 14 genomas vegetales, agrupamientos de genes ortólogos, parálogos y familias de genes y herramientas de análisis

como BLAST para hacer comparaciones contra los proteomas de cada una de las especies representadas en la base de datos. La Tabla 1 condensa un número importante de bases de datos específicas junto con sus accesos Web para facilitar la exploración de las mismas.

Asimismo, existen bases de datos específicas de patógenos que afectan cultivos agrónomicamente importantes. Estas bases aportan información sobre la secuencia del genoma de los patógenos, la cual puede ser asociada a la información derivada de estudios funcionales de transcriptos inducidos durante la interacción con la planta huésped, permitiendo el diseño de estrategias de mejoramiento de los cultivos para lograr resistencia. Entre estas bases de datos se pueden mencionar a: Phytophthora Functional Genomics Database y PathoPlant.

### **Anotación funcional de los datos genómicos**

El proceso de anotación de los datos genómicos a nivel de proteínas consiste en tratar de deducir a partir de los datos de secuenciación nucleotídica, las características funcionales del genoma de un organismo, explorando y describiendo los niveles intermedios de organización como funciones y procesos moleculares, celulares, fisiológicos, compartimentalización de las funciones, etc. Este proceso de anotación demanda una gran integración de los datos e información disponible para la especie o para especies relacionadas, haciendo uso de herramientas de la bioinformática para la comparación y la extracción de datos y de las bases de datos generales y específicas, del conocimiento biológico acumulado en publicaciones y de los análisis transcriptómicos o genómicos disponibles.

Una de las primeras etapas en este proceso es la de tratar de asignar una función molecular a la mayoría de las secuencias incógnitas mediante la comparación con genes de función conocida. La forma más precisa para hacer esta comparación es a nivel de productos de genes, o sea, trabajando con las secuencias de aminoácidos obtenidas a partir de la traducción de las secuencias nucleotídicas a los seis marcos de lectura posibles. Este método de comparación de secuencias, que es central

en el proceso de anotación funcional, se basa en la característica que tienen las proteínas homologas de conservar la misma función y se sustenta con bases de datos de secuencias de proteínas curadas cuidadosamente como por ejemplo SwissProt o PIR.

No obstante, puede asumirse en general que las bases de datos de proteínas asociadas a organismos modelo se encuentran correctamente anotadas y pueden usarse como referencia para inferir posibles funciones moleculares. Sin embargo, para que los procesos de anotación dentro y entre especies en organismos no-modelo sean lo más eficiente y comparable posible, lo óptimo es referir la anotación asignada a vocabularios estructurados u ontologías. Existen diferentes iniciativas para establecer vocabularios controlados, una de las más frecuentemente usadas es la "Ontología de genes" (Gene Ontology, GO), que establece tres categorías o jerarquías: (1) las funciones moleculares, (2) los procesos biológicos y (3) la localización sub-celular o componente celular.

El proceso de anotación es un proceso continuo que debe ser constantemente actualizado con el fin de agregar nueva información y refinar la información existente.

### **Anotación de secuencias en especies de plantas no-modelo**

Acorde a lo mencionado, la anotación genómica llevada a cabo en la mayoría de los proyectos y en especial en los proyectos de pequeña envergadura dedicados a la caracterización de transcriptos o ESTs, se basan en el uso del algoritmo BLAST. La calidad de los resultados en todos los casos depende de los parámetros definidos para recuperar secuencias similares en la comparación usando BLASTX o BLASTP, recomendándose como parámetros óptimos E valores menores de  $e^{-10}$ , identidades mayores a 30% y coberturas mayores o iguales a 50%.

Las bases de datos contra las cuales se realizan las comparaciones juegan también un rol importante en relación al número y calidad de los "hits" recuperados. Algunas de las bases de datos de nucleótidos y proteínas más utilizadas para la anotación de secuencias de plantas son los Índice de Genes (DFCI: <http://compbio.dfci.harvard.edu/tgi/>), Unigenes (NCBI) y PlantGDB

(<http://www.plantgdb.org/>). Las bases de datos de proteínas que se utilizan en general son: NR, SwissProt, Trembl, Uniprot, Uniprotsw, Uniprot-tr, Uniref100 y bases de proteínas de proyectos específicos (Tabla 1).

En base al mejor valor de similitud obtenido en la comparación, se asigna una función génica probable a cada secuencia analizada. Posteriormente, dicha anotación se mapea contra una base de datos de Gene Ontology para finalmente anotar los productos de genes acorde a las normas del Consorcio de Ontología Génica (GO). Otros vocabularios controlados complementarios a GO, son la Comisión de Enzimas (Enzyme Comisión, EC)

que adjudica una numeración basada en una clasificación esquemática de enzimas según la reacción que se encuentran catalizando y el Consorcio KEGG (The Kyoto Encyclopedia of Genes and Genomes) que genera anotación de vías metabólicas para aquellas especies que presentan su genoma secuenciado o en vías finalización. La anotación funcional basada en ontologías puede realizarse usando software libre por ejemplo Blast2GO o GOtcha.

### Expresión de Genes

El análisis de patrones de expresión consiste en la identificación de todos los transcritos presentes en una determinada muestra

**Tabla 1:** Bases de datos genómicas específicas de especie.

Bases de datos generales	URL
NCBI Plant Genomes Central	<a href="http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html">www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html</a>
PlantGDB: plant genome database	<a href="http://www.plantgdb.org">www.plantgdb.org</a>
MIPS plants databases	<a href="http://mips.gsf.de/proj/plant/jsf/genomes.jsp">mips.gsf.de/proj/plant/jsf/genomes.jsp</a>
Bases de datos multiespecie para genómica comparativa	
Phytozome: comparative genomics of plants	<a href="http://www.phytozome.net">www.phytozome.net</a>
Gramene: a Resource for Comparative Grass Genomics	<a href="http://www.gramene.org">www.gramene.org</a>
LIS: Legume Information System	<a href="http://www.comparative-legumes.org">www.comparative-legumes.org</a>
SALAD: Surveyed Alignment and Associating Dendrogram	<a href="http://salad.dna.affrc.go.jp/salad">salad.dna.affrc.go.jp/salad</a>
SGN: Solanaceas genomic network	<a href="http://www.sgn.cornell.edu">www.sgn.cornell.edu</a>
Bases de datos de genomas de iniciativas genómicas completas y parciales	
TAIR: The Arabidopsis Information Resource	<a href="http://www.arabidopsis.org">www.arabidopsis.org</a>
Rice Annotation Project Database (RAP -DB)	<a href="http://rapdb.dna.affrc.go.jp">rapdb.dna.affrc.go.jp</a>
Grain Genes 2.0: a Database for Triticaceae and Avena	<a href="http://wheat.pw.usda.gov/GG2/index.shtml">wheat.pw.usda.gov/GG2/index.shtml</a>
MaizeDB	<a href="http://www.maizegdb.org">www.maizegdb.org</a>
Medicago truncatula: a model for legume research	<a href="http://www.medicago.org">www.medicago.org</a>
Lotus japonicus genome	<a href="http://www.kazusa.or.jp/lotus">www.kazusa.or.jp/lotus</a>
Wheat Applied Genomics	<a href="http://maswheat.ucdavis.edu">maswheat.ucdavis.edu</a>
SoyBase and the Soybean Breeder's Toolbox	<a href="http://soybase.org">soybase.org</a>
SoyMap: An integrated map of soybean for resolution and dissection of multiple	<a href="http://www.soymap.org/">www.soymap.org/</a>
Brassica rapa genome	<a href="http://www.brassicarapa.org/BRGP/index.jsp">www.brassicarapa.org/BRGP/index.jsp</a>
Compositae genomics project	<a href="http://compgenomics.ucdavis.edu/compositae_index.php">compgenomics.ucdavis.edu/compositae_index.php</a>
Cotton Marker Database	<a href="http://www.cottonmarker.org">www.cottonmarker.org</a>
Dendrome: a collection of forest tree genome databases and other forest genetic	<a href="http://dendrome.ucdavis.edu">dendrome.ucdavis.edu</a>
International Grape Genome Project	<a href="http://www.vitaceae.org">www.vitaceae.org</a>
Populus Genome Integrative Explorer	<a href="http://www.popgenie.db.umu.se">www.popgenie.db.umu.se</a>
Bases de datos de diversidad	
PanZea: Molecular and Functional Diversity of the Maize Genome	<a href="http://www.panzea.org">www.panzea.org</a>
ToL: Tree of Life Web Project	<a href="http://www.tolweb.org">www.tolweb.org</a>

de tejido. El desarrollo de tecnologías de alto procesamiento de datos ha permitido abordar estudios de expresión génica en forma concertada para miles de genes a partir de distintos genotipos en determinados órganos, tejidos, y condiciones de crecimiento. El diseño de los experimentos de expresión, así como las herramientas de análisis de la información generada constituyen factores claves para obtener información con significado biológico.

Las distintas estrategias que permiten evaluar perfiles transcripcionales pueden dividirse en dos grupos. Un primer grupo corresponde a estrategias en las que la estimación del nivel de la expresión se basa en la intensidad relativa de una señal de hibridación e incluye los tradicionales experimentos de *northern blot* y los microarreglos de ADNc, en los cuales se toma la intensidad relativa de la señal como medida de la transcripción. El otro grupo de estrategias se basa en una cuenta directa del número de cada transcripto presente en una muestra, incluyendo la secuenciación de secuencias expresadas o “Expressed Sequence Tags” (ESTs), el análisis de la expresión génica en serie o “Serial Analysis of Gene Expression” (SAGE) y la secuenciación masiva en paralelo o “Massively Parallel Signature Sequencing” (MPSS). En este capítulo se discutirán las estrategias de análisis de expresión que permiten la evaluación concertada de un gran número de genes en forma procesiva.

### **Expressed Sequence Tags (ESTs):**

La secuenciación de moléculas de ADN complementario (ADNc) sintetizadas a partir de ARN mensajeros obtenidos de muestra biológica en particular, realizada a través de una única lectura, da origen a un conjunto de secuencias denominadas ESTs, que representan secuencias parciales de las colecciones originales de ADNc.

Actualmente, la división EST de GenBank cuenta con un total aproximado de 6.000.000 secuencias y se encuentran representadas más de 100 especies de plantas.

La secuenciación de ADNc constituyó una herramienta alternativa de los proyectos genómicos para el descubrimiento de nuevos genes para distintas especies, cuando la secuenciación de genomas complejos no contaba con

las actuales técnicas de secuenciación masiva que prometen en un futuro cercano capacidad de procesamiento elevada a costos accesibles a la comunidad en general. Aún con el advenimiento de proyectos genómicos de gran escala, la secuenciación de ESTs permite la identificación de genes que se expresan en pocos tejidos y/o su expresión es muy baja, mediante la construcción de colecciones de ADNc normalizadas o enriquecidas en determinados transcritos como las colecciones derivadas de técnicas de hibridación sustractiva o “Suppressed Subtracted Hybridization” Asimismo, el agrupamiento de ESTs derivados de distintas colecciones de ADNc para una dada especie en grupos o clusters de secuencias no redundantes, utilizando rutinas de análisis como las esquematizadas en la Figura 1 permite estimar el número de genes de una especie.

Respecto al análisis de expresión génica, los ESTs contribuyen a estos estudios a través de dos vías: 1) la proporción de ESTs correspondientes a un determinado gen respecto a un conjunto de ESTs generados en distintas condiciones experimentales refleja el nivel de expresión del mismo; y 2) en segundo lugar las bases de ESTs se utilizan para el diseño de oligonucleótidos en experimentos de microarreglos de ADNc para evaluar expresión génica.

Los experimentos de microarreglos representan una de las herramientas más frecuentes para aproximarse a los análisis de expresión de genes a gran escala. Una vez que se obtienen los resultados surgen interrogantes como ¿Cuáles son los genes sobre o sub-expresados en el experimento?, ¿Cuáles son los perfiles de expresión que pueden revelarse en general a partir de este experimento? En esta sección ejemplificaremos un caso de estudio para ilustrar el proceso de análisis utilizando programas de libre distribución basados en el lenguaje estadístico R (Bioconductor).

En el Instituto de Biotecnología de INTA Castelar (IB) se diseñó un microarreglo de ADNc conteniendo 317 secuencias de tipo ESTs previamente desarrollados en base a secuencias órgano-específicas de una línea de girasol cultivado). Para la impresión de los microarreglos de ADNc sobre soporte de vidrio, se amplificaron por PCR los fragmentos re-

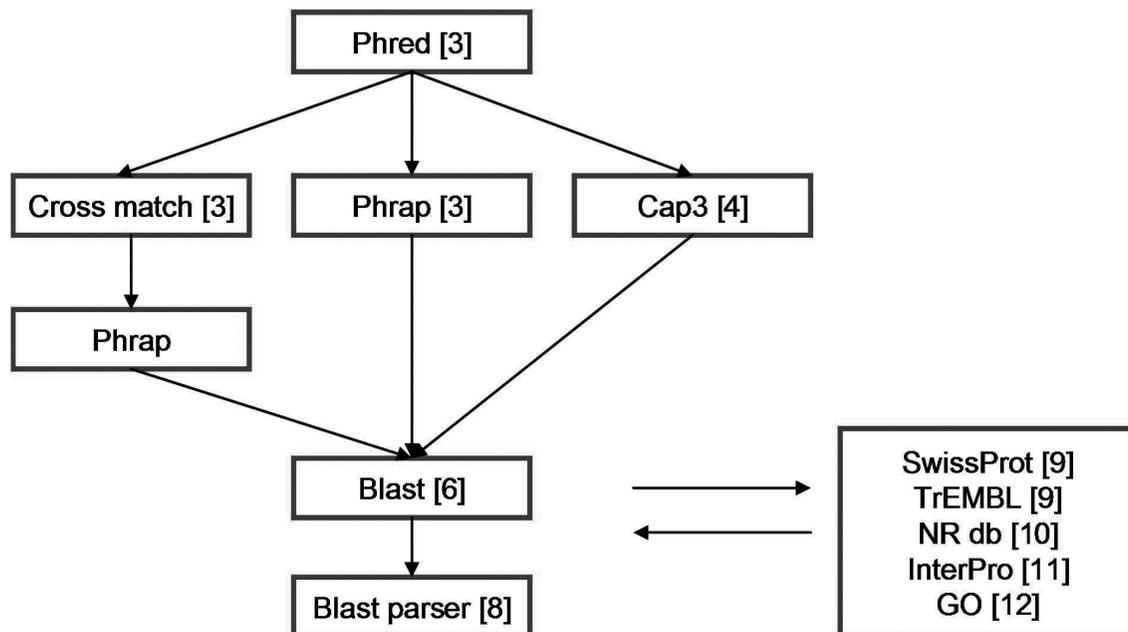


Figura 1: eBiopipeline. A. vista esquemática; B. página inicial, menú de programación y salidas del programa.

representativos de los unigenes anotados en las bibliotecas previamente caracterizadas. Estos microarreglos fueron hibridados con ARN total de plantas crecidas en invernáculo y sometidas a dos condiciones de estrés abiótico diferenciales: frío y salinidad. Una vez impresos, los vidrios fueron escaneados (usando canales para ambos fluoróforos) a tres intensidades diferentes, usando el lector de fluorescencia VersArray Chip Reader (BioRad). Las imágenes obtenidas fueron analizadas utilizando el programa de acceso libre Spotfinder ([www.tm4.org/spotfinder.html](http://www.tm4.org/spotfinder.html)), cuantificándose la intensidad de señal para cada spot. Luego, se realizó una integración de los datos de las imágenes escaneadas (Figura 2).

#### Preprocesamiento de los datos

La imagen obtenida a partir de un microarreglo corresponde a la información cruda de un experimento de estas características. Es así que los algoritmos computacionales convierten la imagen en información numérica que cuantifica dicha expresión: este es el primer paso en un análisis de datos de microarreglos, y es fundamental la calidad y la rigurosidad obteni-

da de la misma para la posterior interpretación que se obtendrá como resultado.

#### Normalización

La normalización es un término genérico que se refiere a la resolución de errores sistemáticos y desvíos producidos en un microarreglo debido a condiciones experimentales inevitables y propias de la plataforma utilizada. La normalización y el análisis de expresión global se realizaron en este caso a través de programas generados utilizando el programa R (University of Auckland) a través de su versión R 1.9.0 (<http://www.rproject.org>). Los datos de cada vidrio fueron normalizados corrigiendo la dependencia por intensidad mediante la utilización del suavizador no paramétrico LOWESS (regresión local ponderada). Se realizó una integración de los datos de las réplicas técnicas dentro de cada soporte (cada *spot* fue impreso 4 veces) y las réplicas técnicas entre soportes (intercambio de fluoróforos o “*dye-swap*”). Luego de este paso el producto obtenido es una matriz de expresión génica cuyo análisis se focaliza en la identificación de genes con expresiones diferenciales.

  
Home

Pipeline of programs  
 Phred

Process name

Browse FASTA file

---

  
Home

**Cross match program**

Options			
Banded Search	Filtering of Matches	Input data interpretation	Miscelaneas
Minmatch <input type="text" value="14"/>	Minscore <input type="text" value="30"/>	Default qual <input type="text" value="15"/>	Indexwordsize <input type="text" value="10"/>
Maxmatch <input type="text" value="30"/>	Vector bound <input type="text" value="80"/>		
Bandwidth <input type="text" value="14"/>	Masklevel <input type="text" value="80"/>		

Screen  Browse FASTA file

---

**Search report: prueba2**

Csv file  
 Hit file in fasta format  
 Go to Primer3

**Query Name:** /var/www/html/pipeline/run\_pipeline/prueba2.Contig1 **Query Accession:** /var/www/html/pipeline/run\_pipeline/prueba2.Contig1 **Query Length:** 759 **Database Name:** Helianthus\_annuus.mRNA.PUT.fasta **Hits:** 250

Hit 1

Name	PUT-157a-Helianthus_annuus-30342507		
Length	911		
Accession	PUT-157a-Helianthus_annuus-30342507		
Description	PlantGDB-assembled Unique Transcript-fragment derived from Helianthus_annuus mRNAs Jan_24_2007 (based on GenBank release 157).		
Significance	0.0		
Hit Score	686		
Hsp Quantity	1		
Hsp 1	Score	346	
	Length	358	
	E Value	0.0	
	Frac identical	0.991620111731844	

**Figura 2:** Imagen de una micromatriz de dos canales (Bello y col. 2006)

### *Limpieza de datos y transformación*

Una vez obtenida la matriz de expresión génica, existe una serie de pasos que se llevan a cabo para asegurar una alta calidad en el análisis. Ellos son: la remoción de *spots* dudosos o conflictivos (*flagged features*), que se resuelve ya sea eliminando los *spots* con error aparente (con el consiguiente riesgo de pérdida de datos valiosos), o marcándolos para luego referirlos a la imagen original en el momento de su análisis; la corrección y/o sustracción del ruido de fondo (*corrección o sustracción de background*). La señal de fondo es indicadora de una hibridación inespecífica siempre y cuando la intensidad del *spot* de interés sea mayor que la intensidad del ruido de fondo. Si ocurre al revés, existe la posibilidad que esté representando un problema local en el microarreglo y la intensidad del *spot* se convierta en información no confiable; y por último la transformación logarítmica de los valores para hacer constante la variabilidad a todos los niveles de intensidad detectados.

### **Análisis estadístico de los datos**

#### *Evaluación de consistencia en las repeticiones*

Para evaluar las diferencias asociadas a las repeticiones biológicas de este experimento se obtuvo una ordenación de las mismas en un plano de ordenación generado mediante análisis de componentes principales aplicado a la matriz de expresiones génicas. Esta aproximación tuvo en cuenta la información de todos los genes simultáneamente y permitió establecer si una repetición se comportaba de manera atípica o no.

#### *Identificación de grupos de genes con expresión diferencial.*

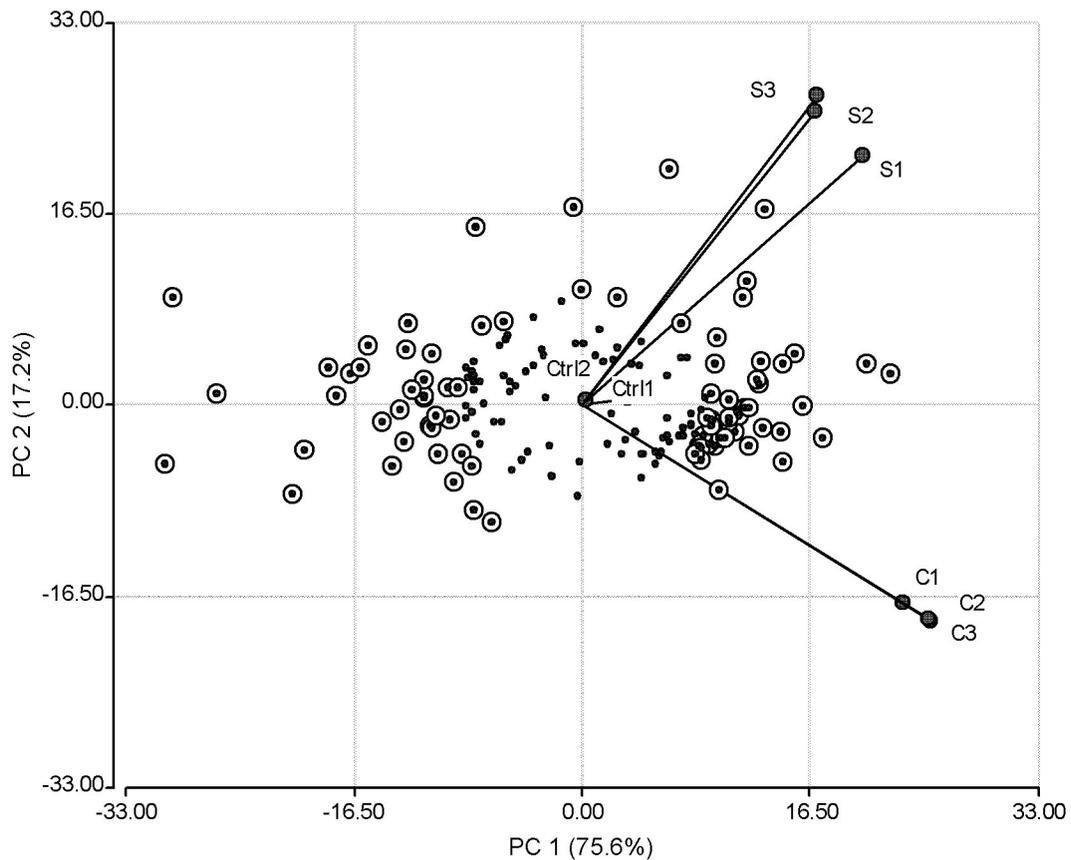
El análisis de la matriz de expresión génica se focalizó en la identificación de genes con expresiones diferenciales. Existen diversas estrategias para obtener un listado de genes candidatos. La estrategia más sencilla es aplicar una prueba de hipótesis tradicional gen a gen. Esta aproximación tiene el inconveniente de generar un gran número de falsos positivos. Los métodos correctivos, basados en el control de

la tasa de falsos positivos, presentan a su vez la dificultad de generar una alta tasa de falsos negativos. Por consiguiente, la metodología utilizada en este trabajo consistió en complementar las técnicas de inferencia clásicas con un criterio de selección basado en el ordenamiento de genes y tratamientos en el espacio generado por los dos primeros ejes principales obtenidos de un análisis de componentes principales de la matriz de expresión génica. Posteriormente, y dentro del conjunto de genes seleccionados en el paso anterior, se realizó el reconocimiento de grupos de genes con perfiles de expresión similar mediante la aplicación de un algoritmo de clasificación no supervisada, k-centroides para nuestro caso de estudio. El análisis transcripcional mediante la aplicación de k-centroides permitió detectar 3 grupos de genes bien diferenciados en sus patrones de expresión. Por una parte el Grupo 2, integrado por 126 genes, no mostró variación en sus niveles medios a través de los tratamientos. En cambio, el Grupo 1 (integrado por 112 genes) evidenció sobre-expresión respecto del control cuando las plantas fueron sometidas a estrés (por frío o por salinidad). De manera análoga, 49 genes conformaron el Grupo 3, pero en este caso se observó una sub-expresión respecto del control en ambos tratamientos.

#### *Obtención de la lista de genes diferencialmente expresados*

Para combinar esta aproximación, basada en la ordenación de genes por ACP y el criterio clásico de prueba de hipótesis para igualdad de medias, se consideraron sólo aquellos genes que para la prueba clásica de análisis de la varianza tuvieran un *p-valor* menor o igual a 0,05 y que estuvieran a un distancia del origen en la Figura 3 mayor que 9 (correspondiente, aproximadamente al percentil 70 de la distribución de distancias al origen). Este criterio de corte se seleccionó teniendo en cuenta el gen T411 (Genbank AN BU671801), que había sido previamente validado experimentalmente.

Para cada uno de los 80 genes seleccionados como genes candidatos se graficó el perfil de expresión según su comportamiento en cada tratamiento a través de un gráfico de colores de expresión diferencial (*heatmap*) de



**Figura 3.** Bi-plot que muestra genes con p-valor para un test  $F < 0,05$  y con una distancia al origen  $<$  percentilo 70 de la distancia al origen de la distribución (círculos rellenos).

modo de analizar cualitativamente el gradiente diferencial por tratamiento y por gen mediante un gráfico de perfiles de expresión individuales. De estos resultados surgieron genes candidatos de interés para su análisis de expresión diferencial, de los cuales 15 fueron seleccionados para validar mediante la técnica de PCR cuantitativa (qPCR).

Este último paso de validación es de una herramienta indispensable en el análisis de microarreglos ya que por todo lo expuesto anteriormente podemos inferir que los genes obtenidos son producto de transformaciones numéricas y una evaluación subjetiva al criterio estadístico aplicado. Es así que podemos concluir que la validación experimental confiere rigurosidad biológica a un conjunto de supuestos genes candidatos obtenidos a partir de una lista de genes diferencialmente expresa-

dos para las condiciones evaluadas, producto final de un análisis de microarreglos.

### **Búsqueda de marcadores funcionales en bases de datos genómicos**

Como mencionamos anteriormente, la secuenciación completa de los genomas de especies modelo o parcial de especies de interés ha generado una rápida acumulación de información biológica que es depositada en bases de datos genómicas y es fácilmente accesible a través de Internet. Esta información, que deriva en muchos casos de la caracterización de líneas elite de una misma especie, de especies de un mismo género o de distintas fuentes relacionadas, constituye el recurso que permite sistematizar el desarrollo de marcadores moleculares potencialmente asociados a la variación fenotípica para caracteres de interés. El

potencial de descubrir marcadores moleculares funcionales analizando las secuencias depositadas en las bases de datos generales, fundamentalmente de ESTs, ha permitido superar una de las limitaciones más importantes de los marcadores moleculares locus-específico que es el costo inicial de desarrollo.

### **Identificación *in silico* de SNPs**

Los polimorfismos de nucleótido simple (SNPs) o los generados por pequeñas inserciones/delecciones (Indels) se han convertido en una de las herramientas más utilizadas para la caracterización genética tanto de plantas como de animales debido a su baja tasa de mutación y a su gran abundancia, ubicuidad y dispersión en los genomas. Su uso se ha extendido a aplicaciones tales como la construcción de mapas genéticos de alta resolución, mapeo de asociación, estudios de diversidad genética y conservación, identificación y análisis de la estabilidad de cultivares, delimitación y asignación de grupos heteróticos, análisis filogenéticos, selección asistida por marcadores y caracterización de recursos genéticos.

Un SNP representa una diferencia en una base nucleotídica entre 2 secuencias de ADN y puede ser categorizada como transición (C/T or G/A) o transversión (C/G, A/T, C/A, o T/G) de acuerdo a la sustitución nucleotídica observada. El uso de marcadores SNP para cualquiera de las aplicaciones mencionadas requiere de una primera etapa de detección y desarrollo que puede ser abordada a partir de una estrategia "de laboratorio" (ej. secuenciación y comparación de genes o regiones candidatas en un conjunto de materiales de interés) o bien estar basada en aproximaciones *in silico* que hacen uso de la información de secuencia disponible en las bases de datos biológicas. En este último caso, el procedimiento de búsqueda generalmente involucra un protocolo de pocos pasos. Primero se identifican secuencias con alta similitud entre múltiples individuos a partir de una o varias bases de datos previamente seleccionadas en función del organismo o tipo de carácter que se desee estudiar, utilizando habitualmente los algoritmos de búsqueda de similitud de secuencias de tipo BLAST (Basic Local Alignment Search Tool). Posteriormente,

se realizan los alineamientos base a base para cada uno de los grupos de secuencias y los mismos son analizados para detectar las diferencias nucleotídicas o SNPs. La rigurosidad y detallada evaluación de esta instancia es fundamental para encontrar verdaderas variantes alélicas dentro de los diferentes grupos. Es por ello que se encuentran disponibles diversas herramientas estadísticas que permiten diferenciar los errores de secuenciación de las verdaderas variantes alélicas basándose en el valor de calidad de la base secuenciada y en medidas de exactitud de la secuencia. Los programas desarrollados para la detección de SNPs, tanto para uso académico (PolyBayes y PolyPhred) como aquellos sujetos a licencias comerciales (Sequencher, Gene Codes Corporation) utilizan los datos de calidad de secuencia para eliminar los falsos positivos o incorporan el concepto de velocidad de mutaciones esperadas para distinguir los verdaderos SNPs. Asimismo, estos programas ofrecen la ventaja de presentar una interfase gráfica y pueden ser usados tanto para comparación de secuencias cortas como para estimar SNPs dentro de largas regiones genómicas. Muchos grupos han explorado además métodos alternativos de detección de SNPs basados en el uso de microarreglos de ADN de alta densidad.

### **Identificación *in silico* de SSRs**

Los microsatélites (repeticiones en tandem de mono, di, tri, tetra o pentanucleótidos) son herramientas muy útiles como marcadores genotípicos debido a que son relativamente abundantes, multialélicos (hipervariables con respecto al número de repeticiones), heredados en forma estable y codominantes (se distinguen los alelos de un locus), permitiendo distinguir materiales estrechamente relacionados. Desde el punto de vista metodológico, son fáciles de detectar mediante PCR, requiriendo poca cantidad de ADN para el análisis y poco equipamiento.

La identificación *in silico* de SSRs a partir de la abundancia de secuencias genómicas disponibles para genomas vegetales, al igual que para los SNPs, se ha convertido en la alternativa más económica de desarrollar este tipo de marcadores haciendo uso de herramientas

computacionales eficientes. Asimismo, es muy frecuente poder asociar una función molecular determinada por similitud a la secuencia que alberga la repetición convirtiéndolo en un marcador funcional.

Existen distintas herramientas computacionales basadas en la web para el descubrimiento de SSRs y el diseño de iniciadores específicos para la amplificación de los mismos. Se pueden mencionar como ejemplos el programa SSRPoly accesible desde el sitio <http://acpfg.imb.uq.edu.au/ssrpoly.php>, que permite la identificación de SSR polimórficos que se distinguen de las variantes monomórficas por la representación de tamaños variables del microsatélite identificado en alineamientos múltiples de secuencias representadas en la bases de datos. Otra herramienta valiosa disponible en la web es CUGI SSR Server ([http://www.genome.clemson.edu/cgi-bin/cugi\\_ssr](http://www.genome.clemson.edu/cgi-bin/cugi_ssr)).

### **Análisis de la variabilidad genética**

Finalizada la etapa de desarrollo de marcadores moleculares, la fase siguiente generalmente consiste en la genotipificación de grandes cantidades de individuos a fin de determinar las variantes alélicas presentes en cada uno de ellos para cada uno de los marcadores detectados. La elección de la metodología más apropiada dependerá de los recursos disponibles y del nivel de procesamiento que se desee alcanzar. Entre las más utilizadas para el caso de los SNPs pueden mencionarse: secuenciación directa, amplificación alelo-específica por reacción en cadena de la polimerasa (PCR), métodos de digestión enzimática, hibridación con oligonucleótidos alelo-específicos, y por último las plataformas de microarreglos GoldenGate e Infinium desarrolladas por Illumina (Illumina, San Diego, CA, USA).

La tecnología de SNPs presenta un gran potencial para la identificación de las formas alélicas que controlan procesos biológicos complejos, como la resistencia a los estreses bióticos y abióticos que afectan la productividad de los cultivos. Sin embargo, para ser aprovechado en su totalidad, este potencial debe ser combinado con metodologías analíticas apropiadas. En los últimos años las metodologías de mapeo por asociación han surgido como com-

plemento de las aproximaciones convencionales (ej. Mapeo de QTLs) y han ido cobrando cada vez más importancia como herramientas para alcanzar un mayor grado de resolución. En pocas palabras, el razonamiento subyacente al mapeo de asociación consiste en utilizar eventos de recombinación históricos a lo largo de un linaje, y no solamente aquellos ocurridos en una determinada población de mapeo, para establecer posibles relaciones entre genotipo y fenotipo. Los polimorfismos responsables de la variación fenotípica no son observados en forma directa, si no que surgen a partir de inferencia estadística. La base para la detección de tales correlaciones es la asociación no aleatoria entre un polimorfismo dado y la variación fenotípica para cierto rasgo de interés, es decir la identificación de desequilibrio de ligamiento (DL) entre las variantes genéticas causales de la variación del carácter y los polimorfismos analizados.

Previo a cualquier estudio de mapeo por asociación, y para realizar un diseño experimental apropiado, es necesario conocer el grado y la estructura genómica del DL en la especie a evaluar, ya que éste es variable tanto entre especies como para las diferentes regiones de un mismo genoma. Las dos medidas preferidas en la literatura para cuantificar DL son el coeficiente de desequilibrio estandarizado ( $D'$ ) y la correlación de frecuencias alélicas al cuadrado ( $r^2$ ). La estimación del DL generalmente involucra grandes cantidades de marcadores, razón por la cual suelen resultar útiles las herramientas de software que permiten sistematizar el cálculo de ambas medidas y representar gráficamente las asociaciones halladas (por ej.: DNAsp, Arlequin, Gold, GoldSurfer).

El desarrollo de un estudio de mapeo por asociación involucra la evaluación fenotípica del carácter agronómico de interés (ej. peso del grano, contenido de aceite, días a floración, AUDPC, etc) y la genotipificación de un conjunto apropiado de SNPs para cada una de las entidades a comparar (líneas, variedades, etc). El número y naturaleza de los polimorfismos a caracterizar dependerá de la estrategia de análisis que se haya seleccionado. Si se desea explorar todo el genoma (genome-wide analysis) el número de SNPs será mucho ma-

yor (entre 10.000 y 100.000) que el requerido si sólo se estudian los polimorfismos correspondientes a un grupo de genes candidatos. Sin embargo, en este último caso es necesario un conocimiento previo de la base genética del carácter fenotípico en estudio. Otro aspecto importante del diseño experimental es la selección de individuos. En humanos generalmente es la utilización de metodologías basadas en poblaciones de enfermos vs. controles sanos (case-control) o aquellas que involucran individuos relacionados o familias en donde algún individuo es enfermo (Transmission disequilibrium tests). En plantas, los individuos de la población en estudio usualmente incluyen germoplasma de distinto origen geográfico con el objeto de abarcar la mayor cantidad posible de variación genética y fenotípica.

Las inferencias estadísticas del mapeo por asociación para rasgos continuos son generalmente realizadas a través de análisis de regresión lineal, si bien también son aplicables otras aproximaciones. Dado que existen diversos factores que pueden llevar al establecimiento de asociaciones espurias, se han desarrollado modelos específicos en donde es posible incorporar las distintas variables intervinientes y controlar su efecto en la detección de asociaciones. Entre los procesos que pueden interferir en el análisis se encuentran: la subestructuración de las poblaciones estudiadas, las interacciones epistáticas y pleiotrópicas, las interacciones entre genotipo y ambiente, el tamaño muestral pequeño, la complejidad del carácter en estudio y la calidad de los registros fenotípicos. El software Tassel es uno de los paquetes más utilizados para los estudios de asociación en plantas ya que incluye metodologías que permiten tener en cuenta la estructura poblacional y el grado de parentesco entre los individuos analizados.

En conclusión, el mapeo por asociación constituye una poderosa herramienta para la disección genética de caracteres complejos, siendo los SNPs marcadores ideales para la implementación de tales estudios dada su alta frecuencia y ubicuidad en los genomas.

## Lecturas recomendadas

- Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol* 215:403-10.
- Ashburner, M., C.J. Mungall, and S.E. Lewis. 2003. Ontologies for biologists: a community model for the annotation of genomic data. *Cold Spring Harb Symp Quant Biol* 68:227-35.
- Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25-9.
- Bairoch, A. 2000. The ENZYME database in 2000. *Nucleic Acids Res* 28:304-5.
- Bairoch, A., and R. Apweiler. 1996. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Res* 24:21-5.
- Balding, D.J. 2006. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 7:781-91.
- Cervigni, G.D., N. Paniago, M. Diaz, J.P. Selva, D. Zappacosta, D. Zanazzi, I. Landerreche, L. Martelotto, S. Felitti, S. Pessino, G. Spangenberg, and V. Echenique. 2008. Expressed sequence tag analysis and development of gene associated markers in a near-isogenic plant system of *Eragrostis curvula*. *Plant Mol Biol* 67:1-10.
- Conesa, A., and S. Gotz. 2008. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *Int J Plant Genomics* 2008:619832.
- Diatchenko, L., Y.F. Lau, A.P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E.D. Sverdlov, and P.D. Siebert. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A* 93:6025-30.
- Fernandez, P., N. Paniago, S. Lew, H.E. Hopp, and R.A. Heinz. 2003. Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project. *BMC Genomics* 4:40.
- Fernandez, P., J. Di Rienzo, L. Fernandez, H. Hopp, N. Paniago, and R. Heinz. 2008. Transcriptomic identification of candidate genes involved in sunflower responses to chilling and salt stresses based on cDNA microarray analysis. *BMC Plant Biol* 8:11.
- Gupta, P.K., J.K. Roy, and M. Prasad. 2001. Single nucleotide polymorphisms: A new paradigm for

- molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr Sci* 80:524-535.
- Kanehisa, M., S. Goto, S. Kawashima, and A. Nakaya. 2002. The KEGG databases at GenomeNet. *Nucleic Acids Res* 30:42-6.
- Martin, D.M., M. Berriman, and G.J. Barton. 2004. GOtcha: a new method for prediction of protein function assessed by the annotation of seven genomes. *BMC Bioinformatics* 5:178.
- Oraguzie N. et al., eds. 2007. *Association Mapping in Plants*. Springer Science+Business Media, LLC.
- Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* 5:94-100.
- Stekel, D. 2003. *Image Processing*, In P. C. University, ed. *Microarray Bioinformatics*, NY, USA.
- Stich, B., and A.E. Melchinger. 2009. Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. *BMC Genomics* 10:94.
- Wold, B., and R.M. Myers. 2008. Sequence census methods for functional genomics. *Nat Methods* 5:19-21.
- Wu, C.H., L.S. Yeh, H. Huang, L. Arminski, J. Castro-Alvear, Y. Chen, Z. Hu, P. Kourtesis, R.S. Ledley, B.E. Suzek, C.R. Vinayaka, J. Zhang, and W.C. Barker. 2003. The Protein Information Resource. *Nucleic Acids Res* 31:345-7.

