




Selección Genética y Genómica
en Agricultura, Ganadería,
Silvicultura y Acuicultura

Informe de Vigilancia Tecnológica



Genoma España



Selección Genética y Genómica
en Agricultura, Ganadería,
Silvicultura y Acuicultura
Informe de Vigilancia Tecnológica



Genoma España

Selección Genética y Genómica en Agricultura, Ganadería, Silvicultura y Acuicultura

La reproducción parcial de este informe está autorizada bajo la premisa de incluir referencia al mismo, indicando:
Selección Genética y Genómica en Agricultura, Ganadería, Silvicultura y Acuicultura. Informe de Vigilancia Tecnológica. GENOMA ESPAÑA.

Genoma España no se hace responsable del uso que se realice de la información contenida en esta publicación. Las opiniones que aparecen en este informe corresponden a los expertos consultados y a los autores del mismo.

Autores:

Dr. Francisco Javier Gallego
Universidad Complutense de Madrid

Dr. Armand Sánchez Bonastre
Universidad Autónoma de Barcelona

Dr. Pere Arús
Instituto de Investigación y Tecnología Agroalimentaria (IRTA)

Dr. Miguel Ángel Toro
Universidad Politécnica de Madrid

Dr. Paulino Martínez Portela
Universidad de Santiago de Compostela

Coordinación:

Miguel Vega García
Fernando Garcés
Olga Ruiz Galán

© Fundación para el Desarrollo
de la Investigación en Genómica
y Proteómica (Genoma España)

Edición: Genoma España
Referencia: GEN-ES11005
Fecha: 2011
Diseño y realización: Creaciones Hazanas, S.L.

Índice de contenido

1. INTRODUCCIÓN

7

2. DESCRIPCIÓN DE TECNOLOGÍAS. TÉCNICAS DE GENOTIPADO DE SNP

15

- 2.1. Protocolos basados en la utilización de equipos de PCR a tiempo real 20
 - 2.1.1. Empleo de sondas alelo-específicas con doble marcado fluorescente 20
 - 2.1.2. Genotipado mediante PCR alelo-específica y detección fluorescente 22
 - 2.1.3. Genotipado mediante la utilización de fluorocromos intercalantes 23
- 2.2. Protocolos de genotipado de alto rendimiento 23
 - 2.2.1. Protocolos basados en ligación alelo-específica 23
 - 2.2.2. Protocolos basados en extensión de base 24
 - 2.2.3. *Microarrays* de alto rendimiento 26

3. APLICACIÓN DE LA SELECCIÓN GENÉTICA Y GENÓMICA EN GANADERÍA

27

- 3.1. Consideraciones generales sobre los programas de mejora genética 27
- 3.2. Biología molecular y mejora genética animal 29
 - 3.2.1. Detección de QTL en las especies domésticas 29
 - 3.2.2. Detección de los genes responsables de los QTL 32
- 3.3. Selección asistida con marcadores (MAS) 33
- 3.4. Una propuesta más radical: la selección genómica 35
- 3.5. Ciencia ficción: el uso combinado de la selección genómica y la manipulación de los gametos y embriones 38

4. APLICACIÓN DE LA SELECCIÓN GENÉTICA Y GENÓMICA EN AGRICULTURA

39

- 4.1. Características específicas de las plantas y sus genomas en relación con su mejora genética 39
- 4.2. Aplicaciones de los marcadores moleculares a la selección y mejora genética de plantas basadas en un número bajo de marcadores 40
 - 4.2.1. Identificación de genotipos y control de calidad 40
 - 4.2.2. Selección asistida por marcadores (MAS) 41
 - 4.2.3. Selección de todo el genoma en programas de retrocruzamiento 42
- 4.3. Nuevas aplicaciones de los marcadores basadas en la cobertura completa del genoma 43
 - 4.3.1. Uso del desequilibrio de ligamiento: genética de asociación 43
 - 4.3.2. Generación de poblaciones especialmente adaptadas al análisis de asociación 44
 - 4.3.3. Nuevas estrategias de mejora especialmente adaptadas a la selección de caracteres de herencia compleja 45
- 4.4. Casos prácticos: estado actual de la selección genética y genómica en varias especies 46

5. APLICACIÓN DE LA SELECCIÓN GENÉTICA Y GENÓMICA EN SILVICULTURA

49

5.1. Estrategia clásica de identificación de QTL	50
5.2. Estrategia de estudios de asociación	51
5.3. Integración de la selección genómica en los programas de mejora forestal	53
5.4. Genética y genómica de <i>Eucalyptus</i> y su aplicación a la mejora	54
5.4.1. Selección genética	55
5.4.2. Selección genómica	56

6. APLICACIÓN DE LA SELECCIÓN GENÉTICA Y GENÓMICA EN ACUICULTURA

59

6.1. Producción y mejora genética en acuicultura	59
6.2. Rastreo genómico mediante mapas genéticos para la identificación de <i>loci</i> de QTL	60
6.3. Acuicultura y mejora genética del rodaballo	62
6.4. Identificación de QTL: aplicación al sector del rodaballo	62
6.5. La detección precoz del sexo en el rodaballo: una aplicación industrial de los desarrollos genómicos	63

7. CONCLUSIONES

67

7.1. La selección genómica es una alternativa a la selección genética (MAS)	68
7.2. Beneficios de la selección genómica	69
7.3. Conclusiones por áreas	70
7.4. Conclusiones finales	73

8. REFERENCIAS BIBLIOGRÁFICAS

76

8.1. Introducción	76
8.2. Descripción de tecnologías. Técnicas de genotipado de SNP	78
8.3. Aplicación de la selección genética y genómica en ganadería	81
8.4. Aplicación de la selección genética y genómica en agricultura	83
8.5. Aplicación de la selección genética y genómica en silvicultura	85
8.6. Aplicación de la selección genética y genómica en acuicultura	88

1. Introducción

Desde mediados del siglo XIX hasta nuestros días se ha producido, mediante selección artificial, una espectacular mejora de las especies de interés para el hombre. Se han conseguido cultivares y razas con mayor producción y calidad, así como más adaptados a estreses bióticos y abióticos. Los mejoradores, que se enfrentan a esa continua tarea del desarrollo de nuevos cultivares y razas, tienen ante sí otro reto. En la actualidad el margen de mejora de las especies de interés es muy escaso, pues la mayor parte los genes más importantes ya han sido seleccionados. En este contexto, la ciencia debe aportar nuevas herramientas que permitan manipular el resto de genes con efectos menores, pero en absoluto desdeñables en su conjunto.

Es bien conocido el papel que puede tener en este proceso de mejora la ingeniería genética mediante la introducción de genes de unas especies en otras. Otra gran área de la biotecnología que pasa más desapercibida es la tecnología de los marcadores moleculares, derivada de la investigación en genética molecular. Los marcadores moleculares pueden emplearse para identificar aquellos individuos portadores de las variantes alélicas más favorables para los genes responsables de los caracteres de interés. La utilización de los citados marcadores puede ayudar a aumentar la eficacia de los procesos de mejora genética y es a lo que se le denomina Selección Asistida por Marcadores (MAS, *Marker Assisted Selection*). La potencialidad del empleo de marcadores ligados a genes de interés ha sido evidente durante años. Se puede simplificar mucho la selección si, en lugar de tener que medir el carácter correspondiente, se puede aplicar un ensayo molecular con un resultado claro y fácil de analizar. El concepto de MAS comenzó a aparecer en los años 90 a raíz de los primeros trabajos sobre análisis molecular de caracteres cuantitativos (Paterson y col., 1988) aunque la primera referencia es muy anterior (Neimann-Sorenson y Robertson, 1961). Desde entonces, son innumerables los artículos científicos que describen la identificación de marcadores moleculares estrechamente ligados a genes de interés, pero, por el contrario, son bastante escasas las referencias de la aplicación en programas de mejora de esos marcadores moleculares. Es por ello que las diversas revisiones que se han hecho sobre este asunto tengan, por lo general, un tono "cautamente optimista" (Young, 1999). Las herramientas están ahí pero no se utilizan tanto como parecería razonable.

Los marcadores genéticos

Los marcadores genéticos pueden clasificarse en tres grupos: morfológicos, cromosómicos y moleculares. Los primeros marcadores genéticos empleados fueron diversos caracteres morfológicos tales como coloración, talla, morfología foliar, etc. Este tipo de marcadores presentan una serie de desventajas que reducen su utilidad. Por lo general son caracteres determinados por varios genes, es necesario el desarrollo del individuo hasta la madurez, su expresión puede estar influenciada por el ambiente y presentan frecuentemente interacciones génicas (epistasias). Pese a esto, hay marcadores de este grupo, que presentan control monogénico y son de fácil clasificación, que son empleados habitualmente.

Dentro del grupo de **marcadores cromosómicos** se encuentran, además de diferentes tipos de bandeos cromosómicos (bandeo C, bandeo G, etc.), mutaciones ta-

les como las translocaciones o las inversiones. Si bien su empleo no está muy extendido alguno de estos marcadores son de gran utilidad por estar asociados a caracteres de interés. En la actualidad se está produciendo un auge en la utilización de marcadores cromosómicos debido al desarrollo de las técnicas de hibridación *in situ* (FISH, *Fluorescence In Situ Hybridization*).

El nacimiento de los **marcadores moleculares** se produce con el desarrollo de las técnicas electroforéticas que permitían el análisis de la variabilidad genética existente a nivel de proteínas. Diferentes variantes alélicas de un mismo gen pueden presentar distinta carga neta debido a modificaciones en la estructura primaria de las proteínas. Las más empleadas son las **isoenzimas**, que son distintas formas moleculares de la misma enzima que muestran idéntica o similar especificidad de sustrato. El uso de este tipo de marcadores permitía un análisis más directo del producto de los genes, sin influencias ambientales. Por primera vez, se podía analizar un elevado número de individuos, en un tiempo relativamente corto y de manera no destructiva. Su principal inconveniente reside en que el número de proteínas que se puede analizar por electroforesis es bastante limitado.

Finalmente, el desarrollo de técnicas para la manipulación del ADN ha permitido el desarrollo de un gran número de marcadores basados en las diferencias en secuencias de ADN para los individuos de una misma especie. Este tipo de estudios tiene la ventaja de analizar directamente el material hereditario y no su manifestación fenotípica. De forma muy general podríamos decir que estos marcadores moleculares se basan, según los diferentes tipos, en el uso de las tecnologías que permiten la manipulación del ADN (secuenciación, hibridación, digestión, amplificación, etc.). Son muy variados los tipos de marcadores moleculares y se encuentran en continuo proceso de evolución.

El primer tipo de marcadores de ADN desarrollado fueron los **RFLP** (*Restriction Fragment Length Polymorphism*) que se basan en el empleo de endonucleasas de restricción para digerir el ADN. La variación se detecta en función del cambio de tamaño de los fragmentos de ADN digeridos. Estos cambios se deben a alteraciones en las dianas que reconocen las endonucleasas o a inserciones y deleciones de piezas de ADN en la región estudiada. El hecho de que el número de sondas utilizables sea prácticamente ilimitado fue la principal característica responsable del extendido uso de los RFLP en estudios genéticos. Por otra parte, esta tecnología supone un proceso largo y costoso, siendo además necesario, en la mayor parte de los casos, el empleo de radiactividad.

La técnica de la **reacción en cadena de la polimerasa** (PCR, *Polymerase Chain Reaction*), que permite la amplificación enzimática del número de copias de un segmento de ADN (Mullis y col., 1986), supuso una revolución en la Biología y, en particular, en el campo de los marcadores moleculares. El principal inconveniente de esta técnica es que requiere el conocimiento previo de la secuencia a amplificar. Para solventar este problema se desarrollaron una serie de variantes de la técnica, como son los **RAPD** (*Random Amplified Polymorphic DNA*) (Williams y col., 1990), y otros, que se basan en el empleo de cebadores de secuencia al azar para los que no es preciso tener información previa de secuencia. El polimorfismo encontrado puede producirse por la existencia de inserciones o deleciones dentro del segmento amplificado, aunque, en general, es consecuencia de mutaciones puntuales que afectan a nucleótidos de las secuencias diana del cebador. El método es sencillo, rápido y con

costes reducidos. Sin embargo, los RAPD presentan problemas de repetibilidad por lo que están en desuso.

Con posterioridad, según se ha dispuesto de más secuencias, se ha desarrollado la técnica llamada **STS** (*Sequence Tagged Sites*) (Olson y col., 1989) que consiste en la amplificación específica de fragmentos de ADN. Entre muchas variantes, están los **PCR-RFLP** que permiten el análisis con enzimas de restricción pero de una forma más rápida y sencilla que la clásica de los RFLP. Además, se han desarrollado otro tipo de marcadores basados en la PCR en regiones con microsatélites (Tautz, 1989; Weber y May, 1989). **Los microsatélites**, o **SSR** (*Simple Sequence Repeats*) son muy abundantes y se encuentran dispersos por el genoma. En este caso, se detecta la variación en el número de repeticiones del tándem mediante amplificación, empleando cebadores complementarios a las regiones flanqueantes al microsatélite. En la actualidad son muy utilizados debido a la gran variabilidad que detectan y a su fiabilidad.

Los **AFLP** (*Amplification Fragment Length Polymorphism*) (Vos y col., 1995) es una metodología que combina características de los RFLP y de los marcadores basados en la PCR. Para la identificación de polimorfismo se amplifican fragmentos previamente digeridos. Su mayor ventaja reside en la posibilidad de analizar cientos de fragmentos de ADN al mismo tiempo. Por contra, esta técnica es más laboriosa. Su uso se ha extendido durante años hasta la aparición de nuevos tipos de marcadores de más fácil estudio, como los SNP (*Single Nucleotide Polymorphism*).

Los **Polimorfismos de Nucleótido Simple** (SNP) son variaciones en la secuencia de ADN que afectan a un solo nucleótido y están convirtiéndose rápidamente en la mejor opción para producir muchos marcadores en poco tiempo. Los SNP son una parte importante de todas las variaciones presentes en los genomas y aparecen en promedio cada 100 a 300 bases. Estas variaciones en la secuencia del ADN pueden o no ser la causa directa de las diferencias funcionales en diferentes alelos de un mismo gen, pero, en cualquier caso, pueden estar estrechamente ligados a esos genes que se pretende mejorar. En la actualidad existen numerosas tecnologías para el genotipado de este tipo de marcador: SNPlex, Taqman, SnapShot, MLPA, *microarrays*, entre otras.

Los marcadores moleculares son herramientas para escrutar la variabilidad presente en el ADN de diferentes individuos, muestreando al azar esa variabilidad. El mejor tipo de marcador molecular será aquel que abarque una mayor parte del genoma estudiado. La nueva revolución de los secuenciadores automáticos ha puesto a nuestra disposición el mejor marcador posible: la secuencia completa de los genomas. En la actualidad ya es posible la secuenciación masiva de genomas completos a un precio razonable, aunque aún inabordable por un programa de mejora. Las compañías (Roche, Applied Biosystems, Illumina, etc.) están en una carrera por conseguir la secuenciación de genomas humanos a 1.000 dólares. Mucho antes de lo que pudiéramos imaginar la secuenciación masiva será una herramienta habitual para la mejora.

De diferentes formas, todos estos marcadores tratan de identificar las diferencias en la secuencia del ADN de los individuos bien por esos cambios en un único nucleótido (SNP, *Single Nucleotide Polymorphisms*) o por la inserción o delección de una pieza de ADN (*Indels, insertions-deletions*). Cada tipo de marcador molecular tiene sus ventajas e inconvenientes, pero desde un punto de vista general, podemos decir que permiten obtener mucha información del ADN con un coste y en un

tiempo bastante razonables. De todo esto hablaremos más adelante y será una de las claves para su utilización en programas de mejora.

Los Mapas Genéticos y la Mejora

Como dijimos previamente, los marcadores moleculares pueden convertirse en piezas clave en el desarrollo de nuevos programas de mejora genética. Los marcadores moleculares sirven para detectar variabilidad genética, información que puede ser de extrema utilidad en muy diferentes aplicaciones. Así, entre otras muchas aplicaciones, podemos emplear la variabilidad para determinar relaciones filogenéticas entre individuos o poblaciones, para proteger materiales mejorados o para identificar genes de interés económico. Es por lo general esta última utilización la que puede desembocar en el uso de marcadores en programas de MAS. En virtud del fenómeno conocido como ligamiento genético, dos piezas de ADN que se encuentran físicamente muy próximas dentro del mismo cromosoma tienden a heredarse de forma conjunta a los hijos de un individuo. Por lo tanto, cuando un marcador molecular está estrechamente ligado a un gen de interés, esa variante del marcador se heredará conjuntamente con la variante del gen que porte ese individuo y podrá servirnos para predecir el genotipo de los individuos de la descendencia mirando únicamente al marcador y no al gen.

La aparición de los marcadores de ADN ha permitido la construcción de mapas saturados. Estos mapas contienen un elevado número de marcadores localizados en puntos concretos del genoma y para los que se sabe la distancia genética relativa entre unos y otros. Es a partir de estos mapas como se puede establecer qué marcadores están más próximos a cada uno de los genes de interés y, por lo tanto, son más útiles en programas de mejora. Así, en la actualidad ya hay mapas de estas características en la mayor parte de especies de importancia económica. Adicionalmente, se han desarrollado diferentes procedimientos para saturar de marcadores específicamente las regiones del genoma de interés. De este modo se reduce notoriamente el tiempo y el dinero invertidos en estos trabajos al no tener que construir mapas genéticos completos. Uno de los campos que están mostrando un mayor desarrollo gracias a la obtención de estos mapas o regiones saturadas de marcadores es el estudio de caracteres cuantitativos. Se han podido conseguir marcadores estrechamente ligados a diversos *loci* de caracteres cuantitativos (QTL, *Quantitative Trait Loci*) de una forma más o menos sencilla.

Hay una serie de consideraciones que deben tenerse en cuenta para la elección de los marcadores moleculares correctos en un programa de MAS:

1. *Fiabilidad*: un marcador debe ser muy fiable, esto es, debemos ser capaces de deducir sin equivocarnos el fenotipo para el carácter a partir del genotipo para el marcador. El riesgo de equivocarnos será mayor cuanto mayor sea la distancia genética entre el marcador y el gen. La mejor solución posible es que el marcador no solo se encuentre muy próximo al gen si no que esté físicamente dentro de él.
2. *Cantidad de variabilidad detectada*: los marcadores moleculares ejercen a modo de lupas que escrutan el material hereditario en diferentes regiones. Como regla general, serán más útiles aquellos marcadores que detectan más variabilidad y que, por lo tanto, permitan observar más diferencias entre individuos.

3. *Requerimiento en cantidad y calidad del ADN*: no todos los marcadores tienen las mismas exigencias en cuanto al material de partida y esto se traduce en un coste en cuanto al tiempo y al dinero necesario para preparar el ADN que se analizará.
4. *Complejidad del proceso técnico*: hay tecnologías que requieren una mayor preparación y entrenamiento por parte del personal que realizará los marcadores y en muchas ocasiones puede suponer un freno al empleo de ese tipo de marcador.
5. *Coste*: es evidente también que cada tecnología tiene un coste para la obtención de los marcadores, tanto para implementar la tecnología como posteriormente para llevarla a cabo en el día a día.

La decisión de qué tipo de marcador debe utilizarse en cada caso particular dependerá de estas cuestiones y, evidentemente, de los datos previos: tipo de marcadores descritos con anterioridad en la especie, mapas genéticos, información sobre ligamientos de algún marcador al carácter o caracteres que se quieran mejorar, etc. Actualmente, en la mayor parte de los casos son los marcadores que identifican por PCR alelos específicos, los microsatélites y los SNP los más empleados en programas de mejora. Un ejemplo es la comercialización del trigo Patwin, la primera variedad comercial desarrollada por MAS en la Universidad de California-Davis, portadora de dos genes de resistencia a la roya estriada y a la roya de las hojas seleccionados con marcadores basados en la PCR (Helguera *et al.*, 2003).

Los caracteres cuantitativos y los marcadores moleculares

Ya sabemos que los marcadores moleculares son más útiles en la medida en la que están más estrechamente ligados a genes de interés. El punto crítico que permitirá diseñar con éxito un programa de MAS es la disección genética del carácter. Es preciso disponer de diferentes poblaciones en las que segregue el carácter para deducir cuántos genes participan en su control, qué papel individual tiene cada uno y dónde se localizan. Todo este trabajo de investigación previo es la parte más crucial que determinará la posibilidad de implementar un programa de mejora apoyado en marcadores. Una parte de los caracteres de interés están controlados por un único gen pero la gran mayoría lo están por varios genes (QTL, *Quantitative Trait Loci*) complicándose por consiguiente la selección de todas las variantes positivas de todos los genes implicados. Aunque no entraremos en detalle sobre el estudio de caracteres cuantitativos sí podemos indicar que conllevan una enorme dificultad pues es preciso trabajar con elevado número de individuos de diferentes procedencias que deben estudiarse durante varias generaciones y en distintos ambientes. Pero es precisamente por la complejidad de ese análisis por lo que en la mejora de caracteres cuantitativos es donde la selección asistida por marcadores puede ser más útil.

La MAS presenta una serie de ventajas genéricas frente a la selección por observación directa del carácter cuantitativo:

1. Es mucho más simple el análisis de marcadores que la medición y estudio de un carácter cuantitativo.
2. El análisis de marcadores puede realizarse en fases muy tempranas del desarrollo de los individuos por lo que no es preciso alcanzar el momento en el que se exprese el carácter.

3. Con los marcadores moleculares se pueden seleccionar individuos portadores de todos los alelos favorables para todos los genes, tarea imposible con el sistema tradicional.

Por lo que hemos visto hasta el momento, los marcadores moleculares parecen tener únicamente aplicación en la selección de los mejores individuos para ser los parentales en cada generación. Sin embargo, las aplicaciones de los marcadores son diversas (Collard y Mackill, 2007):

1. Evaluación del material de mejora: hay toda una serie de aplicaciones útiles para identificar las poblaciones susceptibles de ser incluidas en programas de mejora:
 - a) Estudio de la pureza de las poblaciones.
 - b) Estudio de la diversidad genética entre poblaciones para ampliar la base genética a mejorar.
 - c) Identificación de individuos híbridos que tengan una mayor respuesta en mejora.
2. Retrocruzamientos asistidos por marcadores: es muy común la introducción de nuevos genes (alelos) de interés en líneas élite. Una vez introducidos los genes mediante cruzamientos, es preciso retrocruzar por las líneas élite para recomponer el genotipo de esas líneas donde finalmente solo haya cambiado su genoma en las regiones donde están los nuevos genes. De nuevo, un número suficiente de marcadores moleculares permitirá identificar aquellos individuos en los que la mayor parte del genoma sea el primigenio de la línea élite pero con los nuevos atributos introducidos. Este es un procedimiento que se emplea habitualmente en el caso de introducción de genes por transgénesis.
3. Piramidación asistida por marcadores: en este caso los marcadores serán útiles para introducir varios genes de interés en un programa de mejora. Un ejemplo típico es el de la combinación de varios genes de resistencia a un patógeno. Hasta ahora, con mejora clásica, esto era tremendamente complejo porque los diferentes genes de resistencia muestran generalmente el mismo fenotipo.
4. Selección temprana de individuos favorables: esta es la aplicación más evidente. En cada generación, y en las primeras fases del desarrollo se pueden seleccionar los individuos favorables, descartando el resto, con el consiguiente ahorro de tiempo y dinero.

La práctica ha demostrado que, por muy útil que pueda ser la estrategia de mejora asistida por marcadores, los mejores resultados de mejora se obtienen con la combinación de la mejora tradicional con esta selección molecular. La selección clásica por fenotipos es precisa para confirmar el resultado de la selección por marcadores y además es imprescindible para aquellos caracteres en los que no todos los genes han sido localizados. Hay numerosos ejemplos de los éxitos de estas estrategias de mejora. En arroz, por ejemplo, se mejoraron a la vez los caracteres de número de granos y altura de la planta mediante el uso de marcadores moleculares (Ashikari y col., 2005). En cebada se pudieron localizar y mejorar simultáneamente un gen cualitativo y tres QTL de resistencia a la roya (Castro, 2003). En el caso de los animales también hay numerosos ejemplos del uso de marcadores para mejorar caracteres cuantitativos (Dekkers, 2004).

La selección genética frente a la selección genómica

La naturaleza universal del ADN implica que el desarrollo de marcadores moleculares y su aplicación en programas de mejora es factible en cualquier especie de interés, ya sea ganadera, agrícola, forestal o piscícola. A pesar de las ventajas aquí descritas del empleo de la selección asistida por marcadores en programas de mejora, por el momento parece que no ha tenido un impacto tan grande como cabría esperar. Las causas son variadas y se podrían englobar en dos clases: las relacionadas con el desarrollo de las tecnologías de marcadores moleculares y aquellas que tienen que ver con el análisis de caracteres cuantitativos. Los avances en el desarrollo de marcadores moleculares son espectaculares y se dirigen, como veremos a continuación, a permitir el estudio de los genomas en su conjunto. Además, las compañías están apostando por simplificar y, sobre todo, abaratar esas tecnologías para ponerlas al alcance de laboratorios menos especializados. Por otra parte, el análisis de caracteres cuantitativos, tal como se realiza actualmente, tiene una serie de problemas de difícil solución, entre otros: 1) hay QTL identificados que son poco fiables, pues no se repiten en réplicas o poblaciones diferentes; 2) muchos de los genes menores de un carácter tienen efectos indetectables y 3) el mismo QTL puede mostrar efectos diferentes (incluso contrarios) en ambientes diferentes. Es principalmente por todas estas cuestiones por lo que recientemente se ha propuesto una nueva estrategia para el estudio y mejora de los caracteres cuantitativos: la Selección Genómica.

A principios de este siglo, nace la selección genómica (Meuwissen *et al.*, 2001) como una variante para aumentar la eficacia de la selección asistida con marcadores moleculares (o selección genética). La principal característica de esta metodología es que se emplean una multitud de marcadores distribuidos por todo el genoma y que sirven para escrutar una porción muy significativa de la variabilidad genética presente en ese genoma. La selección genómica se llevó a cabo en primer lugar en ganado vacuno pues, gracias a la publicación de la secuencia completa del genoma de bóvidos, comenzaron a aparecer plataformas de genotipado de alto rendimiento que permitían el análisis de miles de SNP a un coste muy razonable. A título de ejemplo, el genotipado de 50.000 SNP para un individuo cuesta unos 200 dólares.

Aunque estamos describiendo una nueva metodología para mejorar la eficacia de los programas de mejora, conocida como selección genómica, debe quedar claro que las fuentes de información siguen siendo los marcadores moleculares y los datos fenotípicos para los caracteres de interés. Las principales diferencias de la selección genómica frente a la selección genética residen en el aumento del número de marcadores, el cambio de la estrategia de análisis y el diferente tipo de poblaciones estudiadas.

En cuanto a la **estrategia de análisis**, hay multitud de variantes que siguen en la actualidad en periodo de prueba. De forma general, en primer lugar se utiliza una **población de entrenamiento** que será genotipada para todos los SNP y de la que se tienen datos fenotípicos y de genealogías. Posteriormente se genotipa la **población de validación** (de la que no hay datos fenotípicos) y en la que se estima el efecto de cada QTL para cada carácter analizado. Se obtiene un **valor genómico estimado de mejora** (GEBV, *Genomic Estimated Breeding Value*) para cada individuo que define su calidad para el carácter estudiado. La selección genómica está todavía en plena fase de crecimiento pero por el momento todos los resultados publicados parecen mostrar que las estimas deducidas a partir de los datos genómicos son mejores que las obtenidas con la medida clásica de datos fenotípicos.

En relación con las **poblaciones** también se pueden apreciar diferencias notables, pues en el caso de la selección genética habitualmente se emplean cruzamientos mientras que en la genómica se estudian poblaciones. Por lo tanto, el espectro de variación observado es mucho más amplio en el caso de la selección genómica con la consiguiente posibilidad de detectar los efectos de más genes a la vez. La deducción del valor de cada individuo no ofrece información alguna sobre la función de los genes que controlan un carácter pero aportan el mejor criterio de selección posible.

En este informe detallaremos las tecnologías de marcadores moleculares así como su utilización en programas de mejora. Se abordarán las peculiaridades del uso de esta tecnología en función del tipo de organismo. Así, dedicaremos apartados independientes a la ganadería, las plantas cultivadas, la silvicultura y la acuicultura. Hay causas, como la posibilidad de realizar autofecundaciones o el mayor número de descendientes a partir de un único cruzamiento en plantas, que provocan que tanto los planteamientos de programas de mejora como la utilización de marcadores sean muy diferentes en función del tipo de organismo, animal o vegetal. Solo a modo de introducción resumiremos cual es el estado actual de la selección genética y genómica en estas áreas:

1. *Agricultura*. Es el área donde se han producido mayores inversiones por parte de las compañías del sector apostando abiertamente por la MAS (Bernardo, 2008). Hay notables diferencias en función del cultivo. La selección genómica está creciendo en función de los datos de SNP disponibles (Jannink y col., 2010).
2. *Ganadería*. Hay también una clara apuesta por el uso de marcadores moleculares en este sector (Misztal, 2006). Hay numerosos ejemplos de selección de un único gen de gran importancia. También se están realizando esfuerzos para mejorar el análisis de caracteres cuantitativos (producción de leche, prolificidad, etc.) que permitan la aplicación de MAS. El concepto de selección genómica se acuñó en la mejora de especies ganaderas y es donde más está evolucionando en la actualidad (Goddard y Hayes, 2009).
3. *Silvicultura*. Debido a la particular complejidad del análisis de especies forestales, con tiempos de generación mucho más largos, el desarrollo de los programas de mejora están notablemente más atrasados (Butcher y Southerton, 2007). A pesar de ello, ya hay mucha información genética en especies como el eucalipto, pinos o acacias. Precisamente debido a esos largos tiempos de espera hasta que se pueden analizar los caracteres de interés, en árboles la aplicación de la selección genética y genómica se muestra como una herramienta de extrema utilidad (Grattapaglia y Resende, 2010).
4. *Acuicultura*. Hay mapas genéticos en gran parte de las especies de peces pero su grado de densidad es en muchos casos insuficiente (Chistiakov y col., 2006). Hay diferentes genes identificados pero aún se precisa un nivel mayor de inversión y esfuerzo para implementar esta tecnología en peces. La selección genómica está en estado embrionario (Sonesson y Meuwissen, 2009).

Aún queda mucho camino por recorrer para sacarle todo el partido posible al conocimiento de la información de los genomas. En el salto de la selección genotípica a la genómica ya comienza a adivinarse una verdadera revolución en el estudio y manejo de los caracteres de interés, pero el paso definitivo está aún por llegar, más temprano que tarde.

2. Descripción de tecnologías. Técnicas de genotipado de SNP

Durante las tres últimas décadas se ha venido postulando la utilización de marcadores moleculares en los programas de mejora y la implementación de métodos selección asistida por este tipo de marcadores (MAS) en las principales especies animales y vegetales de interés productivo.

Las alozimas de los años setenta dieron paso a los primeros marcadores de ADN en la década de los ochenta, fundamentalmente detectados como polimorfismos a la digestión con endonucleasas de restricción y analizados mediante electroforesis (RFLP). La búsqueda de marcadores asociados a caracteres de interés se enfrentaba con la escasa densidad de los mapas genéticos existentes y el limitado número de genes caracterizados en estas especies. La construcción de mapas genéticos sufrió una rápida transformación al inicio de la década de los noventa mediante la caracterización de marcadores del tipo microsatélite o STR (repeticiones cortas de 2 a 5 nucleótidos en tándem) que permitió la identificación de los primeros QTL (regiones del genoma con efectos significativos sobre un carácter cuantitativo). Sin embargo, debemos admitir que, salvo algunos casos de marcadores asociados a genes de efecto mayor, la utilización de MAS ha contribuido modestamente a la mejora de los caracteres cuantitativos de naturaleza poligénica.

En esta última década los avances en las técnicas de laboratorio para el análisis masivo del genoma han permitido que la caracterización de marcadores genéticos y el conocimiento de la secuencia del genoma de las principales especies de interés agrícola y ganadero hayan experimentado un desarrollo exponencial.

En este contexto, la utilización de paneles de marcadores genéticos en programas de mejora se ha convertido en una práctica habitual en la mayor parte de especies animales y vegetales de interés productivo.

En este capítulo revisaremos las principales técnicas que han permitido estos avances y las metodologías de laboratorio que disponemos actualmente para su utilización, en particular el empleo masivo de marcadores genéticos para su implementación en los planes de mejora que han supuesto un cambio de paradigma frente a los esquemas clásicos basados en las herramientas de la genética cuantitativa.

Así, un suficiente número de marcadores genéticos moleculares analizados, combinados con información fenotípica, permiten realizar inferencias sobre el mérito genético o la predicción de valores fenotípicos para los caracteres de interés en las poblaciones objeto de selección. Estas técnicas cuando se usan de forma masiva con marcadores que cubren todo el genoma han venido a denominarse de selección genómica (SG) (Meuwissen *et al.*, 2001; Lee *et al.*, 2008).

En su aproximación más clásica la SG permite estimar los efectos de un elevado número de marcadores que cubren adecuadamente el genoma a partir de individuos de los que también se dispone de registros fenotípicos para el carácter objeto de interés. A partir de esta información es posible realizar la estima del valor genético de candidatos a la selección directamente mediante el genotipo de sus marca-

dores y sin la necesidad de registros fenotípicos. A las ventajas económicas de reducir los registros fenotípicos esta aproximación permite añadir una considerable reducción de los intervalos generacionales con el correspondiente incremento del progreso genético en las poblaciones objeto de selección.

Por estos motivos, la SG se ha implementado con inusitada rapidez inicialmente en los programas de selección del ganado vacuno lechero ya que permite reducciones de costes de hasta el 90% en los actuales esquemas de selección de sementales (Schaeffer, 2006; Hayes *et al.*, 2009). Esta rápida implementación de la SG ha generado un elevado interés para su desarrollo metodológico y su aplicación en la mejora genética de otras especies animales (Goddart y Hayes, 2009; Hayes y Goddart, 2010), plantas (Heffner *et al.*, 2009; Jannink *et al.*, 2010) y especies de interés acuícola (Sonesson y Meuwissen, 2009; Nielsen *et al.*, 2009, 2010).

El desarrollo de las técnicas de SG ha sido posible gracias a las nuevas tecnologías de secuenciación y genotipado de alto rendimiento que en los últimos años han experimentado una verdadera revolución. Así, en tres décadas el genotipado de marcadores moleculares de ADN ha podido pasar de técnicas laboriosas y de bajo rendimiento basadas en polimorfismos en las dianas de corte de endonucleasas de restricción (RFLP) a los actuales paneles de cientos de miles de polimorfismos de un solo nucleótido (SNP) analizados de forma simultánea en protocolos rápidos y fiables.

Con la secuenciación del genoma de las especies de mayor interés productivo ya completada o en vías de finalización, tanto en animales (pollo, vaca, oveja, cerdo...) como en vegetales (arroz, maíz, trigo, vid, manzana, melón...), las actuales plataformas de secuenciación masiva permiten la resecuenciación completa de un genoma a un coste inferior a los 30.000 \$. Actualmente la disminución de costes en secuenciación va paralela al incremento de necesidades de almacenamiento y análisis de los datos, por lo que resulta probable en un futuro próximo que resulte mas económica que los costes asociados al mero almacenaje de la información producida (Garvey 2010).

Paralelamente, el desarrollo de protocolos de captura selectiva del ADN permite realizar la resecuenciación parcial de regiones del genoma (exones, regiones de interés, etc.) a unos costes que hacen actualmente posible el estudio de un elevado número de muestras necesario para la implementación de esquemas de mejora genética (Turner *et al.*, 2009; Fu *et al.*, 2010).

Los protocolos existentes para la captura selectiva de regiones de interés se basan mayoritariamente en sistemas de hibridación (Albert *et al.*, 2007; Hodges *et al.*, 2007), en *arrays* (*NimbleGen Array Capture*, *Febit HybSelect*) o en solución (*NimbleGen*, *Agilent SureSelect*), mediante miles de oligonucleótidos que cubren adecuadamente las regiones seleccionadas y que permiten capturar regiones de hasta 5-10 Mb. Aunque también se han descrito sistemas basados en PCR múltiple (Porreca *et al.*, 2007; Tewhey *et al.*, 2009) su implementación ha sido escasa debido al elevado número de PCRs necesarias para capturar regiones de un cierto tamaño. Una de las aplicaciones de estos sistemas comerciales ya disponible en el mercado para diversas especies consiste en la captura selectiva del exoma (conjunto de exones del genoma) de un organismo para su ulterior secuenciación (Mamanova *et al.*, 2010).

Los estudios del transcriptoma en especies domésticas, inicialmente realizados mediante la tecnología de *microarrays* para el análisis simultáneo de varios de miles de

genes en plataformas comerciales como los *GeneChip*[®] (*Affymetrix*) se están viendo también sustituidas por aproximaciones de secuenciación masiva (RNA-seq), sin las restricciones inherentes al contenido de los *microarrays*, permitiendo la elaboración de verdaderos catálogos de la expresión génica en tejidos diversos o estados fisiológicos de interés (Harhay *et al.*, 2010). La secuenciación masiva de ARN de una muestra permite también identificar la existencia de patrones alternativos en el ensamblaje de exones de un gen en tejidos o individuos distintos que pueden producir proteínas distintas.

Estas aplicaciones han sido posibles gracias a la aparición en el año 2005 de una nueva generación de plataformas de secuenciación masiva (NGS). La posibilidad de secuenciar de forma rápida y asequible el genoma de un animal o planta abre nuevas perspectivas a la utilización de información molecular en los esquemas de mejora genética.

Las tres plataformas comerciales de secuenciación existentes (*Roche*, *Illumina*, *Life Technologies*) han experimentado una rápida evolución desde su aparición (2005-2007) en sus prestaciones y rendimiento (longitud y cantidad de secuencia) en una carrera probablemente incrementada por la inminente aparición de una tercera generación de plataformas de secuenciación masiva.

Tabla 1
TECNOLOGÍAS DE SECUENCIACIÓN MASIVA

	2006	2007	2008	2009	2010
454/Roche	GS20 20 Mb/run 100 bp/read	GSFLX Standard 100 Mb/run 250 bp/read		GSFLX Titanium 500 Mb/run 500 bp/read	
Solexa/Illumina		GAI 1 Gb/run 32 bp/read	GAIi 8 Gb/run 50 bp/read	GAIix 30 Gb/run 100 bp/read	Hiseq1000 / Hiseq2000 100 Gb/run 200 Gb/run 100 bp/read
Life Technologies		SOLiD 3 Gb/run 35 bp/read	SOLiD 2 6 Gb/run 50 bp/read	SOLiD 3 20-60 Gb/run 50 bp/read	SOLiD 4 SOLiD 5500/SOLiD 5500xl 100 Gb/run 90 Gb/run 180Gb/run 60 bp/read 100 bp/read

En sus inicios las longitudes de secuencia de los equipos de Illumina y SOLiD estaban sobre los 35 pares de bases con un coste de 2\$ por millón de bases secuenciadas (Shendure and Ji, 2008), sin embargo ambas plataformas actualmente pueden producir lecturas de hasta 125 pares de bases (pb) con posibilidad de realizar secuenciación de 85 pb en ambos extremos de fragmentos de 300 pb, generando un total de hasta 180Gb de secuencia que permiten cubrir la resecuenciación completa de un genoma animal o vegetal en un solo experimento. Paralelamente han aparecido en el mercado nuevos equipos de secuenciación mucho más económicos, como el secuenciador *GS Junior*[®] (*Roche*) o el nuevo secuenciador *Ion Torrent*[®] (*Life Technologies*), con rendimientos inferiores (entre 70.000 y 100.000 secuencias de unos 400 pb en el caso del *GS Junior* y actualmente unas 100.000 secuencias de unos 100 pb en el caso de *Ion Torrent*). Sin embargo, existen aplicaciones (rese-

cuenciación selectiva, metagenómica, diagnóstico...) en las que estos niveles de rendimiento resultan suficientes.

En el caso del nuevo secuenciador de *Ion Torrent*, además de un precio asequible para laboratorios de mediano tamaño (unos 70.000\$ incluyendo el servidor para procesar los datos) resulta destacable la velocidad de cada carrera (unas 2h) y el bajo coste de la misma (unos 500\$). Esto es posible al no requerir de marcado fluorescente ni sistemas ópticos para su detección. La secuenciación se realiza en chips convencionales, susceptibles de detectar los cambios de pH que se producen al incorporar de forma secuencial nucleótidos convencionales sobre los fragmentos a secuenciar previamente amplificados por una PCR en emulsión.

Las características de economía y rapidez de esta tecnología resultan particularmente atractivas para muchas de las aplicaciones de diagnóstico genético basadas actualmente en genotipado.

Este rápido desarrollo de estas técnicas de secuenciación de alto rendimiento ha significado una verdadera revolución en las técnicas de genotipado (Mardis, 2008) ya que han permitido la identificación de miles de nuevos SNP de forma rápida y eficiente (Hodges *et al.*, 2007; Van Tassell *et al.*, 2008; Wheeler *et al.*, 2008) además de otros tipos de variantes estructurales del ADN como inserciones, deleciones y variación en el número de copias (CNV) (Campbell *et al.*, 2008). Con el incremento de genomas resecuenciados cabe esperar un aumento exponencial de las bases de SNP y en disponibilidad de sistemas de genotipado con un creciente número de marcadores en las principales especies de interés. Así, en la especie humana la reciente publicación de los primeros resultados del proyecto de secuenciación de 1.000 genomas (*1000 Genomes Project Consortium*, 2010) ha puesto de manifiesto la existencia de unos 15 millones de SNP que permitirán en un futuro inmediato la existencia de chips de genotipado de hasta 5 millones de SNP (*Infinium Omni5 BeadChip*, *Illumina*). Este incremento de SNP en las plataformas de genotipado permite incorporar variantes con menores valores de frecuencia en las poblaciones, incrementando el poder de detección en estudios de asociación (GWAS). En una aproximación parecida se han publicado estudios de resecuenciación en variedades de arroz y maíz. En un trabajo de resecuenciación 1x de 517 variedades de arroz se han podido identificar unos 3,6 millones de SNP que se estima representan un 80% de la diversidad existente en esta especie y elaborado un detallado mapa de haplotipos (Huang *et al.*, 2010). Utilizando este mapa se han realizado estudios de asociación (GWA) para 14 caracteres de interés agronómico detectándose hasta 80 marcadores significativos que en promedio explicaban hasta un 36% de la varianza fenotípica para cada uno de ellos (valores muy superiores a los habitualmente obtenidos en estudios equivalentes de GWA en la especie humana).

En el caso del maíz, la resecuenciación (a un nivel 5,4x) de 6 líneas comerciales ha permitido identificar más de 1,2 millones de SNP y unas 30.000 variantes estructurales del tipo inserción/delección y lo que resulta más sorprendente, la existencia de numerosos genes presentes en unas variedades y ausentes en otras (variaciones de presencia-ausencia, PAV) que quizás podrían explicar junto a otros polimorfismos la heterosis en líneas híbridas observada en esta especie (Lai *et al.*, 2010).

Estos ejemplos ilustran el enorme poder de las técnicas de secuenciación masiva para la caracterización de variantes en especies de interés y el rápido desarrollo de protocolos de genotipado masivo de SNP en las principales especies domésticas.

Los chips de genotipado masivo: una herramienta potente para la implementación de la selección genómica

El primer sistema de genotipado de alta densidad en una especie doméstica fue el chip 10K (de 10.000 SNP) comercializado por *Affymetrix* (The Bovine HapMap consortium, 2009). Sin embargo, este número de SNP resultaba insuficiente para estudios genómicos de asociación (GWAS) o para la implementación de GS poniendo de manifiesto la necesidad de un chip con mayor densidad de marcadores. El chip de *Illumina Bovine SNP50* (Matukumalli *et al.*, 2009) con cerca de 50.000 SNP fue desarrollado por un consorcio de laboratorios mediante la utilización de secuenciación masiva para la caracterización de nuevos SNP en diversas razas bovinas. El protocolo utilizado (Van Tasell *et al.*, 2008), permite el descubrimiento de SNP a partir de secuenciar los fragmentos generados por la digestión de un pool de ADN genómicos con un enzima de restricción. Los fragmentos de ADN son previamente seleccionados para un cierto intervalo de tamaño mediante separación electroforética. Esta aproximación ha sido también utilizada para la caracterización de nuevos SNP en otras especies (Ramos *et al.*, 2009; Myles *et al.*, 2010) demostrando su utilidad para la obtención de chips de genotipado de alta densidad en especies en las que ya se dispone de un genoma de referencia secuenciado.

¿Necesitamos chips de mayor densidad?

En el caso de los mamíferos con un genoma de 3.000 millones de pb con un chip 50K esperamos encontrar un SNP cada 60 Kb, si disponemos de un chip 800K la distancia entre marcadores puede reducirse a intervalos de menos de 4 Kb. Estas distancias entre marcadores permiten estimas mucho más precisas en aplicaciones de GS o estudios de asociación (GWAS). La utilización de chips de alta densidad (en la especie humana próximamente dispondremos de chips de genotipado de hasta 5 millones de SNP) permite obtener resultados de marcadores que pueden estar en baja frecuencia en la población estudiada, permitiendo su utilización en razas o variedades distintas en las que las frecuencias génicas de los marcadores pueden diferir significativamente.

Otra ventaja de los chips de alta densidad consiste en la posibilidad de establecer con una razonable precisión los haplotipos existentes en las muestras genotipadas. La determinación de estos haplotipos permite con una razonable probabilidad la inferencia de genotipos en muestras analizadas para paneles más reducidos de SNP a menor coste. Así en el caso de la especie bovina la reciente aparición de un nuevo chip de genotipado de alta densidad *Illumina BovineHD* (con más de 700.000 SNP) se ha visto acompañada con la de otro chip de baja densidad *Illumina Bovine 3K* (con unos 3.000 SNP) que permite inferir haplotipos con una precisión aceptable y a un coste muy razonable.

En el caso del ganado bovino, la única especie productiva en la que la SG se encuentra completamente implementada, el escenario actual contempla el genotipado mayoritario de toros con el chip de 50.000 SNP complementado en menor medida con el de 700.000 SNP y un creciente número de hembras genotipadas con el chip de baja densidad.

El horizonte de la selección genética y genómica en especies productivas contempla en consecuencia la coexistencia de técnicas de genotipado masivo (fundamentales para la evaluación genómica) con genotipados de baja intensidad que por una parte cubran el diagnóstico de los genes de mayor efecto y que permitan además la realización de inferencias haplotípicas.

La detección de SNP se fundamenta en una de estas 4 aproximaciones:

- (1) Hibridación alelo-específica.
- (2) Extensión alelo-específica.
- (3) Ligación alelo-específica de oligonucleótidos.
- (4) Digestión alelo-específica.

Existen numerosas plataformas y protocolos para el genotipado de SNP basadas en alguna de estas 4 aproximaciones metodológicas. La elección de cada una de ellas depende fundamentalmente de la combinación entre el número de muestras y el número de marcadores que necesitemos analizar.

En esta revisión nos limitaremos a describir las metodologías mayoritariamente empleadas en las especies animales y vegetales de interés productivo en aproximaciones de MAS, GWAS y GS.

2.1. Protocolos basados en la utilización de equipos de PCR a tiempo real

que permiten el genotipado sin requerir el procesamiento ulterior del producto amplificado lo que reduce el riesgo de contaminación, facilitando además su automatización.

En la actualidad, existen en el mercado varios equipos y protocolos para la realización de esta técnica. Desde el punto de vista de la detección del producto amplificado, existen tres métodos principales de detección basados en técnicas de fluorescencia y que se diferencian en el tipo de detección de los productos de PCR. Estos métodos son: el que emplea sondas alelo-específicas con doble marcado fluorescente (*TaqMan*[®]), los basados en el uso de dos sondas que hibridan de forma adyacente a la posición polimórfica en el fragmento que amplificamos (*HybProbes*[®] Roche), ambos basados en el principio de la transferencia de energía entre fluorocromos (FRET) y, por último, el que utiliza fluorocromos intercalantes que se unen a la doble cadena de ADN (*LCGreen* o *EvaGreen*), produciendo emisión de fluorescencia.

Las tres metodologías de detección de producto amplificado en aplicaciones de PCR cuantitativa han sido adaptadas al genotipado de SNP.

2.1.1. Empleo de sondas alelo-específicas con doble marcado fluorescente

El método más difundido, es el que emplea sondas *TaqMan*[®] (Gelfand *et al.*, 1993). Este método se basa en la actividad 5'-exonucleasa de la Taq polimerasa y en la amplificación mediante PCR de una determinada secuencia diana en presencia de dos sondas fluorescentes alelo-específicas (sondas *TaqMan*[®]) que hibridan con la secuencia diana del SNP que estamos amplificando.

La sonda *TaqMan*[®], de un tamaño aproximado de 20-30 bases, tiene unido un fluorocromo en posición 5' (*Reporter*) y un segundo fluorocromo que actúa por interferencia con el primero como amortiguador de fluorescencia en posición 3' (*Quencher*). Además, esta sonda está fosforilada en su extremo 3' para evitar su extensión du-

rante la reacción de PCR. Para reducir el tamaño de las sondas y minimizar la señal fluorescente de fondo se utilizan oligonucleótidos conjugados (*MGB-TaqMan*[®]) que han demostrado una mayor fiabilidad en la discriminación alélica (Kutyavin *et al.*, 2000). Si la secuencia diana (alelo) está presente en la muestra, la sonda *TaqMan*[®] correspondiente hibridará específicamente con ella, situándose entre los dos cebadores. Cuando se produce la etapa de extensión en la reacción de PCR, la actividad 5'-exonucleasa de la Taq polimerasa degrada a la sonda *TaqMan*[®] liberando el fluorocromo, que al quedar separado del amortiguador emitirá una señal fluorescente específica para cada una de las sondas, que podrá ser captada por el sistema óptico del equipo. Este proceso de degradación de la sonda tiene lugar en cada uno de los ciclos y es directamente proporcional al número de moléculas que están siendo extendidas en cada uno de ellos y que podremos visualizar por el incremento de la señal de fluorescencia al final de los ciclos de amplificación.

Además, la Taq polimerasa no digiere la sonda libre sino únicamente la hibridada, por lo que la cantidad de señal fluorescente emitida es proporcional a la cantidad de producto específico acumulado que podremos correlacionar directamente de forma objetiva y reproducible con el genotipo de la muestra analizada. Los ensayos de genotipado mediante este tipo de sondas no permiten el análisis simultáneo de varios SNP (multiplexado) y debido al coste del doble marcaje fluorescente no resultan apropiados cuando el número de SNP que debamos analizar sea elevado. Sin embargo, la simplicidad y fiabilidad del protocolo lo hacen muy recomendable para ensayos rutinarios de un elevado número de muestras.

Aunque menos utilizadas que las sondas *TaqMan*[®], existen otros tipos de sondas con el mismo principio de doble marcado fluorescente (*Reporter-Quencher*) que solo emiten fluorescencia durante la hibridación con la correspondiente secuencia alélica del SNP al presentar una estructura no lineal por la existencia de secuencias complementarias en sus extremos como las sondas *Molecular Beacons*[®] (Tyagi *et al.*, 1998) o las sondas *Scorpion*[®] que además de la estructura en horquilla incorporan uno de los cebadores de la PCR en su extremo 5' (Thelwell *et al.*, 2000).

Actualmente existen además dos plataformas de alto rendimiento en forma de chip para la realización de un elevado número de reacciones de forma simultánea: *OpenArray*[®] de *Life Technologies* y *EP1 System*[®] de *Fluidigm*.

Basadas en el mismo tipo de químicas que las empleadas en los equipos de PCR cuantitativa convencionales estas plataformas utilizan volúmenes de reacción muy inferiores (en la escala de nanolitros), en forma de reactivos liofilizados o utilizando técnicas de microfluídica en el chip, con el consiguiente ahorro de reactivos y disminución del coste de los análisis.

OpenArray[®] (Life Technologies)

El sistema utiliza la metodología TaqMan de genotipado en un sistema de portaobjetos metálicos que presentan 3.072 puntos de análisis (organizados en 48 conjuntos de 64 pocillos de 33nl de capacidad cada uno) en los que se realizará la reacción de amplificación. Los portaobjetos se suministran con las sondas liofilizadas y el sistema permite la carga del resto de reactivos y del ADN problema de forma automatizada. El formato de análisis permite diseñar en cada portaobjetos desde un mínimo de 16 ensayos para 144 muestras hasta un máximo de 256 ensayos para un total

de 12 muestras. El equipo es capaz de realizar el análisis simultáneo de hasta 3 portaobjetos (8.064 reacciones) que equivalen a un total de 84 placas de 96 pocillos con un considerable ahorro de tiempo y coste en reactivos.

EP1 System® (Fluidigm)

Chips basados en tecnología de microfluidos que permiten volúmenes de reacción en la escala de nanolitros. Existen dos formatos de *array* de 48 x 48 y de 96 x 96 (muestras x ensayos) que producen un total de 2.304 y 9.216 resultados por ensayo. Los *arrays* constituyen una plataforma abierta en la que podemos utilizar cualquier aplicación basada en PCR convencional o cuantitativa (Wang *et al.*, 2009).

Fluidigm comercializa adicionalmente otra plataforma (*BioMark*®) que además del genotipado de SNP con sondas *TaqMan* puede utilizarse para otros tipos de aplicaciones (estudios de expresión génica por PCR cuantitativa, PCR digital y preparación de muestras para secuenciación de alto rendimiento).

Una de las ventajas competitivas del sistema radica en la posibilidad de reutilización de los chips para aplicaciones de genotipado. En una colaboración con investigadores del Departamento de Agricultura de Estados Unidos (USDA) Fluidigm ha lanzado recientemente chips capaces de ser utilizados hasta un máximo de 5 veces lo que comporta una considerable reducción en el coste de los análisis, factor crítico en la mayor parte de las aplicaciones de selección genética en especies productivas.

2.1.2. Genotipado mediante PCR alelo-específica y detección fluorescente

Basadas conceptualmente en la posibilidad de discriminar entre dos alelos por PCR mediante la utilización de dos cebadores específicos, complementarios cada uno en su extremo 3' a la variante alélicas del SNP (PCR alelo-específica), recientemente se han desarrollado técnicas que no requieran la utilización de sondas fluorescentes específicas para cada SNP.

Uno de los protocolos más eficientes es la técnica *KASPar*® (*KBiosciences*). Para cada SNP el protocolo requiere la síntesis de un trío de oligonucleótidos no marcados que será utilizado para una PCR alelo-específica. Dos de ellos actuarán como cebadores para cada uno de los 2 alelos y presentan una secuencia específica y distinta para cada uno en su extremo 5' (cola).

El otro componente del protocolo son 4 oligonucleótidos de utilización universal. Dos de ellos marcados fluorescentemente en su extremo 5', uno con el fluoróforo *FAM* y el otro con *CAL Fluor Orange 560*. La secuencia de cada uno de estos oligonucleótidos es complementaria a las colas de cada uno de los cebadores alelo-específicos.

Los dos oligonucleótidos restantes, con secuencia complementaria a cada uno de los anteriores, están marcados en su extremo 3' con fluorocromos que actúan como apagadores.

Al inicio de la PCR no hay emisión de fluorescencia por la complementaridad de estos 2 pares de oligonucleótidos universales. A medida que avanza la PCR y los productos amplificados incorporan la cola de los cebadores.

2.1.3. Genotipado mediante la utilización de fluorocromos intercalantes

Basada en técnicas de discriminación de la cinética de desnaturalización del ADN heteroduplex, recientemente ha ganado popularidad por su bajo coste y simplicidad la técnica denominada HRM (*High resolution melting*) para el genotipado de SNP en amplicones de pequeño tamaño (Witwer, 2009). La técnica requiere de un equipo de PCR a tiempo real capaz de realizar curvas de disociación del producto amplificado y de la utilización de sustancias fluorescentes con propiedades intercalantes sobre el ADN. Este tipo de análisis de disociación del producto amplificado para discriminar variantes alélicas fue inicialmente introducido por Ririe *et al.*, (1997) y ampliamente utilizado en protocolos de PCR cuantitativa en los equipos *LightCycler*[®] de Roche. La estabilidad de los amplicones de ADN era monitorizada por los niveles de fluorescencia emitidos por agentes intercalantes como el SYBR Green I a medida que se producían incrementos de temperatura y se realizaba una cinética de desnaturalización del producto amplificado monitorizada por la correspondiente disminución de la señal fluorescente emitida.

Las características e identidad del producto amplificado podían estimarse mediante la T_m (temperatura a la cual el 50% de las moléculas estaban desnaturalizadas) del mismo. Sin embargo, diferencias mínimas de sólo un nucleótido entre dos fragmentos de ADN se consideraban por debajo de la capacidad de discriminación en las cinéticas de desnaturalización. La introducción de nuevos agentes intercalantes, como *LCGreen* (*Idaho Technology*) o *EvaGreen* (*Biotium, Inc.*), capaces de ser utilizados a concentraciones de saturación sin efectos inhibidores sobre la PCR permitieron mejorar y optimizar el poder de discriminación y su utilización en el genotipado de SNP (Wittwer *et al.*, 2003; Liew *et al.*, 2004). Los sutiles cambios de T_m producidos por un cambio nucleotídico generan diferencias en el perfil de las curvas de disociación para cada uno de los genotipos siendo posible la discriminación entre muestras homocigotas y heterocigotas de forma sencilla y fiable.

Una de las principales ventajas del método reside en su bajo coste al no requerir de sondas específicas marcadas fluorescentemente y en su simplicidad y rapidez, realizándose el análisis post-PCR en el mismo vial de amplificación, lo que minimiza los riesgos de error y contaminación (Farrar *et al.*, 2009). A pesar de requerir una puesta a punto previa de la técnica para cada SNP y a la imposibilidad de realizar genotipados múltiples HRM está ganando popularidad para el análisis rutinario de marcadores y para la detección económica de nuevas variantes en genes de interés en los ámbitos ganadero, acuícola, agrícola y forestal (Cánovas *et al.*, 2010; Borza *et al.*, 2010; Muleo *et al.*, 2009; Li *et al.*, 2010).

2.2. Protocolos de genotipado de alto rendimiento

2.2.1. Protocolos basados en ligación alelo-específica

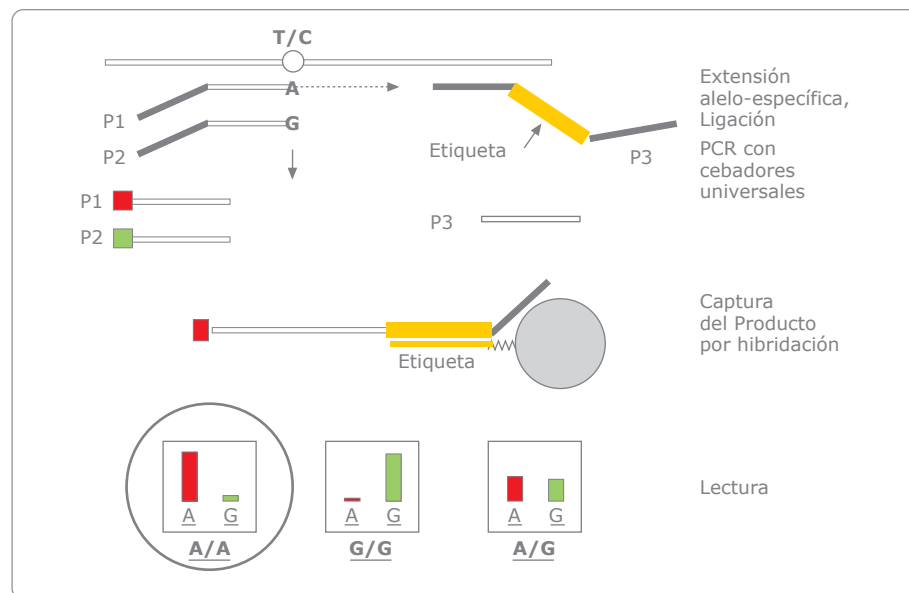
Para proyectos con un elevado número de muestras y/o SNP uno de los protocolos más utilizados se basa en la ligación alelo-específica de nucleótidos y en la detección mediante fluorescencia por hibridación en soportes inmovilizados (Fan *et al.*, 2003). El protocolo comercializado como *Golden-Gate*[®] (*Illumina*) permite el genotipado de entre 96 y 1.536 SNP en formato de 96 muestras mediante la plataforma

iScan[®] y tecnología *BeadArray*[®] o entre 48 y 384 SNP mediante la plataforma *BeadXpress*[®] y tecnología *VeraCode*[®].

El protocolo se inicia con la hibridación sobre ADN genómico de tríos de oligonucleótidos específicos para cada SNP, dos de ellos específicos para cada uno de los 2 alelos que, en caso de complementariedad en su extremo 3' con el ADN molde, podrán ser extendidos y ligados con el tercer oligonucleótidos, específico para cada SNP y que contiene una secuencia "etiqueta" que permitirá identificar cada una de las reacciones. Los productos de ligación son amplificados de forma universal por PCR mediante cebadores que incorporan marcado fluorescente específico para cada uno de los 2 alelos y el producto de amplificación es capturado por hibridación mediante secuencias específicas en esferas que contienen oligonucleótidos complementarios a cada una de las secuencias "etiqueta" que identifican el producto de ligación de cada SNP. En esta técnica cada esfera define el ensayo y cada una de las dos fluorescencias el alelo correspondiente (Figura 1).

En el sistema *BeadXpress*[®] la captura de los productos de hibridación se realiza en el interior de cilindros de vidrio específicos para cada SNP que la plataforma puede discriminar mediante holograma y en los que el tipo de fluorescencia existente permite determinar el genotipo de cada uno de ellos. Mediante este tipo de ensayos es factible la obtención de hasta 300.000 resultados de genotipado por persona y día.

Figura 1
ILLUMINA GOLDEN GATE[®]



2.2.2. Protocolos basados en extensión de base

La mayor parte de métodos utilizan la extensión de un oligonucleótido alelo-específico en una sola base que corresponde a la de la posición polimórfica que deseamos analizar. Conocidos como métodos de "mini-secuenciación" se diferencian en el tipo y formato de detección del producto extendido. Estos protocolos requieren la reali-

zación de diversos pasos hasta llegar al resultado final y resultan lógicamente más laboriosos que los métodos basados en PCR cuantitativa, aunque se caracterizan por la posibilidad de realizar simultáneamente el genotipado de múltiples SNP. Existen diversas plataformas comerciales para la detección y lectura de los productos de extensión: emisión de luz por pirosecuenciación (*Biotage*), citometría de flujo (*Luminex*), fluorescencia por polarización (*PerkinElmer*), electroforesis capilar (*Life Technologies*) y espectrometría de masas (*Sequenom*), siendo estas dos últimas las más utilizadas en el ámbito agrícola y ganadero.

a. Protocolos basados en extensión alelo-específica

El más utilizado recibe el nombre comercial de SNaPshot® (Life Technologies) y utiliza en la reacción de extensión dideoxi-nucleótidos marcados fluorescentemente (ddNTPs). Una vez realizada una amplificación por PCR del fragmento a analizar y eliminados los restos de cebadores y nucleótidos de la misma, se realiza la extensión de un oligonucleótido de secuencia complementaria a la región adyacente a la base polimórfica. El tamaño del producto obtenido es el del cebador inicial más el de la base fluorescente incorporada que, al ser un dideoxinucleótido, impide la incorporación de bases adicionales. El producto se puede visualizar mediante electroforesis capilar en equipos de secuenciación convencionales. Existe la capacidad de realizar la detección simultánea de diversos SNP (normalmente hasta un máximo de 10) mediante el diseño de cebadores de distinto tamaño. Utilizando equipos de secuenciación de 96 capilares es posible generar hasta un máximo de 10.000 genotipos por persona y día.

b. Protocolos con marcado fluorescente y separación electroforética

Basados en el mismo principio de mini-secuenciación mediante la utilización de dideoxinucleótidos fundamenta la detección del producto extendido mediante la utilización de un equipo de MALDITOF MS (*Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry*) capaz de discriminar las diferencias en masa atómica del oligonucleotido extendido sin la necesidad de marcaje fluorescente.

El equipo *MassARRAY® Analyzer 4 System* (*Sequenom*) es capaz de realizar detecciones múltiples de hasta 36 SNP de forma simultánea, en formatos de 96 o 384 muestras lo que permite la realización de hasta 138.000 genotipos/día. El sistema, aunque requiere la adquisición de un equipo de coste elevado, ofrece un alto rendimiento y niveles de flexibilidad para proyectos en el que el número total de SNP a analizar por muestra sea inferior a 3.000.

c. Protocolos con separación del producto extendido por espectrometría de masas

Los protocolos de genotipado masivo *Infinium I* (Gunderson *et al.*, 2005) e *Infinium II* (Stemmers *et al.*, 2006) se fundamentan en reacciones de extensión de base producidas en oligonucleótidos específicos e inmovilizados en esferas depositadas en los chips. Ambos protocolos incorporan una amplificación masiva inicial del ADN (WGA) seguida de hibridación en el *array* que contiene sondas de captura de 50 pb específicas para cada SNP. En el protocolo *Infinium I* las secuencias de captura incluyen en su extremo 3' el nucleótido específico para cada alelo. Una reacción de extensión alelo-específica permite la incorporación de nucleótidos marcados con biotina para la poste-

rior detección de la reacción de extensión. En los ensayos basados en *Infinium II* las esferas contienen sondas específicas para cada locus y la discriminación alélica ocurre mediante una reacción de extensión mediante dideoxinucleótidos marcados fluorescentemente. El sistema de 2 colores utilizados por *Infinium II* restringe el análisis de algunos SNP (los cambios A/T y C/G no pueden ser genotipados mediante este protocolo).

Esta plataforma de genotipado se ha convertido en el sistema de referencia para la implementación de protocolos de selección genómica en especies domésticas.

2.2.3. *Microarrays de alto rendimiento*

Mediante el sistema de *arrays* de genotipado *GeneChip*[®] de *Affymetrix* cada SNP es analizado mediante su hibridación frente a un conjunto de oligonucleótidos cortos posicionados en el *array*. Este conjunto contiene secuencias complementarias para cada uno de los alelos además de oligonucleótidos que incorporan cambios nucleotídicos adicionales para discriminar la intensidad de la señal de hibridación (desde hibridación completa a hibridación parcial por la existencia de una o más bases no complementarias) y poder asignar el genotipo a la muestra problema por discriminación alélica en el proceso de hibridación (Kennedy *et al.*, 2003). Para que el protocolo de hibridación funcione eficientemente se requiere previamente una reducción de la complejidad del ADN genómico mediante su digestión con endonucleasas de restricción (Sty I y Nsp I para chips de 500K o superiores) y los fragmentos obtenidos de entre 400 y 800 pb, previa unión a secuencias adaptadoras, se utilizan para ser amplificados y marcados por PCR para ser hibridados posteriormente con el chip. El sistema permite actualmente la detección simultánea de hasta unos 2 millones de SNP.

3. Aplicación de la selección genética y genómica en ganadería

La especie humana ha adaptado los animales para la producción tanto de alimentos (carne, leche, huevos) como de otros productos (lana, cuero, huesos) desde los inicios de la domesticación hace más de 10.000 años. Se espera que el consumo de productos de origen animal aumente en un 7% anual (la mayor parte en los países en vías de desarrollo) ya que, aunque la tasa de crecimiento de la población está disminuyendo, el número de habitantes del planeta aumenta en 80 millones al año y se espera que alcance la cifra de 9.000 millones en 2050. La mejora genética animal juega un papel crucial en la producción animal y si Europa quiere tener algo que decir en este campo debe desarrollar conocimiento y tecnología en un esfuerzo conjunto de los productores, la investigación y la sociedad en su conjunto.

3.1. Consideraciones generales sobre los programas de mejora genética

La mayoría de los rasgos de interés económico de los animales domésticos son caracteres cuantitativos que muestran una variación hereditaria poblacional debida a la segregación de varios *loci*, no necesariamente muchos, cuya expresión puede ser modificada por acción del medio. Esta mediatización ambiental implica que la correspondencia entre fenotipo y genotipo no sea perfecta, de manera que dos individuos con genotipos idénticos pueden tener fenotipos distintos y, por otra parte, que la igualdad fenotípica no lleva forzosamente consigo la genética.

La descripción de los **caracteres cuantitativos** suele hacerse siguiendo el llamado **modelo infinitesimal** propuesto por Fisher en 1918, que está basado en las técnicas de análisis de la varianza debidas igualmente a este autor. El modelo supone que la desviación del valor fenotípico (P) de un carácter en un individuo con respecto a la media de la población panmíctica de la que éste forma parte, es suma de dos componentes residuales, uno genético (G) y otro ambiental (E). Si cada uno de estos efectos se considera como la consecuencia final de la acción de un considerable número de factores independientes cuya influencia individual es suficientemente pequeña, ambos pueden tenerse por variables aleatorias normales. El **efecto genético** o **valor genotípico desviado** (G) puede a su vez fraccionarse como suma de tres variables aleatorias, normales e independientes que, en términos del análisis de la varianza factorial, corresponden a los efectos principales (**valor aditivo o mejorante A**), las interacciones de primer orden (**interacciones intralocus** o **desviación dominante D**) y las de orden superior (**interacciones interloci** o **epistáticas I**).

La puesta en marcha de un programa de selección consiste básicamente en utilizar como reproductores solamente a aquellos animales cuyo valor mejorante es mayor. Con respecto a un determinado carácter, este valor puede estimarse de varias formas. En la práctica, la forma óptima de evaluación es la llamada BLUP-Modelo Animal, desarrollada inicialmente por Henderson¹ en 1948 para el vacuno lechero pero

¹ Henderson, C. R. (1948). *Estimation of general, specific and maternal abilities in crosses among inbred lines of swine*. Iowa State University Press, Ames.

rutinariamente utilizada en la actualidad en vacuno, porcino, caballos, gallinas y, más recientemente, en peces. El **procedimiento BLUP** (“óptima predicción lineal insesgada”) es una generalización de los índices de selección que estima los efectos fijos y los valores mejorantes simultáneamente. Esta metodología (y la metodología REML que se le asocia para estimar parámetros genéticos) ofrece grandes ventajas por lo que su utilización ha tenido un gran impacto en los modernos programas de mejora. En primer lugar, permite corregir adecuadamente los efectos sistemáticos (rebaño, año, estación, edad, etc.), de forma que pueden compararse animales procedentes de distintos grupos, criados bajo diferentes sistemas de manejo, o que difieren en edad. También pueden compararse animales en el caso de que no se disponga de la misma información para cada uno de ellos (cuando cuentan con distinto número de parientes) o que han estado sometidos a distintas presiones selectivas, siempre y cuando estos efectos sistemáticos estén conectados a través de la matriz de parentesco. En segundo lugar, todos los individuos son evaluados genéticamente de forma simultánea, utilizándose en cada caso la información proporcionada por todos los parientes que se hayan registrado en el programa desde su inicio hasta la generación más reciente. En tercer lugar, la separación entre los efectos genéticos y ambientales hace posible la estima del progreso genético obtenido a lo largo del tiempo. Por último, el conocimiento de las genealogías nos capacita para utilizar métodos de control de la consanguinidad.

Los resultados de la aplicación de los programas de mejora a las especies animales domésticas son muy notables, alcanzándose tasas anuales de progreso genético del orden del 1-3% de la media. Por ejemplo, la producción anual por ponedora ha aumentado de 120 huevos a 340 durante los últimos 50 años, y la producción de leche por vaca y lactación ha pasado de 2.000 kg a más de 9.000 en el mismo periodo. En términos monetarios, el informe de la plataforma europea FABRE² atribuye a la mejora genética animal un beneficio de 2000 millones de euros anuales, ligados a una inversión de 150 millones en investigación y desarrollo que es gestionada en su mayor parte a través de convenios con universidades y otros centros de investigación. Aunque la tasa de progreso genético pueda parecer pequeña en relación con las atribuibles a los avances conseguidos en nutrición o manejo, debe tenerse en cuenta que, a diferencia de estos últimos, la ganancia genética es continua, acumulativa y permanente. Por ejemplo, Harvenstein y col.,³ compararon en 1991 los pesos medios de gallinas de carne a la edad comercial de una línea control y otra seleccionada, fundadas ambas en 1957, en las condiciones definidas por la utilización de uno u otro de los piensos típicos consumidos en esas dos fechas. El suministro del pienso de 1991 en lugar del correspondiente a 1957 produjo un aumento de peso del 19%, mientras que la estirpe seleccionada resultó ser un 210% más pesada que el control al cabo de los 34 años transcurridos. Evidentemente, este espectacular incremento del peso vino acompañado por algunas consecuencias indeseables, entre ellas el aumento de la mortalidad debido a la discondroplasia de la tibia, lo cual indica que cambios genéticos extremos para caracteres de interés económico pueden ocurrir enlazados a otros desfavorables referentes a los atributos relacionados con la eficacia biológica.

² FABRE TECHNOLOGY PLATFORM (2006). *Sustainable Farm Animal Breeding and Reproduction-A vision for 2025*.

³ Harvenstein, G.; Ferket, P. R.; Scheideler, S. E. and Rives, D. V. (1994). Carcass composition and yield of 1991 vs 1957 broilers when fed “typical” 1957 and 1991 broiler diets. *Poultry Science*, 73, 1795-1804.

3.2. Biología molecular y mejora genética animal

Apenas treinta años tras el descubrimiento de la estructura del ADN, la aparición de nuevas técnicas de genética molecular marca el comienzo del nuevo campo de la genómica: la disciplina del mapeo, secuenciación y análisis a nivel genómico de la información del ADN. Aprovechando la existencia de marcadores polimórficos denominados microsateélites, distribuidos por todo el genoma, los investigadores han sido capaces de construir mapas genéticos de las especies domésticas y de localizar regiones del genoma donde residen genes que afectan a los caracteres de interés económico. Con el nuevo siglo, los proyectos genoma están finalizando, primeramente en humanos y posteriormente en vacuno, gallinas, perros y caballos (y muy pronto en cerdos y ovejas). Esto está permitiendo la puesta a punto de chips de miles de marcadores SNP que van a ser utilizados para la trazabilidad individual, familiar y racial, para diagnosticar enfermedades y para practicar selección genómica. En lo que sigue revisaremos algunos de estos avances sin referirnos a aspectos más biotecnológicos que también pudieran ser importantes en el campo de la producción animal (clonación, transgénesis, etc.)⁴.

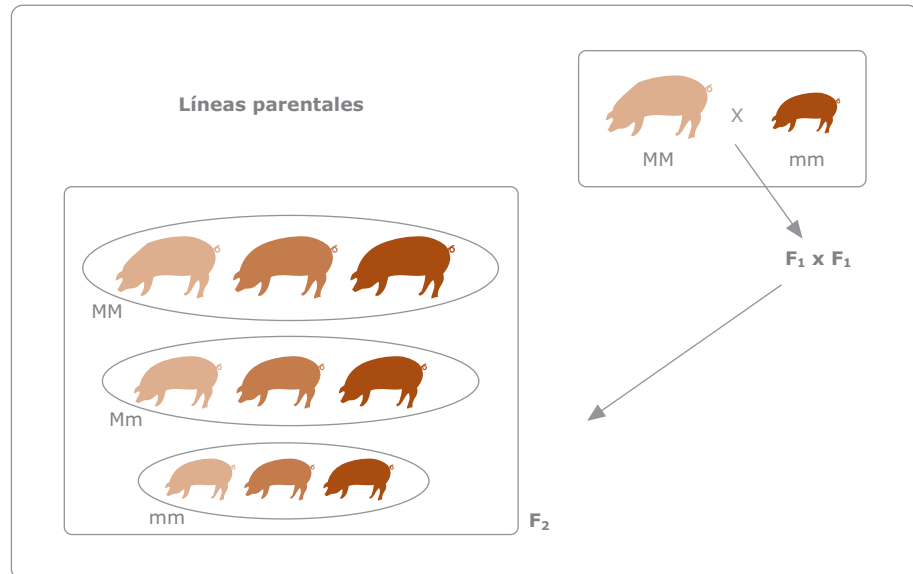
3.2.1. Detección de QTL en las especies domésticas

Denominamos QTL a los *loci* de los caracteres cuantitativos estudiados de forma individual. Un QTL es una región del genoma que alberga uno o varios genes con efecto sobre un carácter cuantitativo. En los años 90 comenzaron los experimentos para detectar QTL utilizando básicamente dos tipos de diseños. El primero utiliza el desequilibrio de ligamiento generado por cruzamientos. Por ejemplo se genera una F1 cruzando dos razas muy distintas: jabalí x cerdo doméstico, cerdo ibérico x cerdo blanco, vacuno de carne x vacuno de leche, razas comerciales de gallinas de puesta x razas de carne, etc. Posteriormente la F1 se cruza para producir una F2 o se retrocruza con una de las razas parentales.

Tal y como se indica en la Figura 2 imaginemos que disponemos de animales de dos razas muy diferentes. Una de ellas blanca y de gran tamaño corporal y otra oscura y de pequeño tamaño. Consideremos ahora un marcador que estaba fijado para alelos alternativos en cada una de las líneas. En el cruzamiento F2 aparecerán animales con todas las combinaciones posibles de alelos y fenotipos. Podemos ahora clasificar a los animales por su genotipo respecto al marcador en las clases MM, Mm y mm. Los animales MM son grandes mientras que los mm son pequeños y los Mm intermedios. Esto sugiere que el marcador M/m está en una región del cromosoma en el que hay algún QTL relacionado con el crecimiento. Este no es el caso para el color, puesto que los distintos colores aparecen con las mismas frecuencias en las diferentes clases genotípicas. Lo mismo haríamos con otros marcadores elegidos de forma que cubran todo el genoma. Naturalmente, en la práctica se requieren técnicas estadísticas más sofisticadas (regresión, máxima-verosimilitud o métodos bayesianos) para estimar tanto la posición como el efecto de ese posible QTL. Los diseños de cruzamientos son bastante potentes y se conocen fórmulas aproximadas para calcular esta potencia. Por ejemplo para detectar un QTL aditivo que explica

⁴ Flint, A. P. F. and Woolliams, J. A. (2008). Precision animal breeding. *Philosophical Transactions of the Royal Society B*, 363, 573-590.

Figura 2
ESQUEMA DE UN DISEÑO DE CRUZAMIENTO F₂ PARA DETECTAR QTL



un 10% de la varianza de una F₂⁵ entre líneas consanguíneas con una potencia del 90% y un error de tipo I de 0,05 se requieren tan sólo 101 individuos si el marcador está completamente ligado y 281 si está a una distancia de 20 centimorgans. El número si duplicaría si el diseño fuera de retrocruzamiento.

En el segundo diseño se utiliza el desequilibrio de ligamiento dentro de familias aprovechando que en algunas poblaciones como el vacuno de leche hay familias de medios hermanos muy numerosas. Matemáticamente, el análisis de de familias de hermanos es muy similar al de cruces entre poblaciones consanguíneas con una excepción: en el caso de familias estimamos un efecto del QTL para cada familia, y hay tantas estimas de efectos y de posiciones como familias. Esto es así porque no se puede garantizar que el QTL esté segregando en todas las familias y porque un alelo del marcador puede estar asociado a efectos del QTL distintos en cada familia. En cualquier caso la potencia es menor que la de los diseños de cruzamientos de familias. Por ejemplo, incluso utilizando 1.000 individuos repartidos en una estructura defamilias de hermanos óptima la potencia para detectar un gen con efecto α de 0,2 es 0,51 y 0,36 si la varianza residual es 0 y 1 respectivamente. La potencia en el diseño doble-haploide es mayor pero también baja: 0,73 y 0,51 si la varianza residual es 0 y 1 respectivamente.

La mayor parte de los estudios de QTL se han realizado con paneles de 100-300 marcadores microsatélites que cubren todo el genoma (con una distancia media entre marcadores de 5-20 centimorgans). Esta actividad ha sido extraordinariamente exitosa. En la base de datos más conocida (<http://www.animalgenome.org/QTldb/>) el número de QTL descritos son **5,986** que afectan a **581** caracteres en cerdos, **4,281** relacionados con **359** caracteres en vacuno, **284** para **229** caracteres en gallinas y **84** para **30** caracteres en ovino.

⁵ Lynch, M.; Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer, Suntherland.

En los estudios de detección de QTL, lo que localizamos es un QTL en una región de unos 20-40 cm en donde puede haber entre 200-400 genes. Por ello el siguiente paso es tratar de refinar la posición del QTL, tarea que se ha revelado bastante difícil. Hay varias estrategias que pueden seguirse. La primera es aumentar el número de individuos genotipados pero se necesitan unos 5.000 individuos para tener una resolución de 5 cM. La segunda es continuar durante más generaciones construyendo una F3, F4, ... Fn de forma que el intervalo de confianza de la posición del QTL se reduce por un factor de $2/n$ respecto a la F2. Las dos estrategias son caras en tiempo y en dinero.

La tercera estrategia para detectar QTL es la denominada del 'gen candidato': buscar genes conocidos que por razones fisiológicas, esto es por el conocimiento de su función, pudieran ser los genes responsables. Y, a continuación, estudiar si existe relación entre la variabilidad fenotípica y la variación en la secuencia de ADN para este gen candidato. A veces puede utilizarse información de otras especies. Por ejemplo si hemos detectado un QTL para crecimiento podemos mirar si en la zona homóloga en humanos o ratones hay descritos genes relacionados con el crecimiento. Aunque este enfoque ha tenido éxito en algunos casos, por ejemplo el receptor 4 de melanocortina (MC4R) como parcialmente responsable del engrasamiento, crecimiento y consumo de alimentos⁶, se han presentado otros muchos poco reproducibles lo que hace que este enfoque resulte problemático⁷. En el enfoque denominado genética genómica (*genetical genomics*) aprovechamos que, hoy en día, hay herramientas (*microarrays*) que permiten cuantificar el nivel de expresión de un gran número de genes en un gran número de especies domésticas y así identificar los polimorfismos responsables de la variabilidad en los niveles de expresión. El fundamento de este enfoque es que los niveles de expresión de cada gen son fenotipos, pero más 'sencillos' que caracteres clásicos como el crecimiento o la resistencia a enfermedades debido a su 'proximidad' al genotipo. Por tanto, debería ser más fácil identificar los polimorfismos causales en este caso; en una segunda etapa, correlacionaríamos los genes cuya expresión varía con los QTL obtenidos de forma clásica para entender mejor la base genética de los caracteres complejos. Sin embargo, este enfoque tampoco está exento de problemas y escollos⁸.

La cuarta es la que se denomina análisis de asociación en una población comercial, también llamado de desequilibrio de ligamiento. 'Desequilibrio de ligamiento' (LD) es uno de los términos más desafortunados en la literatura genética, ya que es bien sabido que puede existir desequilibrio de ligamiento sin ligamiento y que ligamiento no implica desequilibrio de ligamiento. La justificación de este enfoque es que el desequilibrio de ligamiento entre dos marcadores disminuye continuamente debido a la recombinación, ergo, si encontramos dos marcadores en desequilibrio de ligamiento a nivel de toda la población, deberían estar muy próximos, más próximos cuanto mayor sea el desequilibrio de ligamiento. Aplicando este principio a la búsqueda de QTL, el objetivo sería encontrar marcadores con algún o algunos alelos en fuerte desequilibrio

⁶ Kim, K. S.; Larsen, N; Short, T; Plastow, G. and Rothschild, M. F. (2000). A missense variant of the porcine melanocortin-4 receptor (*MC4R*) gene is associated with fatness, growth, and feed intake traits. *Mammalian Genome*, 1, 131-135.

⁷ Georges, M. (2007). Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Annual Review of Genomics and Human Genetics*, 8, 131-62.

⁸ Haley, C. and de Koning, D. J. (2006). Genetical genomics in livestock: potentials and pitfalls. *Animal Genetics*, 37 Suppl 1, 10-12.

de ligamiento con el carácter de interés. La idea es que el LD disminuye rápidamente conforme aumenta la distancia entre el marcador y el QTL. De todas formas, si bien el desequilibrio de ligamiento es una herramienta que nos puede permitir localizar con precisión las mutaciones causales, también nos puede llevar a falsos positivos. Una de las principales causas de falsos positivos es la estratificación o mezcla de poblaciones. Esta técnica es, por tanto, mucho menos robusta que un análisis de ligamiento.

Sin embargo, recientemente, gracias a los avances tecnológicos en la secuenciación de ADN, los estudios de asociación están dando un salto espectacular en lo que ha dado en llamarse GWA (genomewide association). La consecución de los proyectos genoma en varias especies domésticas ha permitido el descubrimiento de un gran número de SNP como parte de la secuenciación o la re-secuenciación. Aunque, en algunas especies, estamos lejos del millón de SNP de los chips de humanos, existen ya "SNP chips" en vacuno (60.000 SNP y muy pronto de 700.000), perros (250.000 SNP), ovejas (56.000 SNP), cerdos (60.000 SNP), caballos (55.000 SNP) y gallinas (60.000 SNP) que pueden genotiparse fácilmente usando la misma tecnología que en humanos y a un precio razonable (100-200 euros). Al aumentar la densidad de SNP podemos asegurar que para 'todos' los genes del carácter de interés habrá algún SNP en desequilibrio de ligamiento con cada uno de ellos.

La última estrategia, más en el espíritu de Darwin, es buscar huellas de la selección. La idea es que cuando ocurre selección en un gen los genes que están a su lado sufren un efecto de arrastre (lo que suele denominarse de forma muy gráfica, efecto autoestopista). De forma que cuando un gen selecciona, bien por selección artificial o natural deja una huella en la región: menor variabilidad dentro de poblaciones, mayor diferenciación entre poblaciones y mayor desequilibrio de ligamiento. Si encontramos un patrón de este tipo en una región del genoma sabemos que esta zona contiene genes que han sido sometidos a selección y, por tanto, genes de posible valor económico o adaptativo. Es interesante hacer notar que en este enfoque no se requieren medidas fenotípicas para localizar genes que puedan ser interesantes sino solamente análisis de secuencias o SNP de la población estudiada. El problema es que podemos detectar la selección pero no saber con certeza sobre qué carácter ha actuado la selección.

Estas estrategias se están implementando actualmente en las especies domésticas⁹ y, de momento, incluso la más prometedora GWA, solo han tenido éxito para caracteres controlados por uno o unos pocos genes. Por ejemplo, las mutaciones en el gen MITF que causa las manchas blancas en perros o los genes SLC6A5 y ABCA12 que provocan la distonia congénita muscular en vacuno. Sin embargo, para los caracteres complejos, que son los más interesantes desde el punto de vista económico los resultados indican que dependen de muchos genes de pequeño efecto y constituyen el gran reto de cara al futuro.

3.2.2. Detección de los genes responsables de los QTL

Lamentablemente, aunque es fácil encontrar QTL para caracteres de interés económico, localizar los genes responsables es una tarea formidable. Se denomina QTN

⁹ Goddard M. E.; Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10, 381-391.

(*Quantitative Trait nucleotide*) a la mutación responsable del efecto de un QTL. Se han descrito varias historias exitosas⁷: DGAT1 y ABCG2 que afectan a la composición de la leche en vacuno y MSTN que influye en la masa muscular de cerdos y ovejas. El gen DGAT1, que se localiza en el cromosoma 14 codifica para la acil-coenzima A: diacilglicerol aciltransferasa que cataliza la última etapa en la síntesis de triglicéridos e influye en la cantidad y composición de la leche. La mutación en el intrón 3-3072(G-A) en el gen porcino con impronta de porcino IGF2 produce un aumento de la masa muscular y una disminución de la deposición de grasa. La mutación Texel MSTN g+6723(G-A) crea una diana ilegítima para los microRNA miR-1 y miR-206 y como resultado los mRNAs MSTN se inhiben durante la traducción lo que provoca unos niveles más bajos de la proteína MSTN y una mayor masa muscular.

Lo difícil que es encontrar la mutación causal puede ilustrarse con algunos ejemplos. En cerdos de los casi 5000 QTL descritos, apenas se han encontrado las mutaciones causales para una docena de ellos. No deja de ser interesante que el primer QTL descrito en especies domésticas en 1994 fue el FAT1 localizado en el cromosoma 4 de porcino y del que todavía hoy no se conoce la mutación causal. En 1996 se descubrió que variantes en el gen ESR estaban asociadas con el tamaño de camada en cerdos¹⁰ pero diez años después se discute si es el gen responsable¹¹. Por último, el primer QTL que se encontró, en el cromosoma 6, con efecto sobre la producción de leche fue en 1995¹² pero todavía en 2006 se discutía cual de los dos posibles genes OPN (Osteopontina) o la proteína ABCG2 era la mutación causal¹³.

3.3. Selección asistida con marcadores (MAS)

Una de las principales razones para la detección de QTL es poder implementar la selección asistida con marcadores (MAS). Clásicamente se considera que la MAS es un proceso en tres etapas. En la primera se detectan uno o varios QTL relacionados con un carácter productivo. En la segunda se encuentran los genes responsables del carácter. En la tercera se diseña un programa de mejora que aumente la frecuencia de los alelos favorables bien sea por selección o por introgresión.

Si conocemos un gen responsable de un efecto importante sobre un fenotipo de interés podemos utilizar esa información en el programa de mejora. Si el alelo deseable no está en nuestra población pero sí lo está en otra, podemos tratar de introducirlo mediante cruzamientos. Es lo que se denomina introgresión asistida por marcadores (MAI) y es el método habitual de mejora en plantas. En este sentido ha habido iniciativas para introducir alelos chinos en poblaciones europeas (Gene +).

¹⁰ Rothschild, M.; Jacobson, C.; Vaske, D.; Tuggle, C.; Wang, L.; Short, T.; Eckardt, G.; Sasaki, S.; Vincent, A.; McLaren, D.; Southwood, O.; van der Steen, H.; Mileham, A. and Plastow, G. (1996). The estrogen receptor locus is associated with a major gene influencing litter size in pigs. *Proc. Natl. Acad. Sci. USA*, 93, 201-205.

¹¹ Alfonso, L. (2005). Use of meta-analysis to combine candidate gene association studies: application to study the relationship between the *ESR PvuII* polymorphism and sow litter size. *Genetics, Selection, Evolution*, 37, 417-435.

¹² Georges, M.; Nielsen, D.; Mackinnon, M.; Mishra, A.; Okimoto, R.; Pasquino, A. T.; Sargeant, L. S.; Soerensen, A.; Steele, M. R.; Zhao, X.; Womack, J. E. and Hoeschele, I. (1995). Mapping QTL controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*, 139, 907-920.

¹³ de Koning, D.-J. (2006). Conflicting candidates for cattle QTL. *Trends in Genetics*, 22, 301-305.

El procedimiento debe ser lento y cuidadoso ya que hay que hacerlo como complemento a la evaluación genética tradicional para evitar una disminución del mérito para éste carácter (ya que dependerá también de otros genes) o para otros caracteres. Si el alelo favorable está presente en nuestra población trataremos de aumentar su frecuencia (selección asistida por genes conocidos o GAS), pero de nuevo como un criterio más junto con las evaluaciones genéticas habituales. Ya hemos señalado que el número de genes conocidos es relativamente pequeño y discutible y no está muy claro cuántos y en qué medida se están utilizando por la industria, más allá de los aspectos publicitarios. Hay algunos ejemplos como la eliminación de genes nocivos como el halotano (HAL) responsable del síndrome del estrés porcino o el gen napole (RN) en cerdos o la utilización del gen booroola de prolificidad en ovejas o el alelo del FMO3 que causa mal olor en los huevos de cáscara marrón.

La información molecular puede también utilizarse aún cuando los genes responsables no hayan sido localizados, basándose en la información de los marcadores y es lo que típicamente se ha denominado selección asistida con marcadores (MAS) y que, desde el punto de vista industrial, podría ser más relevante. La teoría de la respuesta a la selección con MAS y sus posibles ventajas teóricas están bien establecidas. Los beneficios con MAS serán mayores para caracteres de baja heredabilidad, expresados en un solo sexo (reproducción), de expresión tardía (calidad de carne, longevidad), o de medida costosa (calidad de la carne). Desde el punto de vista táctico también es interesante su empleo en la preselección de animales jóvenes candidatos a pruebas de progenie. En general la gran mayoría de estudios de simulación muestran que la ganancia puede ser del 5-50%. Sin embargo, hasta el momento hay pocos ejemplos donde se esté aplicando MAS en esquemas reales de mejora. Una limitación es, como ya hemos indicado, que se aunque se han detectado muchos QTL, se conocen pocos genes responsables. Además, algunas de estas mutaciones pueden tener efectos deseables e indeseables al mismo tiempo. Varias compañías están desarrollando y comercializando ADN test basados en un pequeño número de marcadores¹⁴ (van Eenennaam *et al.*, 2007). Por ejemplo, en cerdos, los marcadores MC4R, HMGA1 y CCKAR (receptor de la colescistoquina tipo A) que mejoran el crecimiento, el índice de conversión y el porcentaje de la canal, el marcador PRKAG3 asociado con el contenido en glicógeno del músculo y la calidad de la carne. El marcador CAST (Calpastina) responsable de la inhibición de enzimas proteasas y, por tanto, relacionadas con la ternura de la carne. El gen ESR, ya mencionado, asociado con el tamaño de camada y el EPOR asociado con la capacidad uterina y el tamaño de camada.

El impacto de la selección asistida por marcadores ha sido modesto, porque los QTL que exceden el nivel de significación solo explican una fracción pequeña de la varianza. A título de ejemplo señalamos los comentarios de varios autores, todos ellos asesores de empresas de mejora¹⁵: 'Despite of the great enthusiasm for breeding companies to be involved there are very few applications of MAS in commercial poultry breeding. They are not convinced about economic feasibility of MAS' (de Koning y Hocking, 2007); 'Although several useful genes (primarily gene-linked markers) have been identified in pigs, their application has been limited and their success inconsistent' (Spotter y Distl, 2006); 'The much anticipated benefits of

¹⁴ Van Eenennaam, A. L.; Li, J.; Thallman, R. M.; Quaas, R. L.; Dikeman, M. E.; Gill, C. A.; Franke, D. E.; Thomas, M. G. (2007). Validation of commercial ADN tests for quantitative beef quality traits. *Journal of Animal Science*, 85, 891-900.

¹⁵ Toro, M. A. (2010). Future trends in Animal Breeding due to new genetic technologies (in press).

ADN-based tools to routinely guide selection decisions in cattle have not been fully met since the origin of this premise' (Sonstegard y van Tassell, 2004); 'Although initial expectations for the use of marker assisted selection were high the current attitude is one of cautious optimism' (Dekkers, 2004); 'Unfortunately, as is too often the case, the immediate promise of genomics was clearly oversold, as it has since become clear that the identification of QTL was only a first baby step in the process to bringing these results to a practicable technology' (Green, 2009).

3.4. Una propuesta más radical: la selección genómica

En 2001 Meuwissen *et al.*,¹⁶ Propusieron un enfoque diferente de la MAS, que denominaron selección genómica. Se parte de dos supuestos que solo recientemente parecen haberse alcanzado. El primero es que paneles de varios cientos de miles de marcadores estén disponibles a un coste razonable. Y el segundo que cubran densamente todo el genoma de forma que todos y cada uno de los genes responsables del carácter estén en desequilibrio de ligamiento con al menos un marcador. Ambos supuestos son ahora mismo una realidad ya que como la consecución de los proyectos genoma en varias especies domésticas ha permitido el descubrimiento de un gran número de SNP como parte de la secuenciación o la re-secuenciación. Aunque en cierto sentido la selección genómica se relaciona con el concepto GWA explicado antes existe una diferencia de enfoque. En la GWA el objetivo es encontrar los genes responsables de un carácter de interés. En la mejora genética lo importante es la predicción de los valores genéticos para elegir a los reproductores más idóneos. En el primer caso buscamos conocimiento, en el segundo predicciones adecuadas. Estas predicciones adecuadas se basan en la información de los marcadores, consideradas como variable explicativas en el sentido estadístico, estén o no relacionadas de forma causal con los verdaderos alelos que controlan el carácter.

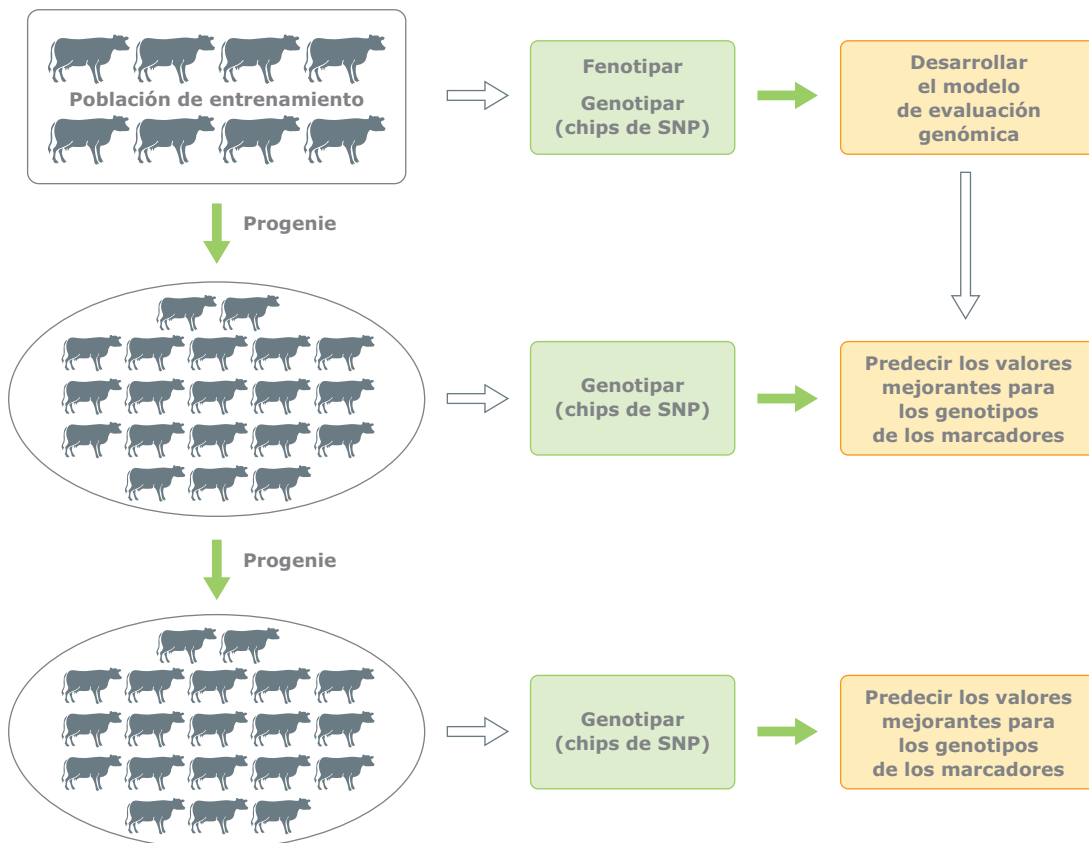
En su versión más simple (Figura 2) la selección genómica es un proceso en dos etapas. En la primera se estiman los efectos de los marcadores (>50,000) en una población de referencia (*training population*) a la que hemos genotipado y en la que hemos medido los caracteres de interés. En la segunda etapa utilizamos esta información para predecir los valores mejorantes en una población de referencia (*testing population*) que únicamente se ha genotipado para la misma batería de marcadores.

La principal diferencia entre la selección genómica y la MAS es que esta última se concentra en unos pocos QTL en los que se ha demostrado una asociación con los caracteres de interés mientras que la selección genómica utiliza un panel de marcadores de alta densidad de forma que todos los QTL estén en desequilibrio de ligamiento por lo menos con un marcador.

La selección genómica ha despertado mucho interés y algunas empresas están rediseñando sus programas de mejora. La idea es que, con la selección genómica podemos predecir los valores mejorantes de los candidatos a la selección al nacimiento con una precisión de 0.8. De esta forma podemos seleccionar animales a una edad temprana y así duplicar la tasa de progreso genético por año.

¹⁶ Meuwissen, T. H. E.; Hayes, B. J.; Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 1819-1829.

Figura 3
ETAPAS EN LA SELECCIÓN GENÓMICA



En vacuno de leche es donde quizás el impacto de la selección genómica ha sido más inmediato. Desde hace más de 50 años la mejora genética del vacuno de leche se basa en las denominadas pruebas de progenie. Los toros jóvenes son apareados con vacas utilizando la inseminación artificial para obtener unas 100 hijas de cada uno de ellos. Estas hijas son evaluadas para la producción de leche lo que permite la predicción del mérito genético de sus padres con una precisión del 90%. Los mejores machos "probados" (aproximadamente un 10% de ellos) son los que ahora utilizaremos de forma masiva en la población general. Este proceso de mejora, aunque eficiente, es muy costoso tanto en tiempo (unos 5 años) como en dinero. En cambio, el diseño óptimo de un programa de mejora basado en la selección genómica sería más o menos como sigue: a) genotipar un gran número de toros jóvenes de la población; b) calcular los valores mejorantes genómicos de estos toros (precisión = 0.8); c) seleccionar un grupo de estos toros en función de su valores mejorantes y vender semen de ellos tan pronto como lo producen. El intervalo generacional se reducirá a unos 2 años aproximadamente con lo que la tasa de progreso genético se duplicará.

La selección genómica en vacuno de leche es ya un hecho. El primer catálogo oficial de predicciones genómicas de toros jóvenes se presentó en enero de 2009. Las dudas que existían respecto a que los ganaderos serían reacios a utilizar semen de toros sin prueba de progenie, esto es sólo evaluados genómicamente, se ha revelado injustificada. Al parecer las ventas de semen de estos toros se ha cuadruplicado frente a los de toros probados. De todas formas, aunque la tecnología se ha aceptado con prontitud, queda por comprobar cuál será el resultado a largo plazo tanto sobre la mejora genética como la re-organización del sector. Una consecuencia no deseable será, probablemente, que debido a la alta intensidad de selección el mantenimiento de la variabilidad genética puede verse comprometido así como la posibilidad de que puedan introducirse en los rebaños de genes que puedan ser favorables para la producción de leche pero negativos para otros caracteres como, por ejemplo, la fertilidad.

En el caso del vacuno de carne la situación es diferente. Una de las razones es que el número de razas involucrada en el sector es mucho mayor y podría ocurrir que un alelo que es un buen predictor del mérito genético en una raza no lo fuera en otra. Otra es que los objetivos de la mejora genética en vacuno de leche (cantidad y composición de la leche) son más sencillos, que en vacuno de carne: composición de la canal, terneza, espesor de grasa, marmoleado... Por último la inseminación está mucho menos extendida que en vacuno de leche.

En la mejora clásica de gallinas de puesta los machos se seleccionan a una edad temprana en función de la información proporcionada por sus hermanas para los caracteres de interés, producción y calidad de los huevos. Sin embargo, todos los hermanos machos tienen la misma evaluación genética por lo que si todos ellos fueran seleccionados la consanguinidad se dispararía. La selección genómica permitiría elegir a los mejores machos, dentro de cada grupo de hermanos, en función de la información proporcionada por los marcadores. Se ha calculado que la aplicación de la selección genómica en gallinas, aumentando la precisión de la estimación en gallinas de carne y acortando además el intervalo generacional en gallinas de puesta podría aumentar el progreso genético en 20-30% (Avendaño *et al.*, 2009). Sin embargo, hay dudas sobre si este progreso sería competitivo en relación a su coste. A nivel más experimental algunas empresas están genotipando con los chip de alta densidad animales resistentes y susceptibles a enfermedades o desórdenes metabólicos con objeto de analizar la base genética de estos caracteres.

En definitiva, la selección genómica es la herramienta de mayor impacto sobre la eficiencia de la selección animal que el mejoramiento genético ha producido en muchos años, particularmente para el bovino de leche. Esto quedó reflejado en el último "World Congress on Genetics Applied to Livestock Production" celebrado en Leipzig en agosto del 2010, de 846 comunicaciones presentadas, nada menos que 276 eran sobre selección genómica. Existen varias incógnitas relacionadas con la implementación de la selección genómica que irán clarificando en los próximos años. En primer lugar, respecto a temas de diseño como la elección de la población de referencia: una o varias razas, número de animales y número de marcadores a genotipar y con qué frecuencia se han de re-estimar los efectos de los marcadores

o introducir nuevos marcadores. La secuencia genómica completa de un animal a un precio razonable será probablemente factible en poco tiempo, pero esto ¿aumentará la precisión o introducirá más ruido? En segundo lugar, respecto a los modelos y enfoques estadísticos. Aunque los modelos bayesianos propuestos inicialmente (Meuwissen *et al.*, 2001) son, de momento, los más aceptados, se están investigando otras alternativas de tipo no-paramétricos tales como métodos de *machine learning*, redes neuronales o redes bayesianas^{17,18}, (Gianola *et al.*, 2006; Gianola and de los Campos, 2008; González-Recio *et al.*, 2008). En tercer lugar, en algunas especies, como cabras y conejos, todavía no se ha secuenciado el genoma y por tanto no existen chips de SNP disponibles. En cuarto lugar, quizás la limitación más importante es la falta de medidas fenotípicas para caracteres difíciles de medir tales como la resistencia a enfermedades y al estrés, la fertilidad, la longevidad y la utilización de nutrientes. Es más, cabe pensar que si los ganaderos piensan que elegir a los animales por su valor genómico es suficiente, tienen menos incentivos para medir los caracteres fenotípicos, aunque en última instancia son imprescindibles para conocer el efecto de los marcadores. Finalmente, lo que nos gustaría conocer es la biología de las interacciones entre los genes y de estos con el ambiente para producir los fenotipos complejos.

3.5. Ciencia ficción: el uso combinado de la selección genómica y la manipulación de los gametos y embriones

Meuwissen *et al.*,⁶ también señalaron que en el futuro la tasa de progreso genético podría aumentar todavía más si la selección genómica se combinara con las tecnologías de manipulación de gametos y embriones. Estas últimas permitirían reducir todavía más los intervalos generacionales y aumentar el número de ciclos de selección por unidad de tiempo. En la denominada *velogenetics* los oocitos se recolectarían del útero de las terneras, se madurarían *in vitro*, se fertilizarían y los embriones se seleccionarían genómicamente. Posteriormente, se implantarían en vacas receptoras. El proceso podría repetirse recolectando oocitos de esta segunda generación con lo que el intervalo generacional se reduciría a unos 3-6 meses. En la propuesta denominada *whizzogenetics*, los embriones, en vez de re-implantarse se cultivarían *in vitro*, se induciría la meiosis y se formaría una nueva generación de embriones. Así podría practicarse varias generaciones de selección genómica en el laboratorio con un intervalo generacional de unos quince días.

¹⁷ Gianola, D.; de los Campos G. (2008). Inferring genetic values for quantitative traits non-parametrically. *Genetics Research*, 90, 525-540.

¹⁸ González-Recio, O.; Gianola, D.; Long, N.; Weigel, K. A.; Rosa, G. J. M. and Avendaño, S. (2008). Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in broilers. *Genetics*, 178, 2305-2313.

4. Aplicación de la selección genética y genómica en agricultura

4.1. Características específicas de las plantas y sus genomas en relación con su mejora genética

Las plantas constituyen un conjunto diferenciado de especies con características biológicas propias, muchas de las cuales determinan el modo en que pueden estudiarse a nivel genético y manipularse para obtener variedades mejoradas, ya que:

- a) Habitualmente es posible obtener un elevado número de descendientes por cruzamiento.
- b) Muchas especies pueden autofecundarse y generar líneas puras que reproducen el mismo genotipo y fenotipo.
- c) La reproducción vegetativa o clonal es posible en muchos casos, algunas veces con gran facilidad, por esquejes, injertos u otros propágulos, lo que permite la evaluación de individuos parcialmente heterocigóticos en ensayos replicados. Muchas especies pueden propagarse o regenerarse a partir de protoplastos o de diversos tejidos por procedimientos *in vitro*.
- d) Las plantas admiten frecuentemente la hibridación interespecífica, la poliploidía y la haploidía, lo que las convierte en organismos particularmente flexibles respecto a su contenido genético lo que permite acceder a fuentes o modos de variabilidad de las que otros organismos carecen.

También las plantas representan un conjunto muy heterogéneo de organismos vivos que difieren en aspectos esenciales de su biología. Uno de ellos es el sistema reproductivo dentro del que se distinguen las especies de reproducción sexual, entre las que se encuentran las típicamente autógamas, como el trigo, la cebada, el arroz, la colza o la soja y las alógamas, como el maíz, la cebolla, la mayoría de los árboles frutales, las especies de reproducción vegetativa como la fresa y muchas ornamentales, y las apomícticas, como el ajo. Muchas especies combinan de manera natural o artificial diversos de estos modos de reproducción, como muchos de los cultivos frutales y ornamentales que suelen reproducirse sexualmente aunque como cultivos se propaguen por la vía vegetativa o varios cítricos que son naturalmente apomícticos pero se suelen propagar vegetativamente.

Otro aspecto diferencial entre diversas especies de plantas es su contenido en ADN. Un genoma pequeño, como el del melocotonero (230 Mbp), es 100 veces menor que el del puerro (24.000 Mbp). Estas diferencias, sin embargo, no son tan grandes en lo que respecta al número de genes, que raramente superan el doble de los ~27.000 estimados para *Arabidopsis* independientemente del tamaño total del genoma (Sterck *et al.*, 2010). El tamaño del genoma sí es en cambio un problema para su secuenciación, que aumenta en complejidad y precio en los genomas mayores. Es por ello que algunos de los cultivos de mayor importancia económica, como el trigo o la cebada, no han sido secuenciados todavía, aunque dado el ritmo de caída del coste de secuenciación de ADN es previsible que todos los genomas del centenar de especies de mayor importancia en agricultura lo sean a medio plazo.

4.2. Aplicaciones de los marcadores moleculares a la selección y mejora genética de plantas basadas en un número bajo de marcadores

La era del uso de marcadores moleculares se inicia en plantas en el último cuarto del siglo pasado con las isoenzimas. A pesar de sus limitaciones en cuanto a número, con ellas se pusieron las bases de la aplicación de los marcadores a la genética y mejora de plantas (Tanksley y Orton, 1983). En el transcurso de los años siguientes, emergieron nuevos marcadores a partir de la mejora de las tecnologías de manipulación del ADN, como los métodos de extracción, el uso de enzimas de restricción, el descubrimiento de la reacción en cadena de la polimerasa y más recientemente el abaratamiento de la secuenciación, particularmente con las llamadas tecnologías NGS (*Next-Generation Sequencing*), que actualmente permiten disponer de juegos de marcadores en números suficientemente grandes, buena calidad y precio razonable como para que sean empleados por prácticamente todas las empresas de mejora de plantas.

Entre los marcadores más usados destacan los basados en microsatélites (o SSR por *Simple-Sequence Repeats*) y los SNP (*Single Nucleotide Polymorphisms*). El hecho de que cada vez sea mayor el número de genomas secuenciados favorece el uso de ambos tipos de marcadores. Aunque los SSR se emplean profusamente en el ámbito aplicado, los SNP van ganando progresivamente terreno dado que existen métodos muy diversos y versátiles para detectarlos, muchos sin necesidad de separación electroforética, que su coste es previsible que siga descendiendo y que pueden basarse en la misma secuencia del gen (o incluso de la mutación) causante de la diferencia fenotípica a seleccionar. Otros marcadores no específicos de secuencia, como AFLP, RAPD o similares son usados para aplicaciones muy puntuales y cada vez más raramente.

4.2.1. Identificación de genotipos y control de calidad

Entre las primeras aplicaciones de los marcadores a los programas de mejora de plantas destacan las relacionadas con el control de calidad de las semillas o plantas comercializadas por las empresas. Una de ellas es la determinación de pureza en semilla de híbridos F1. Este tipo varietal, generalizado en muchas especies como el maíz, la remolacha y la mayoría de las hortalizas, exige asegurar que la semilla no híbrida, habitualmente debida a autofecundaciones de la línea usada como parental femenino, esté por debajo de determinado umbral. El riesgo de lotes de semillas no conformes se agudiza con la globalización de la producción de semilla híbrida, ya que muchas empresas contratan a terceros de todo el mundo la producción de una parte de la semilla que comercializan por cuestiones de precio y disponibilidad durante todo el año. Un test de pureza híbrida puede realizarse fácilmente con el uso de un solo marcador homocigoto para alelos distintos en cada uno de los parentales del híbrido. Se toma para ello una muestra de semillas de cada lote, se determina el genotipo de cada una de ellas y se calcula la proporción de semillas que no tienen el genotipo heterocigoto esperado. La rapidez y eficiencia de este test es un elemento clave desde hace ya muchos años para asegurar la calidad de la semilla híbrida y las empresas dedican una parte considerable de su esfuerzo en marcadores a esta labor. El mismo principio se puede aplicar de una manera general al control de identidad de lotes de semilla (híbrida o no) o de planta (en especies de re-

producción vegetativa), para evitar poner en el mercado lotes erróneos o mezclas de plantas o semillas. En este caso el test se hace con un conjunto de marcadores suficiente (5-10 SSR o 10-30 SNP) para poder separar todas las variedades con las que exista riesgo de mezcla o confusión.

Otra aplicación importante en este ámbito se refiere al uso de los marcadores para la protección de derechos de obtentor. La reproducción y venta de semilla y planta de una variedad sin autorización de su propietario legal, el obtentor que ha protegido dicha variedad, está penalizada por la legislación de muchos países. La situación es particularmente grave en variedades línea pura, de polinización abierta o de reproducción vegetativa que no tienen un sistema de autoprotección como los híbridos F1. La demostración de si dos muestras problema pertenecen o no a la misma variedad puede ser larga, difícil y, a veces, sin una conclusión suficientemente rotunda cuando se basa en caracteres morfológicos. En especies arbóreas como los frutales o la vid, el largo período necesario para tales ensayos puede resultar en la imposibilidad real de la defensa de los derechos de los mejoradores. Una solución es el uso de marcadores moleculares para identificar en días o incluso horas cada variedad por su genotipo para un juego de SSR o SNP. España es pionera en el uso de tales técnicas que han sido aplicadas especialmente en frutales de hueso y de una manera particular en melocotonero y nectarina, donde la producción es elevada y el fraude importante. La acción conjunta entre la empresa GESLIVE AIE, la organización de los mejoradores que operan en España para la gestión de las licencias varietales, y el servicio de marcadores moleculares IRTAGen del IRTA ha permitido reducir considerablemente el fraude, utilizando los marcadores como elemento clave para el test de identidad varietal. Estas pruebas han sido aceptadas por los tribunales españoles como prueba relevante en los juicios realizados. Actualmente existe experiencia y capacidad de identificación GESLIVE-IRTAGen para 14 cultivos, la mayoría frutales y ornamentales, aunque también en algunos cereales y leguminosas.

4.2.2. Selección asistida por marcadores (MAS)

En su acepción más sencilla, la MAS consiste en utilizar el genotipo de un marcador en las proximidades de un gen mayor que determina un carácter de interés para seleccionar este carácter en lugar de hacerlo usando el fenotipo. En plantas es frecuente que haya genes mayores interesantes como las resistencias a enfermedades u otros caracteres que a veces provienen de líneas exóticas o de especies silvestres intercompatibles con la cultivada. Con la MAS se trata de seleccionar antes, con mayor precisión o a menor precio que los requeridos cuando se usa el fenotipo. Uno de los ejemplos pioneros de esta aplicación fue el uso de un gen de una fosfatasa ácida del tomate (Aps-1) para seleccionar para el gen Mi de resistencia a nematodos procedente de una especie silvestre relacionada (Medina-Filho, 1980).

Para poder realizar la MAS es preciso primero disponer de al menos un marcador en las proximidades del gen a seleccionar, lo que puede conseguirse de diversas maneras. Dos de las más habituales se describen a continuación:

- a) *Mapeo en descendencias* tipo F2, BC1, o similares que segreguen para el gen de interés. Con ellas se construye un mapa de marcadores y se estudia la co-segregación de estos con el carácter de interés. La localización del gen mayor en el mapa supondrá también la identificación de marcadores en su proximidad. Pues-

to que se suele usar un número limitado de marcadores para construir estos mapas, habitualmente seleccionados de mapas de referencia para asegurar que el genoma quede totalmente cubierto a distancias parecidas y grandes (15-25 cm por marcador), el gen estudiado suele estar alejado del marcador más próximo (para la MAS es deseable que la distancia entre marcador y gen sea ≤ 1 cm). Por ello es habitualmente necesaria una segunda ronda de selección de marcadores que puede hacerse a partir de otros transferibles ya mapeados en mapas de referencia. Si se dispone de un mapa físico o de la secuencia del genoma en la especie de interés o en otra muy próxima, se puede usar esta información para buscar otros marcadores más cercanos que puedan emplearse en MAS.

- b) *Uso de mezclas de ADN* en poblaciones segregantes. Este método descrito por Michelmore *et al.* (1991) como *bulked segregant analysis* (BSA) consiste en obtener una población (F₂, BC₁, etc.) que segregue para el gen de interés, fenotiparla y, a continuación, seleccionar dos grupos de plantas (~6-14 individuos por grupo) con fenotipos alternativos (p. ej., uno de plantas resistentes y otro de susceptibles a una enfermedad), extraer su ADN y mezclar el de los individuos con el mismo fenotipo. Las dos mezclas de ADN se usan entonces para obtener el mayor número posible de marcadores moleculares (los AFLP y RAPD son muy usados para este propósito) en busca de diferencias entre los patrones de bandas de las dos mezclas. Estas diferencias pueden corresponder a fragmentos de ADN próximos al gen de interés. Se aíslan entonces estos fragmentos y se elaboran marcadores polimórficos cuya segregación se estudia en la población. Aquellos suficientemente próximos al gen de interés se usan en MAS.

4.2.3. Selección de todo el genoma en programas de retrocruzamiento

Uno de los métodos más utilizados en mejora vegetal es el del retrocruzamiento cuyo objetivo es incorporar (introgresar) en una línea élite un gen que se encuentra en una variedad exótica o en una especie próxima. Para ello se cruzan las dos líneas y se utiliza la línea élite como parental recurrente, es decir, para retrocruzar en generaciones sucesivas. En cada generación de retrocruzamiento (se suele llegar hasta la seis u ocho) se selecciona a favor de la presencia del gen a introgresar y de las plantas que tengan mayor parecido parental recurrente. Al final del proceso se llega a una línea idealmente idéntica al parental recurrente con la excepción de un fragmento de ADN de la línea donante que contiene el gen que hemos introgresado.

Para acortar este proceso se usa un conjunto de marcadores que cubran el genoma de la especie estudiada a distancias relativamente grandes (20-30 cM) y se seleccionan aquellos individuos de la primera generación de retrocruzamiento (BC₁) que tengan el mayor número posible de marcadores homocigotos como en el parental recurrente, con la excepción del gen de interés que debe seleccionarse en heterocigosis (usando un marcador muy próximo o seleccionando previamente los individuos que muestren el carácter, si se trata de un alelo dominante). En generaciones posteriores se completa la selección en homocigosis de los marcadores aún heterocigotos. El proceso normalmente termina en la BC₂ o BC₃, lo que supone un ahorro de 3-6 generaciones con respecto al procedimiento convencional. Esta es una de las aplicaciones más usadas por las empresas, particularmente para las especies que se comercializan como líneas puras o híbridos F₁.

4.3. Nuevas aplicaciones de los marcadores basadas en la cobertura completa del genoma

El desarrollo de las tecnologías de NGS ha sido una de las causas del abaratamiento de la secuenciación de ADN que ha ocurrido en los últimos años y que previsiblemente continuará en el futuro (ver Apartado 2). Una de las consecuencias de este avance es la posibilidad de la secuenciación *de novo* del genoma de muchos cultivos para los que hasta ahora esto era económicamente prohibitivo. Después de *Arabidopsis* el año 2000 han seguido, o seguirán en los próximos meses, especies como el arroz, *Populus*, papaya, vid, *Sorghum bicolor*, pepino, patata, ricino, manzano, maíz, melón, fresa diploide, melocotonero y tomate, la mayoría obtenidas en el período 2008-2010, a las que se unirán en un futuro muy próximo las de cualquier especie que tenga una mínima importancia económica, incluidas las que tienen grandes genomas.

La búsqueda de genes y marcadores para su selección en una especie determinada se simplifica cuando se puede acceder a su genoma y, aunque sigue siendo necesario establecer la posición del gen/QTL de interés por la vía, que podríamos llamar convencional, del análisis genético con mapas de marcadores, la operación de desarrollar mapas de alta precisión y detectar posibles genes candidatos o marcadores estrechamente ligados se limita esencialmente a una búsqueda *in silico* que debe después ser refrendada en poblaciones segregantes.

4.3.1. Uso del desequilibrio de ligamiento: genética de asociación

Un elemento crucial de las tecnologías NGS ha sido su capacidad de generar una gran cantidad de datos sobre variabilidad intraespecífica que han permitido descubrir un gran número de SNP en muchas especies. Cuando la secuencia del genoma está disponible en la especie estudiada, estos marcadores se pueden mapear con muy poco esfuerzo. Estos SNP son la base de chips de gran resolución (decenas de miles de marcadores por ensayo) con los que se puede acceder a los bloques cromosómicos conservados entre individuos separados entre sí muchas generaciones y, por consiguiente, a hacer uso del desequilibrio de ligamiento (DL) como elemento de análisis genético (Rafalski, 2010). Con ello es posible un análisis genético de una amplitud y precisión muy superiores a las que pueden obtenerse con el análisis convencional basado en el estudio de segregaciones biparentales. El análisis conjunto de un gran número de marcadores en una colección de individuos procedentes de colecciones de germoplasma y, por tanto, de pedigrí desconocido o incompleto que han sido fenotipados para diversos caracteres permite identificar asociaciones estadísticas entre ciertos marcadores o haplotipos y estos caracteres. A esto se llama mapeo por asociación o por DL. Las ventajas del mapeo por asociación son: primera, que en una población de 100-200 individuos de una especie hay una muestra mucho mayor de la variabilidad genética para cada carácter, tanto en número de QTL como de alelos de estos QTL, lo que permite un análisis mucho más exhaustivo que en una descendencia biparental tipo F₂, donde hay solamente dos alelos por locus segregante; y segunda, que el mapeo es más preciso ya que la probabilidad de que dos *loci* ligados haya recombinado es mayor, al haber ocurrido un número de meiosis más elevado, lo que significa que solamente los marcadores muy cercanos a los genes de interés, incluyendo los que ocurren dentro de estos genes, estarán asociados a ellos.

La precisión del mapeo será mayor cuanto menor sea el nivel de conservación del DL en la población analizada, lo que depende de la historia de cada especie. La conservación del DL decae con la distancia genética, el número de generaciones desde un progenitor común y el nivel de recombinación del genoma o zona genómica estudiados. Por el contrario, el DL se conserva en especies autógamas (la homocigosis reduce la eficacia de la recombinación) y en especies que han sufrido importantes cuellos de botella poblacionales en su historia reciente. En plantas existen casos extremos de especies de baja conservación del DL, particularmente en alógamas como el maíz (Yu y Buckler, 2006) donde el DL desaparece en pocas kbp, hasta especies como el melocotonero (autógama y con un reciente evento de deriva genética al inicio de la mejora moderna) en la que este puede conservarse durante varias Mbp (Aranzana *et al.*, 2010).

Cuando la conservación del DL es baja, es posible con el uso de una elevada densidad de marcadores en el conjunto del genoma o en la región de interés, localizar con gran precisión la posición del gen de interés o verificar si la variabilidad de la secuencia de un determinado gen candidato correlaciona con el carácter al que supuestamente afecta. Cuando el DL es elevado, es posible hacer rastreos del genoma completo (*whole-genome scans*) para genes o QTL de interés con un número mucho menor de marcadores, aunque la precisión del mapeo de cada gen/QTL sea menor. Puesto que diferentes colecciones de la misma especie pueden en función de su historia tener diferentes niveles de DL, puede escogerse la población y el ensayo de marcadores en función del objetivo: en cebada, Caldwell *et al.*, 2006 observaron que el DL era muy superior en variedades comerciales que en colecciones de materiales ancestrales, lo que permite esta estrategia a dos niveles en los que la primera colección sirve para localizar la posición del gen/QTL de interés y la segunda permite proceder a su mapeo fino o eventual clonación.

La existencia de estructura poblacional en la especie estudiada, favorecida por la deriva genética y la selección direccional, genera asociaciones espurias entre marcadores y caracteres no ligados, lo que complica el análisis de asociación. Es preciso conocerla de antemano a partir de análisis de variabilidad previos y corregirla tanto en la selección de la colección de genotipos a analizar como en el método de análisis de asociación posterior. Otro elemento limitante es la presencia de alelos raros, que difícilmente serán detectados si su efecto no es muy grande. Para analizar estos alelos el mapeo con poblaciones biparentales sigue siendo imprescindible.

4.3.2. Generación de poblaciones especialmente adaptadas al análisis de asociación

Dada la importancia y precisión del análisis de asociación y la versatilidad de las plantas para poder crear poblaciones adecuadas para estudios genéticos, se han desarrollado diversas colecciones experimentales de plantas que superan las limitaciones de las descendencias biparentales y las colecciones de genotipos, permitiendo un análisis genético de todo tipo de caracteres de un modo más eficiente y riguroso.

En primer lugar, es posible mejorar la resolución de mapeo de una población biparental por el procedimiento de forzar una serie de generaciones de polinización cruzada a partir de la F2 (Darvasi y Soller, 1995), seguidas a veces por autofecundaciones hasta obtener líneas recombinantes consanguíneas (RIL). Con ello se aumenta el número de meiosis y por tanto de sobrecruzamientos en la población a fenotipar, disminuyendo el DL y permitiendo la localización de la posición de marcadores, genes o QTL con mayor precisión.

Una versión más elaborada, que implica la creación de poblaciones con múltiples parentales con el objeto de acceder a una variabilidad mayor, es la llamada MAGIC (*Multiparent Advanced Generation Intercross*) que fue propuesta inicialmente en ratones, y recientemente ha empezado a utilizarse en plantas, donde Kover *et al.* (2009) desarrollaron una población en *Arabidopsis*, basada en 527 RIL obtenidas del intercrucamiento entre 19 fundadores que mostró su eficacia en el mapeo fino de QTL de varios caracteres de desarrollo. A partir de estos conceptos, que tratan de complementar el mapeo de QTL convencional y de asociación, McMullen *et al.* (2009) desarrollaron un nuevo concepto de mapeo basado en la población llamada *Nested Association Mapping* (NAM), que consiste una colección de RIL obtenidas a partir de individuos F1 entre 25 líneas puras de maíz de origen diverso cruzadas con una línea pura de referencia, B73, a razón de unas 200 RIL por familia (5.000 en total). En esta población se ha elaborado un mapa con 1.106 marcadores que ha permitido detectar un total de 136.000 sobrecruzamientos, que supondrían de promedio tres sobrecruzamientos por gen en el genoma del maíz. La población NAM permite, por tanto, el análisis genético de caracteres de herencia compleja con una inusitada precisión, facilitando la localización de la mayor parte de los genes implicados en su expresión, así como de sus alelos, sus interacciones con otros genes y el ambiente, así como una vía hacia la identificación de las secuencias responsables de la variabilidad encontrada. La población NAM ha sido ya usada para el análisis de la época de floración en maíz (Buckler *et al.*, 2010), detectando un elevado número de QTL distribuidos por todo el genoma, con efectos menores cada uno de ellos, varios alelos por *locus* procedentes de diferentes líneas, modo de acción esencialmente aditivo con poca epistasia e interacción con el medio. Esta población es un recurso de enorme valor para el objetivo inicial de la aplicación a la mejora, que es la identificación y caracterización de todos los puntos del genoma donde existen genes que afectan a un carácter determinado. Parece probable, a pesar de que el coste de creación y mantenimiento de este tipo de poblaciones no es desdeñable, que se creen otras parecidas en otras especies con el objetivo de realizar análisis genéticos en profundidad de caracteres hasta ahora difíciles de tratar con las aproximaciones convencionales.

4.3.3. Nuevas estrategias de mejora especialmente adaptadas a la selección de caracteres de herencia compleja

La disponibilidad de juegos de marcadores con cobertura de todo el genoma a unos costes por marcador bajos empieza a generar métodos de mejora que utilizan esta información. Si la metodología de selección de todo el genoma descrita anteriormente (apartado 4.2.3) era posible con una cobertura del genoma de muy baja densidad y permitía en esencia la recuperación del fondo genético de una línea con la selección de un solo gen o QTL, ahora es posible plantearse esta misma estrategia seleccionando para varios caracteres al mismo tiempo, para caracteres determinados por varios *loci* o para ambas situaciones. Dos de las nuevas metodologías propuestas, particularmente para la selección simultánea de diversos QTL, son:

- *MARS (Marker-Assisted Recurrent Selection)*: En esencia consiste en aumentar la frecuencia de alelos favorables en uno o varios caracteres de herencia compleja a base de seleccionar marcadores asociados a los QTL que los determinan. Aunque los esquemas de mejora usados pueden ser muy diversos, el método parte del genotipado y fenotipado de una población segregante con un número de marcadores que garanticen la cobertura de todo el genoma. Los datos obtenidos se

usan para identificar un conjunto más pequeño de marcadores estrechamente asociados a QTL de efectos importantes. Estos marcadores permiten la posterior selección de los individuos con mayor frecuencia de los alelos favorables de los QTL, que se intercrucan nuevamente para obtener una siguiente generación que estará enriquecida en los alelos que nos interesa seleccionar, repitiéndose este ciclo las veces que se considere adecuado (Bernardo, 2008; Ribaut *et al.*, 2010). Una ventaja de este procedimiento es que una vez identificado el juego de marcadores relevante, lo que se hace en el medio donde la nueva variedad quiere implantarse, pueden realizarse varios ciclos de selección, únicamente basados en el genotipo, en medios que permitan acortar el período intergeneracional. De este modo se produce un avance más rápido y efectivo hasta que en las últimas fases de selección las descendencias segregantes se vuelven a evaluar en las condiciones ambientales para las que se quería mejorar.

- *Selección genómica o GWS (Genome-wide Selection)*: La cobertura de todo el genoma con marcadores que aseguren el seguimiento de todos los bloques de ligamiento es posible cada vez en un mayor número de especies. Por lo tanto pueden encontrarse marcadores asociados a cada uno de los QTL de un carácter o del conjunto de caracteres interesantes en un cultivo. El genotipo para estos marcadores de cada individuo permitiría estimar tanto su valor fenotípico como su valor de mejora cuando fuera usado como parental en el proceso de obtención de una nueva variedad. Estos conceptos que son la base de la selección genómica (Goddard y Hayes, 2007) ya usada en mejora animal, tienen similares aplicaciones en plantas, aunque actualmente su uso está en una fase muy preliminar. Una primera aproximación, consiste en seguir un esquema igual que en MARS, pero utilizando todos los marcadores en cada ciclo sin ningún análisis previo de QTL (Bernardo, 2008). El valor de cada individuo se estima usando el conjunto la información generada por los marcadores con BLUP (Best Linear Unbiased Prediction) o con estimadores Bayesianos. Los individuos con mejores valores genotípicos serían seleccionados como parentales para la siguiente generación. De este modo se seleccionarían, además de los QTL mayores (los que se seleccionan en MARS), otros menores que pueden permitir un aprovechamiento más exhaustivo de la variabilidad existente para el carácter o caracteres de interés. Tanto MARS como GWS, pueden representar un gran avance en especies con largos períodos intergeneracionales, como muchas especies arbóreas, al permitir el avance de generaciones basado únicamente en la selección genotípica, sin tener que esperar una evaluación fenotípica que puede llevar largos años, así como la estima del valor de mejora de los fundadores de los programas de mejora.

4.4. Casos prácticos: estado actual de la selección genética y genómica en varias especies

A continuación se resume la situación actual de la aplicación de las tecnologías de marcadores moleculares en varias especies cultivadas que representan un amplio rango de la casuística posible en cuanto a diversidad genómica, nivel de conocimientos, características vegetativas y reproductivas. En todas ellas se usan extensivamente los marcadores como herramientas para identificación varietal con objetivos de control de calidad o defensa de los derechos de obtentor, además de las aplicaciones relacionadas con la selección que se citan en mayor detalle a continuación:

- **Trigo:** Una de las fuentes básicas de la alimentación humana. Planta herbácea, anual y autógama, originaria del Creciente Fértil, con un genoma hexaploide ($2n=6x=42$) de gran tamaño (~ 16.000 Mbp por genoma haploide). Las variedades comerciales son generalmente líneas puras. Su genoma no ha sido aún secuenciado, pero existen consorcios científicos activos que permiten pensar en un avance muy rápido durante los próximos años. En cualquier caso, la información disponible sobre secuencia de EST, BAC y mapa físico es de las más detalladas en plantas, además de una importante colección de materiales y fenotipos desde muy antiguo y acorde con una comunidad científica internacional numerosa y bien organizada.

La MAS se practica en programas públicos y privados para más de 65 *loci* donde se encuentran genes/QTL mayores, muchos de los cuales determinan la resistencia a diversas enfermedades y plagas, juntamente con algunos genes de calidad del grano o resistencias a estrés abiótico (Gupta *et al.*, 2010). Habitualmente la selección se hace para varios de estos caracteres al mismo tiempo. No siempre existen marcadores suficientemente cercanos a estos *loci*, ni tampoco es posible asegurar su polimorfismo en la mayoría de poblaciones a analizar. Aunque ya existen mapas de alta resolución basados en marcadores tipo SSR, RFLP y DArT, el futuro desarrollo de chips de SNP de alta densidad y la creciente información de secuencia basada en NGS facilitarán la resolución de estos problemas.

Se están construyendo poblaciones MAGIC para el trigo (Cavanagh y Morell, 2008), de 1.500 y 5.000 RIL a partir de cuatro y ocho fundadores, respectivamente. Estas poblaciones serán de ayuda para mejorar el conocimiento actual sobre diversos caracteres de herencia compleja, particularmente los relacionados con altura de la planta y época de maduración, aunque otras poblaciones deberán ser construidas para otros caracteres importantes. Aproximaciones de mejora tipo MARS o GWS no han sido aplicadas aún al trigo.

- **Melón:** Una de las hortalizas más cultivadas en el mundo, con centro de origen primario probable en el Cuerno de África y centro secundario en la India. Aunque fue uno de los primeros modelos para la comprensión de las leyes de la herencia (Sageret, 1926), su conocimiento genético se estancó y no ha reverdecido hasta las dos últimas décadas, gracias a una pequeña y activa comunidad científica y al gran interés comercial que la mayor parte de las empresas de semillas hortícolas tiene por esta especie. El melón es una especie alógama, con un nivel de variabilidad intermedio, sin otras especies próximas con las que se pueda intercrossar, diploide ($2n=2x=24$) y con un genoma pequeño (450 Mbp). Este genoma está en las últimas fases del proceso de secuenciación, que ha sido posible gracias a la tecnología NGS y a la existencia de una fuerte base previa, particularmente un mapa genético saturado, un mapa físico y una colección importante de EST, todos ellos en una base de datos de acceso público (<http://www.icugi.org/cgi-bin/ICuGI/misc/project.cgi>).

Se han descrito más de 100 genes en melón, de los que al menos 29 se habían localizado en el mapa de la especie (Pitrat, 2006). De ellos, al menos 10 se usan regularmente para la MAS, la mayor parte genes que confieren resistencia a enfermedades. Entre estos últimos, cuatro (dos resistencias a enfermedades y dos genes relacionados con la determinación del sexo) han sido clonados posicionalmente. Puesto que el mayor esfuerzo de mejora se hace en la obtención de híbridos F1, se usa frecuentemente la vía de las líneas diplohaploides (que se identifican con marcadores) y la mejora por retrocruzamiento cuando se trata de incorporar algún gen

procedente de una línea exótica, lo que se facilita con la selección de todo el genoma usando marcadores que cubran el genoma a baja densidad.

Existe una colección de líneas casi-isogénicas (NIL) basada en fragmentos de una variedad coreana en el fondo de variedad Piel de Sapo, la más cultivada en España (Eduardo *et al.*, 2005). Esta colección ha sido muy útil para el análisis de caracteres de herencia compleja como la forma y tamaño del fruto, la época y tipo de maduración y algunas resistencias a enfermedades poligénicas como el CMV (virus del mosaico del pepino). En la actualidad, se están resecuenciando algunas líneas antiguas y comerciales de melón que proporcionarán una enorme cantidad de SNP que permitirán ensayos para la caracterización del DL, como primer paso para la aplicación de la genética de asociación para el análisis genético.

- **Melocotón:** El frutal de clima templado más cultivado después del manzano y el peral. Perteneció a un género (*Prunus*) en el que se encuentran más de un centenar de especies entre las cuales algunas cultivadas como el albaricoque, cerezo, ciruelo y almendro, y otras que se usan como ornamentales o portainjertos. Algunas de estas especies son intercompatibles y producen descendencia fértil, lo que supone un raudal de variabilidad potencial que casi no ha sido usada. El melocotonero es un árbol originario de China cuyo sistema de reproducción comercial es por la vía vegetativa (injerto), aunque tiene un sistema de reproducción sexual operativo y suele autofecundarse. Si se reproduce por semillas su período intergeneracional es relativamente corto (2-4 años). Su genoma es diploide ($2n=2x=16$) y muy pequeño (230 Mbp), su nivel de variabilidad bajo, con un reciente evento de deriva genética que en conjunto determinan la existencia de estructuras subpopulacionales y una elevada conservación del DL (Aranzana *et al.*, 2010).

Existe información sobre varias decenas de genes mayores, de los que 28 están integrados en el mapa de alta densidad de referencia en la especie (Dirlewanger *et al.*, 2004), la mayoría son genes relacionados con las características del fruto y de la planta. El nivel de aplicación de marcadores en mejora está en alza pero es todavía escaso, con solamente cuatro genes, dos relacionados con el fruto, uno de forma y otro de acidez de la pulpa, y dos más de resistencia a nematodos (en este caso solamente para portainjertos), que sean usados habitualmente en MAS. Otra aplicación habitual es la caracterización de parentales y su selección basada en el genotipo de marcadores para predecir distancia genética, nivel de segregación o presencia de determinados alelos asociados con caracteres de interés. La introgresión de genes de otras especies (almendro concretamente) está también en marcha con aproximaciones de selección de todo el genoma.

Una de las restricciones de la aplicación de la MAS es el propio sistema de mejora en esta especie, influenciado por el largo período intergeneracional y la reproducción vegetativa del melocotonero, en el que las nuevas variedades se seleccionan de las descendencias del cruzamiento entre dos parentales parcialmente heterocigotos y más o menos relacionados, lo que implica que la recombinación es muy limitada. Por otra parte, los caracteres más interesantes son de tipo cuantitativo (relacionados con la calidad del fruto). En estas circunstancias, la aplicación de aproximaciones de selección tipo MARS o de selección genómica a la búsqueda de parentales de alto valor de mejora parecen opciones de gran valor en el futuro. Estas aplicaciones están al alcance de la mano desde el punto de vista de recursos disponibles puesto que el genoma del melocotonero acaba de ser secuenciado y hay en curso varios proyectos de resecuenciación que producirán en los próximos años las herramientas necesarias para realizarlas (<http://www.rosaceae.org/>).

5. Aplicación de la selección genética y genómica en silvicultura

Los árboles viven muchos más años que el resto de seres vivos. Esta obviedad es la principal razón por la que los planteamientos de programas de mejora genética forestal son muy diferentes al de otras plantas y animales. Esos programas son mucho más recientes que los de otras especies y las poblaciones mejoradas están a muy escasa distancia genética de las poblaciones naturales, manteniendo un alto grado de variabilidad susceptible de ser utilizado. Hasta hace poco tiempo los incrementos en producción se conseguían ampliando las plantaciones aunque en la actualidad ya son numerosos los programas de mejora en todo tipo de árboles. La meta de la mejora forestal es el incremento de la cantidad y la calidad de la madera y los productos derivados.

El uso de marcadores que correlacionen las diferencias fenotípicas con variaciones genéticas puede acelerar esos procesos de mejora. Los tiempos largos de generación y la escasa correlación observada entre las características de interés para individuos jóvenes y maduros son causa de que los marcadores moleculares tengan aún un mayor interés en el caso de la mejora forestal.

En principio hay dos estrategias para identificar genes de interés. La estrategia clásica se basa en la búsqueda de QTL en cruzamientos entre árboles con comportamientos diferentes para el carácter en estudio. Hay un planteamiento más reciente que consiste en buscar genes por estudios de asociación (correlación) entre la variación para un marcador molecular y la variación fenotípica para el carácter. Como se verá más adelante, la aproximación clásica tiene una serie de problemas que reducen drásticamente su eficacia por lo que actualmente se tiende a emplear la estrategia de estudios de asociación.

El empleo de herramientas moleculares en mejora forestal ha ido creciendo en la medida que se ha ido avanzando en el estudio de sus genomas. Se han construido mapas genéticos para un gran número de especies forestales (<http://dendrome.ucdavis.edu>), principalmente para identificar QTL. Por su mayor complejidad, hay menos trabajos que comprueben el grado de conservación de los QTL entre especies, aunque hay dos notables excepciones. En el primer caso, se ha podido confirmar la conservación para QTL de floración entre *Quercus* y *Castanea* (Casasoli y col., 2006), mientras que en el segundo se ha visto para QTL de densidad de la madera entre especies de *Pinus* (Telfer y col., 2006).

Por el momento, el único genoma completamente secuenciado es el de una especie de chopo (*Populus trichocarpa*) pero están en fases muy avanzadas las secuenciaciones de diferentes especies en géneros como *Eucalyptus*, *Pinus* o *Quercus*. El resto de especies tienen en mayor o menor medida información sobre mapas saturados de marcadores moleculares o secuencias parciales. Estos datos estructurales de los genomas forestales (dónde están los genes y qué secuencia tienen) son además complementados con la información funcional de esos genomas (qué genes se expresan en un determinado momento y para qué) gracias a la utilización de la tecnología de *microarrays*, que permite el análisis de miles de genes a la vez, y de estudios proteómicos que analizan el conjunto de proteínas que se encuentran en la célula en un determinado momento. Hasta el momento, se han identificado nume-

rosos genes candidatos a controlar todo tipo de caracteres, como los de crecimiento, de simbiosis, de embriogénesis o de biosíntesis de la pared celular. Así, en la actualidad hay herramientas para mejorar la mayor parte de especies forestales basándose en la combinación de la caracterización fenotípica de los árboles y la información molecular. Si sabemos donde se localizan los genes y tenemos una idea de su función podremos mejorar muchos de los QTL que por el momento se resisten a ser manipulados.

5.1. Estrategia clásica de identificación de QTL

La mayor parte de QTL localizados por el momento están relacionados con la adaptación y el crecimiento de los árboles y con características madereras. Entre otros, se han localizados algunos en relación con las propiedades físicas de la madera que afectan a la dureza (densidad de la madera, ángulo de las microfibrillas, etc.) o con las propiedades de la pasta de celulosa para hacer papel (producción de pasta, proporción de celulosa, longitud de las fibras, etc.). Además, se han identificado QTL para resistencia a enfermedades, tiempo de floración, resistencia al frío, composición en aceites de las hojas, y muchos más (Butcher y Southerton, 2007).

Por lo general, los QTL se han identificado en descendencias segregantes pequeñas, de 100-200 individuos. En este tipo de poblaciones los efectos de los QTL tiende a sobreestimarse. Su efecto se ve además notablemente afectado por el fondo genético, el ambiente y las interacciones con otros QTL. Por estas razones la localización de los QTL descritos hasta el momento en árboles es más imprecisa de lo que sería deseable si se desea identificar los genes que los regulan. Debe tenerse en cuenta que en otras especies animales o vegetales el número de cruzamientos que se pueden analizar es mucho mayor, permitiendo el estudio del mismo QTL en diferentes ambientes o fondos genéticos. Los pocos casos en árboles en los que se ha podido estudiar un QTL en diferentes condiciones son poco satisfactorios ya que al menos la mitad de ellos no eran fiables (Wilcox y col., 2007).

La otra vía aplicada para mejorar la fiabilidad de los QTL es la búsqueda en las regiones genómicas correspondientes de genes candidatos. Mediante mapeo comparativo con otras especies vegetales se han identificado numerosos genes candidatos para caracteres de calidad de madera, floración o estreses bióticos en especies de *Populus*, *Eucalyptus* o *Pinus*. Los resultados de mapeo comparativo son más esperanzadores porque se ha podido comprobar que se conserva la localización y el orden de muchos genes en especies alejadas dentro del reino vegetal. Es pues una fuente de información con gran potencial para el estudio y mejora de QTL en árboles.

En resumen, esta estrategia de MAS basada en la detección de QTL parece por el momento limitada para QTL muy concretos: de efectos muy grandes, en caracteres que sean muy costosos o difíciles de evaluar en campo, que impliquen la destrucción de individuos o que hayan demostrado escasas correlaciones entre el carácter en estados juveniles y adultos. El sistema mejora su eficacia si se encuentra el gen candidato acertado que puede utilizarse como marcador molecular diagnóstico de alta fiabilidad.

5.2. Estrategia de estudios de asociación

En la introducción de este informe ya se ha descrito que ante las limitaciones del análisis clásico de QTL se ha planteado el estudio directo de la correlación entre las variaciones observadas en los fragmentos de ADN y las variaciones para el carácter de interés. En el caso de la mejora forestal la aplicación de esta estrategia es muy prometedora debido a las peculiaridades de estas especies que dificultan la búsqueda clásica de QTL:

- Su gran tamaño y larga vida implica un coste en la creación y mantenimiento de descendencias segregantes. En el caso de las coníferas, por ejemplo, una generación puede conllevar unos diez años, y en el de algunas angiospermas como los robles incluso más.
- Tardan mucho tiempo en expresar de forma fiable los caracteres de interés en mejora, lo que conlleva muchos años de espera en la caracterización fenotípica de los árboles.
- Hay un alto grado de conservación en el orden de las secuencias de ADN dentro de los cromosomas (sintenia), que facilita el intercambio de datos entre especies.

En la estrategia de búsqueda de asociaciones se emplean poblaciones naturales. Los marcadores moleculares asociados al carácter se pueden emplear posteriormente para seleccionar individuos en poblaciones de mejora. Frente al concepto de selección asistida por marcadores (MAS), que se refiere al método clásico de identificar genes de interés en cruzamientos, se contraponen el término selección asistida por genes (GAS, del inglés) que utiliza esta nueva forma de localizar genes por asociación de la variación en las secuencias con la variación del carácter en estudio.

La idea de seleccionar individuos basándose en la variación para las secuencias de ADN no es nueva; de hecho, la MAS se basa en ese mismo principio. En cualquier caso, hay notables diferencias entre la mejora asistida por genes y la asistida por marcadores que son muy relevantes en relación con los sistemas de mejora forestal. En primer lugar, la MAS busca los genes de interés dentro de familias, mientras la GAS lo hace entre familias, ampliando así la probabilidad de éxito. Además, los marcadores del tipo de los AFLP o los microsatélites suelen emplearse para construir los mapas que se utilizarán para la MAS, mientras que en el caso de la GAS es indispensable el empleo de marcadores muy próximos o incluso dentro de genes (SNP o indels).

Ha sido a partir de la aparición de tecnologías de producción de grandes cantidades de marcadores moleculares cuando se han abordado los estudios globales de asociación. Recientemente, se han identificado marcadores del tipo SNP asociados a un gen de la ruta de biosíntesis de lignina en *Eucalyptus nitens*. En primer lugar se detectaron dos haplotipos significativamente asociados a la dureza de la madera y esos resultados fueron confirmados mediante análisis en descendencias segregantes de *E. nitens* y *E. globulus* (Thumma y col., 2005). Hay otros ejemplos de estudios de asociación global de SNP que han identificado genes de la ruta de biosíntesis de la pared celular, de extremo interés, tanto para mejorar la calidad de las maderas como para facilitar el proceso de producción de papel. Además, toda la información obtenida en especies tan alejadas de los árboles como arroz o *Arabidopsis* puede

aportar datos muy valiosos en estudios de asociación en especies forestales. De este modo, se han identificado distintos tipos de genes en relación con defensa de patógenos, tolerancia a sequía o tiempo de floración (Wilcox y col., 2007).

Hay unanimidad en la idea de que en los programas de mejora forestal es más provechosa la aplicación de una estrategia de asociación que la basada en identificar QTL en cruzamientos. Veamos brevemente cuales son los requisitos necesarios para la aplicación de los estudios de asociación:

1. Integración en programas de mejora: Aunque ya hemos visto que en principio no se emplean cruzamientos, los estudios de asociación deben incorporarse en un programa en marcha de mejora clásica donde se puedan estudiar caracteres verdaderamente interesantes desde el punto de vista económico. Una vez identificados los genes de interés se pueden estudiar y seleccionar ya en los cruzamientos de los programas de mejora.
2. Diseño experimental: Se emplean diferentes diseños experimentales, con complejos análisis estadísticos basados en el estudio de las poblaciones no estructuradas en combinación con descendencias segregantes emparentadas.
3. Acceso a tecnologías de genotipado de alta resolución: en el caso de la investigación forestal existe una tradición bastante alejada del trabajo en un laboratorio de genética molecular. En la actualidad ya no se precisa la puesta a punto de tecnologías o infraestructuras para el desarrollo de marcadores pues existen numerosas compañías y laboratorios públicos que ofrecen servicios de esa índole. En cualquier caso, sigue siendo necesaria una mejor comunicación y educación mutua entre mejoradores y genéticos moleculares.
4. Selección de los marcadores moleculares: es obviamente imprescindible elegir correctamente los marcadores adecuados y las fuentes pueden ser varias:
 - a. SNP analizados en plataformas de *microarrays*.
 - b. Secuenciación (resecuenciación de genomas completos).
 - c. Secuenciación (resecuenciación de porciones de genomas).
 - d. Marcadores previamente identificados por MAS.
 - e. Genes candidatos.

La elección de los marcadores está directamente relacionada con el avance en el conocimiento de los genomas en las especies forestales y en la medida en la que se pueden extrapolar datos de otras especies de plantas. Nos detendremos en la selección de genes candidatos pues existen diversos criterios que pueden emplearse y que describiremos brevemente. En todas las especies forestales más estudiadas existen colecciones de secuencias de genes que se expresan en un momento determinado en un tejido concreto (EST, del inglés *Expressed Sequence Tags*) y esta es la fuente primaria de información para identificar genes candidatos. Además, se pueden identificar genes procedentes de otras especies (*Populus*, *Arabidopsis*, arroz, etc.). Es significativo que en el trabajo en *Eucalyptus nitens*, donde pudieron identificar el papel de un gen de Cinnamoyl CoA Reductasa (CCR) en la calidad de la madera, a partir de la información en una especie tan alejada de la fisiología de los árboles como es *Arabidopsis thaliana* (Thumma y col., 2005). En cualquier caso, los estudios genómicos y proteómicos de expre-

sión global de genes en especies forestales son ya muy comunes. Normalmente, los resultados de expresión se combinan con experimentos clásicos de análisis de QTL. Cuando la segregación de un QTL concreto coincide con la diferente expresión de un gen se consigue un magnífico candidato para su caracterización en detalle. Por ejemplo, se ha identificado un gen de Dehidrina que presenta diferencias de transcripción asociadas con la densidad de la madera y la tasa de crecimiento en *Pinus radiata* (Cato y col., 2006).

5.3. Integración de la selección genómica en los programas de mejora forestal

La mejora forestal tiende a utilizar poblaciones heterogéneas donde se encuentran segregando diversos genotipos en forma jerarquizada. Sobre la base de las poblaciones naturales hay unas *poblaciones mejoradas* en las que se ha ejercido un primer nivel de selección y a partir de los mejores individuos de esas poblaciones mejoradas se obtienen las *poblaciones productoras*. Como puede observarse, se conforma una estructura piramidal con mayor diversidad en la base. Al principio de un programa de mejora las poblaciones mejoradas y las productoras son la misma y es sobre las mejoradas sobre las que se ejerce una continuada selección de los alelos favorables. Todo el esquema suele complicarse enormemente porque por lo general la selección se realiza simultáneamente para muy diversos caracteres. La selección genética y genómica debe incorporarse en los esquemas de estos programas para reducir costes (en tiempo y dinero) y aumentar la eficiencia del proceso.

1. Selección de los árboles "plus".

Antes de iniciar un programa de mejora se pueden utilizar los marcadores para seleccionar los árboles más diferentes desde el punto de vista genético, manteniendo así un alto grado de diversidad. Además, los marcadores asociados a caracteres de interés se podrán utilizar para seleccionar los mejores árboles. Sobre esa preselección de árboles se puede a continuación estudiar sus fenotipos para los caracteres de interés con un ahorro considerable de tiempo y dinero. En el caso de múltiples caracteres el beneficio es aún mayor.

2. Selección sobre las *poblaciones mejoradas*.

Las poblaciones mejoradas están conformadas por muchos individuos de orígenes diferentes y sin relaciones de parentesco para minimizar el efecto de la endogamia. Las poblaciones suelen estructurarse en familias o sublíneas para evitar esa consanguinidad. Los mismos marcadores empleados en la selección de los árboles "plus" pueden utilizarse en las diferentes familias de las poblaciones mejoradas para intensificar la selección. En el caso de varios caracteres puede decidirse en cuáles de ellos es más rentable el uso de marcadores y en cuáles es mejor medir el carácter.

Lo más habitual, salvo en caracteres muy complicados o costosos de medir, es que a la preselección previa realizada mediante los marcadores moleculares le siga, más adelante, una selección final por el dato fenotípico de la medida del carácter. De este modo, se pueden reducir costes, o aumentar la intensidad de selección al estudiar más individuos.

En los programas en los que se tengan que incorporar, por ejemplo, resistencias a patógenos, de las que hay mucha información sobre marcadores moleculares ligados, se evita la necesidad de infectar las familias para determinar cuáles son portadoras de resistencias.

3. Selección en las poblaciones productoras.

Las poblaciones productoras son una parte pequeña de la variación presente en las poblaciones mejoradas y proveen las semillas para su plantación o son utilizadas para su propagación vegetativa. Si por razones diversas quiere añadirse un carácter nuevo en el programa de mejora se puede emplear la selección genómica para identificar los mejores árboles en este material final. En este caso ya no se realiza una presión selectiva sobre las poblaciones y los marcadores se encuentran al final del programa de mejora.

Finalmente los marcadores pueden ser herramientas en el paso final del proceso de producción del material a comercializar. En una gran parte de los casos, el material es propagado mediante tecnologías de cultivo *in vitro*. A partir de pequeños trozos de hojas u otros órganos del árbol se pueden multiplicar muy eficientemente los genotipos a comercializar. Teniendo en cuenta el gran valor de cada uno de los árboles generados, los marcadores moleculares pueden servir para confirmar que todos los individuos son realmente clones del árbol original y además se podrá certificar la naturaleza y características de dichos árboles.

En resumen, la selección con marcadores moleculares puede incorporarse a los programas de mejora reduciendo los costes e incrementando la intensidad de la selección. De todas formas, cada programa de mejora debe estudiar muy detalladamente el coste-beneficio del uso de marcadores carácter a carácter. A este valor de los marcadores en la mejora forestal hay que sumarle su utilidad en el proceso de post-mejora donde puede incrementar el valor añadido del producto final mediante su certificación.

La investigación forestal es una de las actividades donde más beneficio se puede conseguir con la implementación de las tecnologías de análisis genómico. A pesar de ello, la aportación de las herramientas moleculares a la mejora forestal no está todo lo extendida que cabría esperar. Hay aún un largo camino por recorrer que previsiblemente se irá reduciendo poco a poco. A diferencia de la situación en plantas herbáceas y animales, el amplio margen de mejora aún presente en las poblaciones forestales, recientemente domesticadas, es el marco idóneo para sacar el máximo provecho de la información genómica. Los avances en la producción de marcadores de alta resolución y en la secuenciación masiva de genomas, unidos a la estrategia de análisis de asociación son ya responsables de la identificación de un elevado número de genes de interés que hasta ahora escapaban a los análisis clásicos de QTL. A continuación, a título de ejemplo, se describirá el caso concreto del avance en el conocimiento del genoma de *Eucalyptus*, y su aplicación a la mejora.

5.4. Genética y genómica de *Eucalyptus* y su aplicación a la mejora

Eucalyptus es el género de árboles más plantado en el planeta, debido a su alta eficiencia en la producción de pasta para papel, madera y energía. El incremento de esa productividad y la mejora de la calidad de sus productos es un planteamiento

habitual en los programas de mejora. Los avances en el conocimiento de su genoma (mapas genéticos y físicos, análisis de QTL, secuenciación, genómica funcional, etc.) están permitiendo profundizar en el conocimiento de esos caracteres de interés y su aplicación en la mejora (Myburg y col., 2007).

El eucalipto es una angiosperma de origen australiano que comprende varios cientos de especies diferentes. Tras su descubrimiento en el siglo XVIII, se extendió rápidamente su empleo en plantaciones forestales en muy diferentes regiones del mundo. Estos árboles son, por lo general, de fecundación cruzada, marcada autoincompatibilidad y polinización por insectos. Las especies de eucalipto presentan una gran variación fenotípica que ha permitido grandes avances en la mejora mediante la creación de híbridos interespecíficos, que también aparecen de forma natural. Esa gran variabilidad hace adivinar la existencia de muchos alelos favorables en baja frecuencia que pueden ser localizados con las nuevas estrategias genómicas.

En la segunda mitad del siglo pasado se iniciaron los programas de mejora en diferentes países. Los principales caracteres seleccionados han sido aquellos que tienen que ver con el crecimiento y la calidad de la madera y la producción de pulpa para papel. Además, se manejan otros aspectos relacionados con estreses bióticos y abióticos que tienen un efecto sobre esos caracteres principales. La mejora tradicional ha alcanzado notables ganancias en la fase de domesticación, mediante la selección de individuos. Además, la propagación vegetativa por esquejes de árboles seleccionados ha conseguido homogeneizar las poblaciones y capturar esa ganancia. La propagación clonal y la generación de híbridos han sido las herramientas clave para la mejora en eucalipto (Poke y col., 2005).

Los programas de eucalipto han apostado por mejorar la calidad de la madera en detrimento de la velocidad de crecimiento. El objetivo final es aumentar el rendimiento produciendo una madera más densa que contenga más celulosa, aunque se tarde más en obtenerla, porque los ahorros en toda la cadena de producción de papel son considerables (Ikemori y col., 1994). Desde el punto de vista genético, el eucalipto está aún en fase de domesticación, sobre todo si lo comparamos con plantas como el trigo o el maíz u otros árboles. Como vimos anteriormente, las poblaciones de mejora están muy próximas a sus parientes silvestres, manteniendo una gran parte de la variabilidad natural sobre la que se pueden aplicar las herramientas genómicas.

5.4.1. Selección genética

Desde la aparición en escena de los marcadores moleculares hay numerosos ejemplos de su empleo desde las fases iniciales de los programas de mejora de eucalipto. Los marcadores se han utilizado para caracterizar la variabilidad de las poblaciones así como para establecer relaciones entre poblaciones. Una de las áreas en las que han tenido más éxito los marcadores ha sido en la identificación de individuos y sus clones para su propagación, pues un error de etiquetado puede provocar retrasos de años en un programa. Con esta finalidad se han empleado marcadores RAPD (Keil y Griffin, 1994), AFLP (Gaiotto y col., 1997) o *microarrays* (Lezar y col., 2004) pero, como en otros casos, los microsatélites han sido el sistema utilizado con más éxito (Ottewell *et al.*, 2005). Además, se han utilizado los marcadores para caracterizar y gestionar la variabilidad en colecciones de germoplasma o determinar las mejores combinaciones en híbridos interespecíficos (de Aguiar y col., 2007).

Desde el primer mapa genético de eucalipto a principio de los años 90 hasta la actualidad, se ha producido una evolución de los tipos de marcadores utilizados. Hay un mapa consenso con 234 microsatélites que abarca un 90% del genoma y que ha permitido combinar los datos de muchas de las especies del género que habitualmente se usan en mejora (Brondani y col., 2006). Como se presuponía, el nivel de conservación y colinealidad de los genomas es muy elevado y permite la realización de muy diversos híbridos interespecíficos. A pesar del moderado nivel de saturación de esos mapas, se han podido realizar numerosos trabajos de análisis de QTL ligados a algunos de esos marcadores (Grattapaglia y Kirst, 2008). Aún se precisa incrementar la saturación de los mapas genéticos para poder validar los QTL de unas especies en otras.

Se han localizado numerosos QTL mayores para todos los caracteres de interés en eucalipto aunque, debido al tamaño de las poblaciones, con insuficiente precisión para utilizarse en mejora. El éxito en la localización de tantos QTL se debe sobre todo a la escasa domesticación del eucalipto y a la gran variabilidad presente en los híbridos interespecíficos que se usan en mejora. Hay QTL controlando caracteres juveniles como altura de la plántula, área de la hoja y tolerancia al frío de la plántula (Vaillancourt y col., 1995; Byrne y col., 1997a,b). Se han descrito QTL que regulan caracteres de capacidad de propagación vegetativa (Grattapaglia y col., 1995; Marques y col., 1999). También se han localizado QTL de floración (Missiaggia y col., 2005), de contenido en aceites (Shepherd y col., 1999; Henery, 2007) y, sobre todo, de propiedades de la madera (Grattapaglia y col., 1996; Verhaegen y col., 1997; Thamarus y col., 2004; Kirst y col., 2004; Freeman y col., 2009; Thumma y col., 2010). Además, diferentes QTL de resistencia biótica se han identificado (Grattapaglia y Kirst, 2008). Pese a ese importante número de QTL, por el momento la selección de muchos de ellos en programas de mejora se ha complicado por la gran heterogeneidad presente en las poblaciones de eucalipto y por la dificultad de autofecundar individuos para reducir dicha heterogeneidad. Además, sería preciso mejorar la resolución de los mapas genéticos mediante el estudio de más individuos y marcadores reduciendo así el número de genes candidatos presentes en las regiones de los QTL. Los mapas comparativos de QTL en diferentes especies de eucalipto están permitiendo validar diferentes caracteres. Por ejemplo, se han identificado QTL de gran efecto en las propiedades de la madera en diferentes especies de eucalipto (Thamarus y col., 2004) y de producción de raíces adventicias en distintas poblaciones de *Eucalyptus globulus* (Marques y col., 2005).

5.4.2. Selección genómica

La situación está mejorando, como en el resto de especies forestales, gracias a los recientes planteamientos de identificación de genes con estudios de asociación a partir de marcadores de alta resolución. Los estudios realizados por el momento son escasos pero evidencian el futuro prometedor de esta estrategia. Se han identificado SNP asociados al ángulo de las microfibrillas en *Eucalyptus globulus* (Thumma y col., 2005) o el gen COBRA-like asociado al contenido en celulosa en *Eucalyptus nitens* (Thumma y col., 2009). En el momento en que esté completada la secuenciación del genoma será más viable el estudio de los genes presentes en cualquier región cromosómica y de las posibles variantes alélicas responsables de diferentes fenotipos.

Hasta hace unos pocos años la mayor parte de los proyectos de secuenciación eran de iniciativa privada y no eran de acceso público. Varias compañías apostaron por in-

vertir en estos proyectos (Genesis en Nueva Zelanda, Arborgen y Dupont en USA, Genolyptus en Brasil o ForEST en Brasil y Japón) y es prueba del interés desatado por acceder a la información genómica. Los programas de mejora de eucalipto de esas compañías ya están sacando provecho de esa información, aunque evidentemente no hacen públicos sus resultados. En el último lustro se han ido publicando secuencias de diversos laboratorios pero el momento clave ha sido el inicio del proyecto de secuenciación del genoma completo de *Eucalyptus grandis*. Desde mayo de 2010 es de acceso público el borrador 8X de la secuencia completa (Eucalyptus Genome Database, <http://eucalyptusdb.bi.up.ac.za/>), y en breve se publicará la secuencia finalizada. A todo ese caudal de datos hay que sumar todos los resultados acumulados en el estudio funcional de genes en eucalipto y otras especies próximas, tanto de genómica funcional como de proteómica. Es de esperar que en la próxima década se puedan identificar una gran parte de las variantes alélicas presentes en muchos de los genes de eucalipto. Como ejemplo del avance continuo en el conocimiento genómico y del acceso a nuevas herramientas, se ha presentado una nueva plataforma de *microarrays* (DArT) que reduce el coste de análisis de las variantes alélicas en 7.896 genes de eucalipto a tan solo 50 dólares por muestra (Sansaloni y col., 2010). En estos momentos empieza a ser más limitante la cantidad y calidad de las poblaciones experimentales caracterizadas morfológicamente que la información genómica.

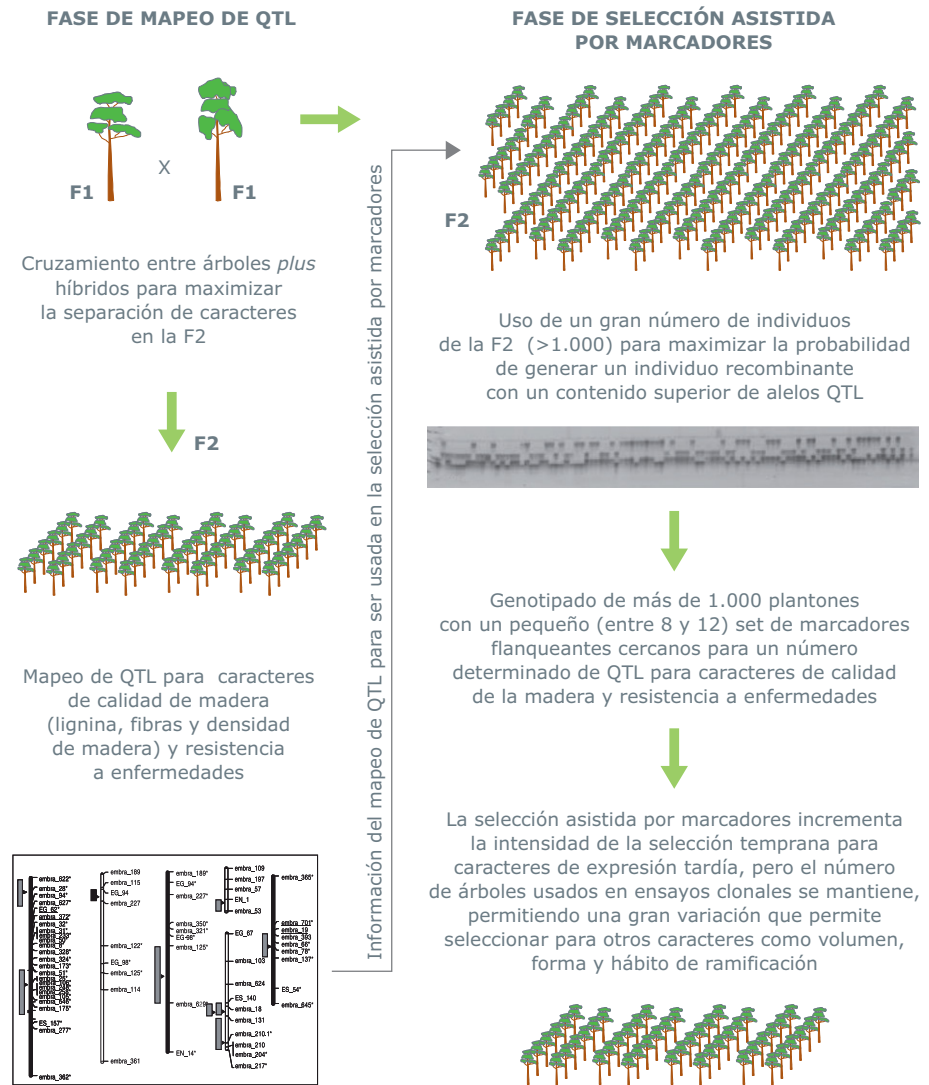
Un ejemplo de la importancia de la aplicación de tecnologías moleculares en los programas de mejora en eucalipto en España lo constituye el Grupo Empresarial ENCE, S.A., que se dedica a la producción de celulosa y energía a partir de madera de eucalipto. Desde hace más de 25 años, ENCE está desarrollando un programa de mejora genética de eucalipto para destino celulósico. Los resultados conseguidos le han permitido seleccionar variedades con mejor adaptación a las condiciones ecológicas de cultivo, lo que ha permitido duplicar la producción de madera.

Como soporte de este programa y para complementar su estrategia, ha realizado varios proyectos de desarrollo y utilización de marcadores moleculares para la identificación de variedades, reconstrucción de paternidad, la asociación de regiones genómicas que controlan caracteres de interés económico, la detección de genes asociados a rasgos de importancia y su posterior uso en selección masal asistida por marcadores moleculares. ENCE continúa apostando por la selección molecular en sus programas de mejora.

A modo de resumen, la situación actual de la mejora en eucalipto se resumiría en los siguientes puntos:

- Está muy extendido el empleo de marcadores moleculares para la selección clonal y para la identificación de QTL de interés.
- La mayor parte de programas de mejora emplean los marcadores moleculares donde son claramente más ventajosos que medir el carácter de interés: caracteres complicados de medir y que requieren llegar a la madurez de los individuos.
- La apuesta de diversas compañías por la secuenciación y las herramientas genómicas disponibles demuestran que la selección genómica en eucalipto ya es una realidad.
- La secuenciación del genoma de eucalipto y las herramientas de genotipado de alta resolución disponibles van a permitir identificar alelos de gran efecto en baja frecuencia en las poblaciones.

Figura 4
ESQUEMA PARA LA SELECCIÓN ASISTIDA POR MARCADORES
DE LOS ÁRBOLES PLUS EN UN PROGRAMA DE MEJORA FORESTAL



Fuente: Gratapaglia, D. (2008). Genomics of Eucalyptus, a global tree for charge pages and wood. In: Moore, P.; Ming, R., eds. Genomics of tropical plants. New York, 259-298.

6. Aplicación de la selección genética y genómica en acuicultura

6.1. Producción y mejora genética en acuicultura

La producción de la Acuicultura ha crecido de manera espectacular desde menos de 1 millón de toneladas a mediados del siglo pasado, hasta cerca de 70 millones de toneladas documentadas en 2007. Mientras que la producción de la pesca extractiva se ha estancado durante la década de los 90 como consecuencia de la sobreexplotación de los caladeros y el deterioro medioambiental, la contribución de la acuicultura al suministro mundial de peces, crustáceos y moluscos para consumo alimentario ha crecido desde el 5% en 1970 hasta casi el 50% en 2007. Casi la mitad de la producción mundial de la acuicultura actual es de origen piscícola, pero el incremento de la producción ha tenido lugar en todos los grupos, incluyendo vegetales, moluscos y crustáceos (APROMAR, 2009).

En contraposición a las especies domésticas tradicionales, en las que la mayor parte de la producción se obtiene de un reducido número de especies de animales, la producción en acuicultura está enormemente diversificada. Actualmente se crían en el mundo cerca de 500 especies acuícolas, 250 de las cuales alcanzan producciones por encima de las 100 t, aunque las 10 primeras especies representan cerca del 50% de la producción total. A pesar de que la antigüedad de la acuicultura se remonta a más de 2000 años en China, el fuerte crecimiento de este sector es reciente, especialmente en comparación con los animales domésticos tradicionales. En consecuencia, los recursos genéticos en las especies de acuicultura se encuentran en poblaciones salvajes. Esto representa una ventaja, ya que es posible fundar los stocks de reproductores y planificar los programas de selección a medio y largo plazo desde su inicio y utilizando toda la diversidad presente en las poblaciones naturales (Hulata, 2001).

La elevada fecundidad de los peces junto con la existencia de puesta natural en muchas especies, ha propiciado que programas de selección masal hayan sido de aplicación usual en acuicultura. Esto ha originado, en no pocos casos, problemas de depresión consanguínea. Programas de selección familiar existían en 2005 en no más de 30 especies, debido a la complejidad de su desarrollo por las necesidades de espacio y al coste asociado a la trazabilidad genealógica con marcadores moleculares. Sin embargo, la buena respuesta a la selección y la creciente competitividad en este campo de la producción, están determinando una demanda cada vez mayor de la tecnología genética desde las empresas del sector. Por ejemplo, para tasa de crecimiento, uno de los caracteres principales de selección en peces, se están obteniendo valores de progreso por generación próximos al 10% (rango 5-20%) (Gjedrem, 2005).

La acuicultura se ha beneficiado de la elevada plasticidad genómica y fisiológica de peces y moluscos para la incorporación de técnicas de manipulación cromosómica y de reversión sexual en planes de selección. La manipulación del número de conjuntos cromosómicos ha permitido, mediante una serie de operaciones básicamente sencillas, la obtención de un conjunto de productos de interés en acuicultura

por su esterilidad (triploides) o por su sexo (ginogenéticos), que están siendo utilizados a nivel industrial en diversas especies (Piferrer *et al.*, 2009). Igualmente, mediante técnicas de reversión sexual se están produciendo de forma rutinaria poblaciones monosexo de machos o hembras, dependiendo de las ventajas que ofrece cada uno de los sexos para la producción (Piferrer y Guiguen, 2008).

En los últimos años, la irrupción de las estrategias genómicas junto con el progresivo abaratamiento de sus costes está propiciando un cambio de escenario en el campo de la mejora genética. Aunque de forma más lenta que en otras especies domésticas, la acuicultura ha ido incorporando las herramientas para el rastreo genómico mediante mapas genéticos; los *microarrays* para el análisis de expresión masiva de genes; la identificación de genes candidatos mediante estrategias de genómica comparada; y más recientemente, la utilización de las nuevas tecnologías de secuenciación masiva que están incrementando los recursos genómicos de forma exponencial. Además de disponer de secuencias más o menos completas de genomas de 5 especies modelo, en la actualidad se están secuenciando los genomas de varias especies de acuicultura, la lubina y el rodaballo entre otras. Por otro lado, la cantidad de etiquetas de genes (EST) gestionadas en bases de datos en peces de acuicultura ha crecido de forma muy notable (más de 1 millón) y actualmente se dispone de plataformas de *microarrays* y mapas genéticos en más de 20 especies (Bouza y Martínez, 2007; Canario *et al.*, 2008).

6.2. Rastreo genómico mediante mapas genéticos para la identificación de *loci* de QTL

La mayor parte de los caracteres de interés en producción son de naturaleza cuantitativa, es decir, muestran amplias distribuciones fenotípicas más o menos continuas. Estos caracteres están controlados por una importante cantidad de genes (poligenes) y sujetos a la influencia del ambiente. Aunque los poligenes tienen habitualmente efectos aditivos sobre el fenotipo, la dominancia y las interacciones entre poligenes y con el ambiente, son también importantes para entender la base genética de estos caracteres y su respuesta a la selección. En consecuencia, la posibilidad de abordar la mejora a través de su disección genética es una tarea ardua y costosa, por lo que las estrategias de la mejora genética tradicional se centraron en la descomposición genética de la varianza fenotípica utilizando correlaciones entre parientes. Con esta información se pueden estimar las heredabilidades, parámetro que permite predecir la respuesta a la selección en cada escenario particular (Falconer y Mackay, 2001).

Los desarrollos genómicos han permitido una disección más detallada de la base genética de los caracteres cuantitativos, aunque se está aún lejos de conocer toda la información necesaria para interpretar de forma precisa las relaciones entre genotipo y fenotipo. Sin embargo, una estrategia posible es la localización en el genoma de los efectos alélicos más notables asociados a un carácter mediante el rastreo con mapas genéticos. Esto es posible porque algunos poligenes (o grupos de poligenes ligados; QTL) tienen una mayor influencia sobre el carácter, lo que hace posible su detección siempre que se disponga de un mapa genético de densidad suficiente y de un diseño experimental y herramientas de análisis apropiadas. Para inferir la existencia de un QTL en una región genómica concreta, se evalúan las

asociaciones estadísticas entre marcadores del mapa y los fenotipos del carácter bajo estudio, teniendo en cuenta que los genes/marcadores localizados en la proximidad en los genomas tienden a transmitirse juntos (Mackay, 2001).

Para la identificación de QTL se utiliza la cartografía por intervalos (IM, *Interval Mapping*), que evalúa la significación estadística de cada intervalo (constituido por un par de marcadores) a lo largo de todo el mapa para proponer la existencia de un QTL en una región particular del genoma utilizando métodos de máxima verosimilitud o técnicas de regresión. El rastreo por intervalos es necesario, dada la incertidumbre que produce el manejo de dos variables simultáneamente cuando se trabaja con marcadores individuales (distancia entre el marcador y el QTL, y peso del QTL sobre el carácter). Para cuantificar la evidencia de un QTL suele aplicarse el método de las puntuaciones LOD, calculado como el log₁₀ de la razón de las probabilidades de la obtención de los datos fenotípicos observados asumiendo la existencia del QTL y sin el QTL. La significación estadística se obtiene a partir de la distribución de probabilidades generada mediante simulación, permutando aleatoriamente los marcadores un número elevado de veces. El tipo de cruzamiento más potente para la detección de QTL son los retrocruzamientos o cruzamiento F₂ obtenidos a partir de líneas homocigóticas de fenotipos extremos para el carácter bajo estudio. Este esquema es altamente informativo, ya que es de esperar la segregación simultánea de varios QTL y además, una gran parte de marcadores mostrarán asociaciones específicas con los alelos alternativos de los QTL en su proximidad. Esta estrategia es factible en las especies domésticas tradicionales en las que existen este tipo de líneas, pero no son frecuentes en especies de acuicultura. En éstas, sin embargo, es posible la obtención de líneas puras en una generación mediante técnicas de manipulación cromosómica (mitoginogénéticos o mitoandrogenéticos). La opción aplicada más frecuentemente en acuicultura es la de los cruzamientos entre hermanos, estrategia favorecida por la elevada prolificidad de ambos sexos. La diversidad interfamiliar, tanto en el número de QTL segregantes como en las asociaciones con los marcadores adyacentes, determina una pérdida de potencia respecto de los cruzamientos entre líneas homocigóticas descritos anteriormente (Pérez-Enciso y Toro, 2007).

Una vez detectado un QTL consistentemente, existe la posibilidad de aplicar esta información en programas de selección asistida por marcadores (MAS, *Marker Assisted Selection*) o intentar afinar la localización del mismo, ya que la precisión del análisis de ligamiento para estimar la posición de un QTL es muy baja. Esto serviría, bien para optimizar el marcaje en programas MAS o, eventualmente, para intentar identificar el gen o genes responsables del mismo. La idea en este caso es incrementar la densidad de marcadores en la región del QTL y realizar un análisis de asociación basado en el desequilibrio de ligamiento entre el QTL y los marcadores adyacentes en una población panmíctica. Dado que el desequilibrio se rompe con la distancia y con el paso de las generaciones, aquellos marcadores que muestren mayor desequilibrio estarán más cercanos al QTL. Esta premisa puede verse alterada por fenómenos demográficos específicos en la población bajo estudio, como la existencia de cuellos de botella o la mezcla de poblaciones diferenciadas, y también como consecuencia de fenómenos selectivos recientes (Pérez-Enciso y Toro, 2007). En todos estos casos, se producirían desequilibrios en amplias regiones genómicas limitando la eficacia del análisis. Una vez acotada la región genómica es factible la realización de un mapa físico de la misma a partir una librería de BAC o utilizar la información genómica, caso de existir, para, mediante estrategias bioin-

formáticas, detectar los genes presentes en la región. En último término, se trataría de identificar el gen/es responsable/s usando aproximaciones funcionales (Anderson y Georges, 2004).

6.3. Acuicultura y mejora genética del rodaballo

El cultivo de rodaballo se ha incrementado notablemente durante los últimos años hasta cerca de las 9.000 t. España, y en particular Galicia, es el país que más contribuye a la producción mundial, representando el 80% de la misma (APROMAR, 2009). El rodaballo es una especie con unas condiciones de cultivo complejas y que precisa cerca de 2 años para alcanzar la talla comercial. Asimismo, los ejemplares de mayor tamaño tienen un valor de mercado superior. Por otro lado, las hembras de esta especie tienen un crecimiento superior al de los machos, alcanzando la talla comercial varios meses antes. Finalmente, las condiciones de la acuicultura intensiva favorecen la transmisión de enfermedades, lo que ha obligado al desarrollo de vacunas para la protección masiva de los alevines, con el consiguiente incremento de costes. Por todo esto, los principales objetivos de la selección genética en rodaballo son el incremento de la tasa de crecimiento, la consecución de poblaciones todo-hembras y la selección para resistencia a patologías de interés industrial (Martínez, 2005).

Desde principios de los 90 se están desarrollando programas de selección familiar en algunas empresas del sector del rodaballo. Estos programas utilizan marcadores moleculares hipervariables como soporte para la trazabilidad genealógica. En esta especie se han realizado varios estudios genéticos para la catalogación de sus recursos en poblaciones naturales, y además se han puesto a punto técnicas de manipulación cromosómica y/o tratamientos hormonales para la obtención de individuos estériles y poblaciones monosexo (Martínez, 2005). Los recursos genómicos se han incrementado de forma muy notable en los últimos años. Se ha desarrollado un mapa genético de densidad apropiada para el rastreo de QTL (Bouza *et al.*, 2007) y diseñado un *oligo-microarray* (Millán *et al.*, 2010) utilizando una amplia base de datos de EST construida a partir de los principales órganos relacionados con la inmunidad (Pardo *et al.*, 2008). Finalmente, como se indicó anteriormente, el genoma de rodaballo está secuenciándose utilizando las nuevas estrategias de secuenciación masiva.

6.4. Identificación de QTL: aplicación al sector del rodaballo

Alrededor de 100 marcadores distribuidos homogéneamente por el genoma de rodaballo están siendo utilizados para el rastreo de QTL. Esto supone una densidad genómica aproximada de un marcador cada 10 cm y cada 6 Mb. Para el rastreo se están utilizando un número importante de familias de hermanos (150 individuos/familia) y los caracteres objetivo son la determinación del sexo, la tasa de crecimiento y la resistencia a furunculosis y ciliatosis, dos patologías de fuerte impacto industrial (Martínez *et al.*, 2009). En todos los caracteres estudiados se han detectado QTL significativos, algunos de ellos concordantes en diferentes familias, particularmente, para la determinación sexual y la tasa de crecimiento, lo que denota su consistencia. En el caso de la resistencia a patologías se han observado QTL de resistencia en las mismas regiones del genoma para diferentes patógenos, lo que sugiere la existencia de gen/es relacionados de forma genérica con la inmunidad.

Este hecho es relevante, por cuanto, tal como se ha venido argumentando, la selección para resistencia a un patógeno podría determinar sensibilidad frente a otros patógenos.

Paralelamente al rastreo de QTL, se están desarrollando estrategias para identificar los genes responsables de las diferencias observadas dentro de familias. Esta no es una cuestión sencilla y son pocos los casos en mejora animal en las que se ha logrado identificar el gen responsable asociado a un QTL (Andersson y Georges, 2004). Los esfuerzos se están dirigiendo por un lado, al incremento de la densidad del mapa genético, utilizando, en este caso, marcadores ligados a genes (microsatélites y SNP). Estos marcadores, además de aumentar la densidad de marcaje, identifican genes particulares, lo que facilita la localización de candidatos si se localizan en regiones próximas a los QTL. Por otro lado, el *microarray* desarrollado está permitiendo detectar genes regulados en respuesta a los mismos patógenos (furunculosis, ciliatosis). Estos genes están siendo mapeados a partir de marcadores o mediante aproximaciones de genómica comparada con especies modelo, y aportan candidatos consistentes cuando se localizan próximos a los QTL. Finalmente, se está realizando clonaje posicional a partir de una librería de BAC en aquellos QTL más consistentes y se está avanzando en la secuenciación del genoma del rodaballo (Proyecto Consolider Ingenio Aquagenomics).

6.5. La detección precoz del sexo en el rodaballo: una aplicación industrial de los desarrollos genómicos

Los peces son el grupo de vertebrados que muestran mayor versatilidad en cuanto a la reproducción, existiendo especies gonocoristas (sexos separados), hermafroditas e incluso, unisexuales. El sexo del adulto viene determinado por mecanismos (genéticos y/o ambientales) establecidos desde la concepción, pero también por los procesos de diferenciación gonadal, desde una gónada indiferenciada hasta la gónada madura definitiva. A diferencia de mamíferos y aves en los que el sexo está fuertemente determinado desde la concepción, los peces muestran una gran labilidad de desarrollo, de manera que factores como la temperatura o los tratamientos hormonales pueden cambiar el desarrollo gonadal. Este hecho ha facilitado la reversión sexual para la obtención de poblaciones monosexo a nivel industrial. Por otro lado, la determinación sexual puede ser tanto genética como ambiental, existiendo entre ambos extremos toda una gama de situaciones intermedias. Entre los mecanismos genéticos, se ha descrito la existencia de un solo gen determinante del sexo, de unos pocos genes o de múltiples poligenes, situación próxima a los caracteres complejos. Finalmente, la labilidad y versatilidad del sexo de los peces se manifiesta también a nivel evolutivo, de manera que especies próximas, o incluso poblaciones de una misma especie, muestran mecanismos de determinación sexual diferentes (Penman y Piferrer, 2008).

Como consecuencia de su dinamismo evolutivo, menos de un 10% de especies de peces han demostrado un heteromorfismo cromosómico asociado con el sexo. Este hecho ha dificultado el estudio de los mecanismos de determinación cromosómica del sexo en peces de acuicultura. Así en el rodaballo, no se han detectado diferencias cariotípicas entre machos y hembras ni con cromosomas mitóticos, ni con los más extendidos y apropiados para este abordaje, cromosomas meióticos. Es de suponer, por tanto, que la región diferencial determinante del sexo se circunscribe a

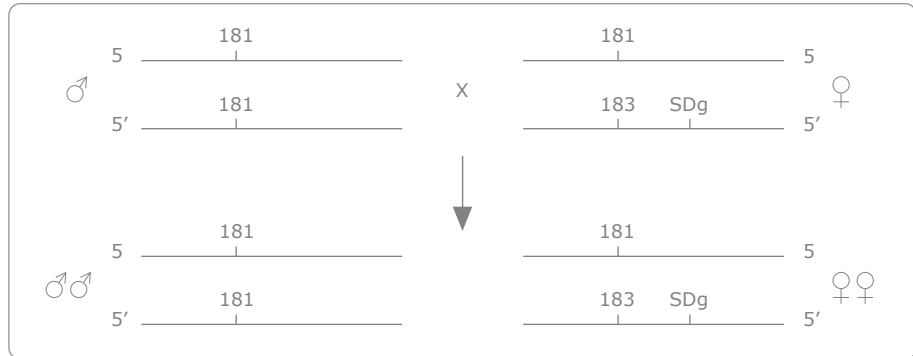
una pequeña región genómica. Con el desarrollo de la genómica disponemos de una gran cantidad de marcadores ordenados en mapas genéticos y colecciones de EST para el rastreo genómico. Los estudios de asociación y el rastreo de QTL han sido aplicados para acotar la región/es determinantes del sexo en peces. Igualmente, los *microarrays* han sido utilizados para identificar la cascada de expresión génica a lo largo del proceso de diferenciación gonadal. Finalmente, el incremento de los recursos genómicos en especies modelo ha posibilitado estrategias de genómica comparada y de genes candidatos para la identificación del gen/es determinantes del sexo en peces. Todas estas estrategias han permitido la identificación de la región determinante del sexo en ocho especies de peces, siendo medaka (*Oryzias latipes*) la única especie en la que se ha identificado inequívocamente el gen responsable (DMY; Matsuda *et al.*, 2002).

La búsqueda de QTL relacionados con el sexo en rodaballo reveló una región del grupo de ligamiento 5 fuertemente asociada con la determinación sexual. En todas las familias analizadas se detectó una asociación altamente significativa con la región proximal de este cromosoma, entre los marcadores Sma-USC30 y Sma-USC270. En realidad, el gen maestro determinante del sexo (*Sex Determining gene*: SDg) en esta especie estaría localizado a 2.6 cM (1.3 Mb) del marcador Sma-USC30. Utilizando este marcador fue posible clasificar correctamente el 98.4% de los individuos en 4 de las 5 familias analizadas. Efectos ambientales u otros factores genéticos menores explicarían la menor asociación detectada en una de las familias, lo cual sugiere la necesidad de nuevos estudios para resolver definitivamente la cuestión. El análisis de segregación del marcador Sma-USC30 en las familias demostró además, que el sexo de la progenie dependía del genotipo de la madre exclusivamente, no teniendo el padre aparentemente ninguna influencia. Esta observación es congruente con un mecanismo de determinación del sexo de tipo ZZ/ZW (machos ZZ y hembras ZW), de manera que si el hijo recibe el cromosoma Z de su madre será un macho, y si es el W, una hembra. Este mecanismo es el opuesto al observado en mamíferos (XX/XY) y en otras especies de peces como la carpa común, la trucha arcoíris o el salmón chino (Martínez *et al.*, 2009).

Esta información ha sido clave para el desarrollo de una herramienta molecular para la determinación precoz del sexo en rodaballo utilizando el marcador Sma-USC30, cuya solicitud de patente ha sido depositada. Tengamos en cuenta que la discriminación sexual no es posible hasta la madurez sexual (3-4 años). Partiendo de un pequeño fragmento de la aleta caudal del individuo de interés se extrae ADN, y mediante una PCR se amplifica el marcador Sma-USC30 para inferir su sexo. Teniendo en cuenta que entre el marcador y el gen maestro existe una cierta distancia, la asociación no existe a nivel de la especie pues se rompe por recombinación, y solo se puede testar a nivel familiar. Para poder establecer la asociación específica entre los alelos del marcador y el sexo en cada familia es preciso conocer el genotipo del marcador en otros parientes sexados de la misma familia.

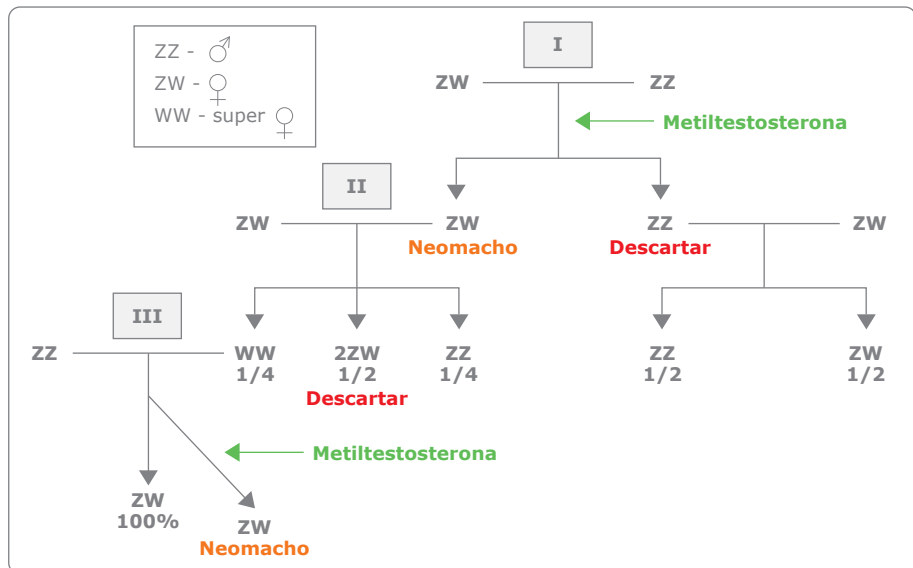
Al realizar una PCR en el marcador Sma-USC30 en una familia producto de un cruzamiento entre un padre 181/181 x una madre 181 x 183, todas las hijas resultaron heterocigotas 181/183, mostrando dos picos en el secuenciador, y los hijos fueron homocigotos, 181/181, mostrando un único pico. A continuación se muestra el esquema del cruzamiento en dicha familia:

Figura 5
CRUZAMIENTO ENTRE UN PADRE 181/181 Y UNA MADRE 181/183



Actualmente, esta herramienta está siendo aplicada en planes de selección genética con empresas del sector. Aunque el coste de la técnica es bajo (en torno a 1€), no es posible su utilización directa para seleccionar precozmente solo hembras para el cultivo, porque sería necesario además el marcaje individual de todos los peces (entre 2-3€ cada uno) hasta que se estableciera su genotipo. Esto comportaría un coste añadido que no compensaría la ganancia por el crecimiento superior de las hembras. Su aplicación, por tanto, tiene interés en planes de selección genética. La selección para crecimiento se realiza usualmente a partir de mezclas de familias antes del año de edad, momento en el que empezarían a manifestarse diferencias de crecimiento entre machos y hembras. Este es un dato empírico, que podría variar dependiendo de las familias. Además, el número de individuos seleccionados en cada grupo y, especialmente, en cada familia no es elevado, por lo que podría haber desviaciones importantes de las proporciones sexuales como consecuencia del muestreo. En estos programas, el genotipado para el marcador Sma-USC30 permite evaluar si la muestra seleccionada está compensada en cuanto a las proporciones sexuales. Sin embargo, la aplicación de mayor interés está en la obtención de poblaciones todo-hembras, combinando técnicas de reversión sexual con la información molecular del marcador.

Figura 6
ESQUEMA PARA LA OBTENCIÓN DE POBLACIONES TODO HEMBRAS EN RODABALLO



En la Figura 6 se muestra el esquema seguido para la obtención de poblaciones todo-hembras en una especie con un mecanismo de determinación ZZ/ZW, como el rodaballo. Esta técnica implica conseguir neomachos ZW utilizando metiltestosterona durante la etapa de desarrollo en que la gónada está aún indiferenciada (cruzamiento I). Estos neomachos serían machos fisiológicos (con testículos) pero hembras genéticas (productores de gametos Z o W). Una vez obtenidos los neomachos, se cruzarían con una hembra normal ZW, y de estos cruzamientos surgirían un 25% de superhembras WW (cruzamiento II). El cruzamiento de estas superhembras con machos normales daría lugar a progenies todo-hembras (cruzamiento III). El mayor problema que existía hasta ahora era diferenciar los neomachos ZW de sus hermanos ZZ, tras el tratamiento hormonal. Lo mismo diferenciar las superhembras WW de sus hermanas ZW en el cruzamiento II. La única solución era realizar un test de progenie individual para cada uno de los peces tratados. Si al cruzar los descendientes revertidos con hembras normales ZW su descendencia era 25% machos y 75% hembras, se trataría de un neomacho; o en el cruzamiento II, si al cruzar las hembras con machos normales produjeran solo hembras, se trataría de una superhembra. Obviamente, este abordaje supone un coste elevado y sobre todo un tiempo adicional importante para poder evaluar las proporciones sexuales en los test de progenie. La utilización del marcador Sma-USC30 permite identificar los neomachos y superhembras con una confianza elevada realizando únicamente una PCR, lo que supone ganar más de dos años en tiempo y a un coste mínimo por individuo.

En resumen, los recientes desarrollos genómicos realizados en rodaballo, han permitido el rápido desarrollo de una aplicación industrial para la detección precoz del sexo. En esta línea, se está avanzando en la localización al gen maestro responsable de la determinación sexual, mediante búsqueda de asociación tras un mapeo más fino de la región y clonaje posicional a partir de una librería de BAC. Por otro lado, algunos de los QTL identificados para crecimiento han evidenciado un peso sobre la varianza del carácter próximo al 40%, lo que los convierte en candidatos idóneos para su aplicación en programas MAS. Finalmente, los QTL de resistencia a patologías, aunque mostraron efectos mucho menores sobre el fenotipo, se están combinando con estrategias de genómica funcional y comparada, lo cual permitirá detectar genes candidatos que expliquen las diferencias entre individuos, y por tanto, ser utilizados para lograr stocks con mayor resistencia a estas patologías.

7. Conclusiones

Con la llegada de los marcadores moleculares parecía que se iban a poder diseñar programas de mejora que redujeran los costes y ampliaran las ganancias en la mejora de todo tipo de caracteres. Con el paso del tiempo se ha podido comprobar que el impacto de la selección asistida por marcadores (MAS) no ha sido tan notorio como cabía esperar y que, al menos en parte, sigue requiriendo del trabajo de campo y el análisis morfológico de los caracteres de interés. Es cierto que dicho impacto está subestimado porque, si bien los trabajos de identificación de QTL son de investigación pura y se publican en revistas científicas, en la mayor parte de los casos los resultados de su aplicación por parte de las empresas no se divulgan. Además, a pesar de que los marcadores moleculares comenzaron a utilizarse en los años 80, no fue hasta mediados de los 90 cuando se comenzaron a utilizar con asiduidad marcadores más fiables y útiles, como los microsatélites o los AFLP. Por lo tanto, también es esperable un incremento de las publicaciones según pase el tiempo y mejoren las tecnologías y la información genómica.

Desde la perspectiva de un mejorador, la selección genética o genómica no debe ser solo eficiente si no más eficiente que los métodos tradicionales de mejora. Debe por tanto tenerse en cuenta que siempre se está comparando la selección con marcadores con la selección fenotípica estándar.

Como una conclusión preliminar, podemos en primer lugar separar aquellos casos en los que los marcadores moleculares no son cuestionables. Estas son las situaciones en las que, en comparación con la mejora clásica, se consigue un ahorro en tiempo o en dinero debido a que el carácter se puede seleccionar en las fases juveniles y a que la medida del carácter es más cara que el genotipado del marcador. El ejemplo más claro es aquel que se produce cuando para medir el carácter hay que destruir al individuo. En el resto de situaciones, el éxito de la selección molecular dependerá en gran medida de la complejidad del carácter que se desea mejorar. Por ejemplo, su aplicación será más eficiente para un carácter monogénico que para uno cuantitativo, será mejor para uno cuantitativo con pocos QTL que para otro con muchos QTL, y será mejor para uno con un QTL mayor que para otro con varios QTL e interacciones epistáticas. La discusión evidentemente debe realizarse carácter a carácter.

La precisión con que se identifica un QTL es clave para poder aplicarlo en un programa de MAS, principalmente si trabajamos con caracteres con muchos QTL de pequeños efectos. Los factores que afectan esa precisión son variados, pero están en gran medida relacionados con el número de individuos muestreados y el número de marcadores analizados. Cuanto mayor sea ese tamaño muestral más fiable será la validación del QTL en diferentes fondos genéticos y ambientes, y menor será el tamaño del intervalo genómico donde se localice.

Es bien sabido que de un QTL identificado en una población no se pueden hacer previsiones de su presencia o efecto en otras poblaciones. Esto no significa que los QTL no sean válidos sino que en diferentes fondos genéticos se producen diferentes interacciones génicas y por tanto el fenotipo final será distinto. Cuando se intentan identificar QTL se realizan cruzamientos entre individuos con comportamientos en los

extremos de la distribución de un carácter y de este modo es fácil lograr el éxito. En muchas ocasiones, cuando esos QTL tratan de identificarse en las poblaciones de mejora, generadas a partir de líneas élite (que posiblemente tendrán un parentesco nulo con el material de mapeo empleado previamente), se fracasa en el intento.

La mayor parte de los QTL son consistentes en diferentes ambientes, aunque se registran notables variaciones en la magnitud de los efectos. Hay incluso casos perfectamente detallados en los que pueden aparecer efectos contrarios en función del ambiente, principalmente en QTL de efectos pequeños. Por lo tanto, en cada caso es preciso estudiar con detenimiento el componente ambiental asociado a cada carácter.

Además de ampliar al máximo posible el tamaño de las poblaciones analizadas, también es preciso aumentar el número de marcadores moleculares. Solamente con un número suficiente de marcadores estrechamente ligados al gen responsable del fenotipo podremos acotar al máximo el intervalo donde se localice el QTL correspondiente. Los marcadores deben ser lo más polimórficos que sea posible para poder encontrar marcadores muy próximos a cualquier gen. El tipo de material utilizado en muchos programas de mejora, principalmente en agricultura y ganadería, está en contra de este precepto porque dichas poblaciones han sufrido ya un severo proceso de selección genética que ha reducido también su variabilidad. Con estas condiciones, actualmente la decisión está entre los SSR y los SNP. Los SSR son más variables pero los SNP son más abundantes y fáciles de analizar.

Como vemos, la MAS presenta diferentes problemas para identificar los QTL, localizarlos con precisión o cuantificar su efecto. Por el momento, para la mayor parte de los caracteres complejos, la MAS solo ha sido capaz de identificar una parte mínima de toda la variación observada para cada carácter. Parece evidente que una buena parte de los genes importantes escapan a la manipulación con las herramientas y diseños experimentales de la selección genética.

7.1. La selección genómica es una alternativa a la selección genética (MAS)

La identificación de todos los genes (y sus funciones) que gobiernan un carácter pueden tener dos tipos de aproximaciones: la de genética directa o la de genética inversa. En muchos casos, con caracteres complejos, el reto es prácticamente inalcanzable con las herramientas tecnológicas y de análisis actuales. La gran aportación de la selección genómica estriba en que ofrece nuevas soluciones para seleccionar los mejores individuos en un programa de mejora respecto de uno o varios caracteres. La selección genómica puede parecer una variante de la selección genética pero el planteamiento estratégico es bien distinto. La MAS pretende identificar nuevos QTL, con la mayor precisión posible, para localizar en esas regiones genes responsables de la variación de caracteres que puedan, finalmente, ser mejorados. En el caso de la selección genómica, con mayor pragmatismo, no se pretende identificar genes sino utilizar los marcadores moleculares para predecir valores de mejora sin tener que medir los caracteres. No cabe duda de que muy posiblemente una vez localizadas secuencias asociadas a la variación se pueden encontrar genes regulando los caracteres, pero esa parte de la investigación no atañe al mejorador.

Los trabajos de simulación de QTL con diferentes condiciones (número de QTL, heredabilidades, efectos, etc.) indican que la selección genómica proporciona mayor precisión en los programas de mejora que la información única del fenotipado. En cualquier caso, aunque no se abandone el análisis fenotípico, una selección previa con marcadores moleculares siempre acelera el proceso de selección. Los trabajos con datos reales confirman que la selección genómica predice mejor que la selección fenotípica pero no tan bien como en las simulaciones. Algunas conclusiones se pueden sacar del análisis de esos trabajos:

- Por lo general se encuentran asociaciones entre marcadores moleculares y caracteres de importancia pero el tamaño de los efectos es habitualmente menor de lo esperado.
- La proporción de variación fenotípica explicada por la variación genética observada es función del número de marcadores estudiados. Es decir, cuanto mayor es la proporción del genoma analizado más piezas del puzzle que completa un carácter son encontradas.
- La acumulación de datos de secuencias y el abaratamiento de los costes de genotipado son claves para generalizar el uso de la estrategia de asociación en los programas de mejora.

7.2. Beneficios de la selección genómica

La selección genómica pretende una selección más temprana, más barata y con mayor intensidad que la fenotípica. A continuación discutiremos brevemente algunos aspectos sobre estas ventajas:

- Selección temprana: la selección genómica tiene potencial para alterar radicalmente la estructura de los programas de mejora reduciendo los intervalos entre generaciones al seleccionar los individuos mucho más jóvenes. Incluso no es descabellado pensar que en breve se puedan seleccionar gametos, en lugar individuos, para ser los "padres" de la siguiente generación. Por el momento ya se consiguen acortar mucho las generaciones al seleccionar los individuos en fases muy tempranas del desarrollo.
- Coste: el avance tecnológico está reduciendo los costes a la vez que aumenta la cobertura del genoma, facilitando el uso de las herramientas moleculares en los programas de mejora. La selección genómica implica un ahorro por varias causas. En primer lugar se ahorra al no tener que fenotipar individuos, aunque la exactitud de la estima genómica del valor de un individuo no sea muy superior a la fenotípica. En segundo lugar se ahorra porque se reducen el número de individuos por generación y la duración de esas generaciones. Debe tenerse en cuenta que el coste de los test fenotípicos no solo debe medirse como el gasto en evaluar el carácter si no también, y principalmente, como el gasto del mantenimiento de los individuos y del establecimiento de réplicas.
- Incremento en la intensidad de la selección: incluso con los sistemas más moderados de genotipado de alta resolución se aumenta considerablemente la capacidad para analizar más individuos que con los test de campo. Es posible aumentar la intensidad de selección, sobre todo en el caso de selección de varios caracteres simultáneamente.

Además de estas ventajas, la selección con marcadores moleculares permite una mayor flexibilidad para la evaluación y selección de los individuos ya que el ADN puede analizarse sin las premuras de tiempo que en muchos casos requiere los análisis morfológicos. No hay tampoco que olvidar que los datos obtenidos son muy valiosos para identificar los genes que controlan los caracteres. La caracterización de la variación última responsable de la variación de un carácter puede aportar información clave, junto a muchos otros datos, sobre la función de los genes y su papel en la regulación de los caracteres.

7.3. Conclusiones por áreas

Las diferentes áreas de la mejora, debido a muchos factores particulares, se encuentran en situaciones bien distintas en cuanto al desarrollo y uso de las herramientas genotípicas y genómicas. Ahora repasaremos algunos datos por áreas, dónde estamos, dónde es previsible llegar y qué es preciso mejorar.

Ganadería

No está muy claro cuántos genes y en qué medida se están utilizando por la industria en programas de MAS. Varias compañías están desarrollando y comercializando test basados en un pequeño número de marcadores. El impacto de la selección asistida por marcadores ha sido modesto, porque los QTL que exceden el nivel de significación solo explican una fracción pequeña de la varianza.

La secuenciación de los genomas en varias especies domésticas ha permitido, entre otros logros, el descubrimiento de un gran número de SNP. Existen ya *microarrays* de SNP en vacuno, perros, ovejas, cerdos, caballos y gallinas que pueden genotiparse fácilmente usando la misma tecnología que en humanos y a un precio razonable (100-200 euros). Al aumentar la densidad de SNP podremos, en breve, asegurar que para todos los genes de un carácter de interés habrá algún SNP en desequilibrio de ligamiento con cada uno de ellos.

El enfoque de la selección genómica ha despertado mucho interés en numerosas empresas que están re-diseñando sus programas de mejora. La idea es duplicar la tasa de progreso genético por año. En vacuno de leche es donde quizás el impacto de la selección genómica ha sido más inmediato. El proceso de mejora clásico, aunque eficiente, es muy costoso tanto en tiempo (unos 5 años) como en dinero. El intervalo generacional se reducirá a unos 2 años aproximadamente con lo que la tasa de progreso genético se duplicará. La selección genómica en vacuno de leche es ya un hecho. El primer catálogo oficial de predicciones genómicas de toros jóvenes se presentó en enero de 2009. Las dudas que existían respecto a si los ganaderos utilizarían semen de toros solo evaluados genómicamente, se ha revelado injustificada pues las ventas de semen de estos toros se ha cuadruplicado frente a los de toros probados fenotípicamente.

Una consecuencia no deseable será, probablemente, que debido a la alta intensidad de selección el mantenimiento de la variabilidad genética puede verse comprometido así como la posibilidad de que puedan introducirse en los rebaños genes que puedan ser favorables para la producción de leche pero negativos para otros caracteres como, por ejemplo, la fertilidad.

En definitiva, la selección genómica es la herramienta de mayor impacto sobre la eficiencia de la selección animal que el mejoramiento genético ha producido en muchos años. Existen varias incógnitas relacionadas con la implementación de la selección genómica que irán clarificando en los próximos años: la elección de las poblaciones de referencia, los modelos matemáticos de análisis de datos (modelos bayesianos, métodos de *matching learning*, redes neuronales, redes bayesianas, etc.) o la mejora en las medidas fenotípicas para caracteres difíciles de medir.

La tasa de progreso genético podría aumentar todavía más si la selección genómica se combinara con las tecnologías de manipulación de gametos y embriones. Estas últimas permitirían reducir todavía más los intervalos generacionales y aumentar el número de ciclos de selección por unidad de tiempo. Los oocitos se recolectarían del útero de las terneras, se madurarían *in vitro*, se fertilizarían y los embriones se seleccionarían genómicamente. Posteriormente, se implantarían en vacas receptoras. El proceso podría repetirse recurrentemente. Incluso es factible pensar que los embriones, en vez de re-implantarse se cultivarían *in vitro*, se induciría la meiosis y se formaría una nueva generación de embriones. Así podría practicarse varias generaciones de selección genómica en el laboratorio con un intervalo generacional de unos quince días. Evidentemente, la selección genómica en animales está un paso delante del resto de áreas.

Agricultura

Los marcadores moleculares llevan utilizándose desde mediados de los 90 en las plantas cultivadas. Las utilidades, asociadas a la MAS o en etapas previas y posteriores a la mejora, pueden resumirse en los siguientes apartados:

- Determinación de pureza en semilla de híbridos F1: con marcadores moleculares del tipo microsatélite las compañías están caracterizando el grado de pureza de sus semillas híbridas.
- Control de identidad de lotes de semilla o de planta: para evitar poner en el mercado lotes erróneos o mezclas se realizan test con un conjunto de marcadores (SSR o SNP) para poder separar todas las variedades con las que exista riesgo de mezcla o confusión.
- Protección de derechos de obtentor: además de proteger los derechos en base a la normativa de la UPOV, los obtentores defienden sus obtenciones mediante el uso de marcadores moleculares (generalmente SSR).
- Selección de todo el genoma en programas de retrocruzamiento; El uso de marcadores supone un ahorro de 3-6 generaciones con respecto al procedimiento convencional. Esta es una de las aplicaciones más usadas por las empresas, particularmente para las especies que se comercializan como líneas puras o híbridos F1. De nuevo, lo más habitual es el uso de SSR o SNP.
- MAS propiamente dicha: principalmente en el caso de resistencias a plagas o enfermedades, u otros caracteres controlados por un único gen, se utilizan habitualmente los marcadores moleculares.

El estado actual de la selección genética y genómica varía también mucho según hablemos de cultivos herbáceos o leñosos. En herbáceos la MAS se practica en pro-

gramas públicos y privados para *loci* donde se encuentran genes/QTL mayores, muchos de los cuales determinan la resistencia a diversas enfermedades y plagas, juntamente con algunos genes de calidad o resistencias a estrés abiótico. En leñosas el nivel de aplicación de marcadores en mejora está en alza pero es todavía menor, aunque varios genes son usados habitualmente en MAS. Otra aplicación usual es la caracterización de parentales y su selección basada en el genotipo de marcadores para evaluar distancia genética, nivel de segregación o presencia de determinados alelos asociados con caracteres de interés.

En el apartado de la selección genómica, la agricultura ha ido a remolque de los desarrollos en ganadería. Hay diversos trabajos que demuestran la posibilidad de identificar secuencias asociadas a caracteres de interés pero se desconoce la aplicación de estos desarrollos por parte de las compañías. Por el momento no hay resultados positivos con caracteres complejos.

Es preciso un esfuerzo por diseñar poblaciones de mejora que consigan sacar el máximo provecho de esta nueva estrategia de mejora. Las poblaciones con múltiples parentales están permitiendo el acceso a una variabilidad mayor en plantas donde estudiar caracteres de herencia compleja con una inusitada precisión. Se facilita la localización de la mayor parte de los genes implicados en un carácter, así como de sus alelos, sus interacciones con otros genes y el ambiente, así como una vía hacia la identificación de las secuencias responsables de la variabilidad encontrada detectando un elevado número de QTL distribuidos por todo el genoma, con efectos menores cada uno de ellos.

Acuicultura

La mejora en acuicultura está en estado embrionario. Hay programas de selección familiar para poco más de 30 especies, debido a la complejidad de su desarrollo por las necesidades de espacio y al coste asociado a la trazabilidad genealógica con marcadores moleculares. Sin embargo, la buena respuesta a la selección y la creciente competitividad en este campo de la producción, están determinando una demanda cada vez mayor de la tecnología genética desde las empresas del sector. Los marcadores moleculares se están utilizando principalmente en tareas muy concretas en los programas de mejora. Un ejemplo claro es el la detección precoz del sexo en rodaballo, que ha sido descrito en este informe, en el que mediante una PCR se amplifica un marcador para inferir el sexo de los individuos. Actualmente, esta herramienta está siendo aplicada en planes de selección genética con empresas del sector.

Aunque de forma más lenta que en otras especies domésticas, la acuicultura está incorporando las herramientas genómicas. Hay mapas genéticos, mapas de expresión y genes candidatos. El mapeo y co-localización de genes candidato y QTL de caracteres de interés comienza a ser una tarea habitual en acuicultura. Más recientemente, se están incrementando los recursos genómicos. Además de disponer de secuencias de genomas de especies modelo como el pez cebra o el pez medaka, en la actualidad se están secuenciando los genomas de varias especies de acuicultura, la lubina y el rodaballo entre otras. Por otro lado, la cantidad de ESTs en bases de datos de peces ha crecido de forma muy notable (más de 1 millón) y actualmente se dispone de plataformas de *microarrays*. En cualquier caso no hay por el momento suficientes SNP para realizar estrategias de análisis de asociación genómica global.

Por lo tanto, en acuicultura prácticamente todo está por hacer en cuanto a la selección genética y genómica. Se requieren diseños específicos, poblaciones de referencia y mucha información de resecuenciación para la producción de chips de SNP. El progreso depende en gran medida de las inversiones de compañías y organismos públicos.

Silvicultura

En árboles se han localizado muchos QTL relacionados con la adaptación y el crecimiento de los árboles y con características madereras. Entre otros, se han localizado algunos en relación con las propiedades físicas de la madera que afectan a la dureza o con las propiedades de la pasta de celulosa para hacer papel. Además, se han identificado QTL para resistencia a enfermedades, tiempo de floración, resistencia al frío, composición en aceites de las hojas, y muchos más. La aplicación de estos conocimientos en MAS parece por el momento limitada para QTL muy concretos, como son las resistencias a enfermedades y plagas.

Los estudios globales de asociación han identificado marcadores del tipo SNP asociados a caracteres de interés. Se está integrando la selección genética y genómica en los programas de mejora forestal en diferentes facetas: selección de los árboles "plus", selección sobre las poblaciones mejoradas y en las poblaciones productoras. Los estudios realizados por el momento son escasos pero evidencian el futuro prometedor de esta estrategia. A modo de resumen, repasaremos la situación actual en silvicultura:

- Está muy extendido el empleo de marcadores moleculares para la selección clonal y para la identificación de QTL de interés.
- La mayor parte de programas de mejora emplean los marcadores moleculares donde son claramente más ventajosos que medir el carácter de interés: caracteres complicados de medir y que requieren llegar a la madurez de los individuos.
- La apuesta de diversas compañías por la secuenciación y las herramientas genómicas disponibles demuestran que la selección genómica es una realidad.
- No hay ejemplos de la identificación y utilización en mejora de varios genes controlando parte de la variación de un carácter cuantitativo.

En el caso de la mejora forestal aún se conserva la mayor parte de la variabilidad genética presente en las especies silvestres. Por lo tanto, la selección genómica encuentra una situación muy favorable en la que identificar numerosos genes de interés.

7.4. Conclusiones finales

En la actualidad se emplean habitualmente las herramientas moleculares de selección genética para realizar diversas tareas en relación con la mejora de especies. Entre otras, podemos destacar las siguientes:

- Selección de parentales de poblaciones que abarquen la mayor diversidad posible.
- MAS para caracteres simples, como por ejemplo genes de resistencia a plagas o enfermedades.

- Piramidación de genes dentro de un programa, de nuevo principalmente para genes de resistencia.
- Tareas posteriores a la mejora propiamente dicha, como puede ser la certificación genética del producto final o la protección de los derechos de los obtentores.
- Otras tareas, como la determinación del sexo en peces.

Para alcanzar un mayor grado de utilización de la selección genómica es preciso alcanzar mejoras científico-tecnológicas en algunos apartados concretos, entre otros:

- Los estudios genómicos globales no deben apartar la mirada de los análisis de caracteres cuantitativos concretos. Sigue siendo muy útil la disección precisa de los QTL que conforman cada carácter.
- La información genómica estructural y funcional será imprescindible para la manipulación completa de los caracteres complejos. Los avances en genómica funcional, proteómica o metabolómica deben cooperar necesariamente en el estudio de esos caracteres.
- Las tecnologías que son la base del desarrollo de la genómica deben conseguir aún un abaratamiento de los costes.
- Se requiere un desarrollo de las bases de datos para que sean más potentes, globales y públicas.

En consonancia con todo lo dicho hasta el momento, se necesita un mayor grado de colaboración en grupos multidisciplinarios donde fluya la comunicación entre la biología molecular y la mejora. Los estudios de asociación suelen realizarse en universidades o centros de investigación mientras que los programas de mejora se realizan en estaciones experimentales o compañías privadas. En muchos casos se producen dificultades en la transferencia de información (marcadores, poblaciones, etc.) si no existe una estrecha colaboración en el trabajo. Por lo general, los científicos sienten el trabajo finalizado cuando publican sus resultados en materiales de laboratorio y no tienen demasiado interés en el proceso posterior que implica el programa de mejora. Por contra, las compañías privadas no tienen mayor interés en publicar sus resultados. Además, existe el problema de comunicación entre los genéticos y biólogos moleculares con los mejoradores. Las tecnologías de marcadores moleculares y la disección de los caracteres cuantitativos son conceptos que manejan los primeros y que en muchos casos se escapan a la formación de los segundos. La misma idea pero en sentido contrario es aplicable a las estrategias de mejora. En este contexto se hace indispensable la creación de grupos interdisciplinarios que permitan afrontar los programas de forma global, y no como compartimentos estancos.

Los resultados descritos en este informe apuntan a que la selección genómica va a marcar un hito en los programas de mejora genética, modificando sus pautas y prácticas y mejorando su eficacia. La selección genómica no trata de dilucidar que genes participan o qué papel tienen en la expresión de un carácter pero es más eficiente señalando los mejores individuos de una población. Aún estamos en una etapa de continuo crecimiento y desarrollo científico. A pesar de ello, la selección genómica está comenzando a dar resultados que sobrepasan las expectativas en muchos casos pues permite identificar con precisión y rapidez genes de interés. Es una herramienta que debe integrarse en la mejora clásica y no sustituirla. En la actualidad, la mejora de diferentes caracteres cuantitativos es primordial en muchos programas de mejora. Por el momento la selección genómica solo ha tenido éxito para caracteres controlados por uno o pocos genes. Por lo tanto, el estudio y manipulación de esos caracteres complejos de interés económico constituyen el principal reto de cara al futuro.

8. Referencias bibliográficas

8.1. Introducción

- Ashikari, M.; Sakakibara, H.; Lin, S.; Yamamoto, T.; Takashi, T.; Nishimura, A.; Angeles, E. R.; Qian, Q.; Kitano, H. and Matsuoka, M. (2005). Cytokinin Oxidase Regulates Rice Grain Production. *Science*, 309:741-745.
- Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science*, 48 (5):1649-1664.
- Butcher, P. and Southerton, S. (2007). Marker-assisted selection in forestry species. En: Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish p. 283-305. FAO.
- Castro, A. J.; Chen, X. M.; Hayes, P. M. and Johnston, M. (2003). Pyramiding quantitative trait locus (QTL) alleles determining resistance to barley stripe rust: Effects on resistance at the seedling stage. *Crop Science*, 43:651-659.
- Chistiakov, D. A.; Hellems, B. Volckaert. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, 255 (1-4):1-29.
- Collard, B. C. Y. and Mackill, D. J. (2007). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 17:1-16.
- Dekkers, J. C. (2004). Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J Anim Sci.*, 82 E-Suppl:E313-328.
- Goddard, M. E. and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet*, 10:381-391.
- Grattapaglia and Marcos, D. V. Resende (2010). Genomic selection in forest tree breeding. Online first <http://dx.doi.org/10.1007/s11295-010-0328-4>.
- Helguera, M.; Khan, I. A.; Kolmer, J.; Lijavetzky, D.; Zhong-qi, L. and Dubcovsky, J. (2003). PCR assays for the Lr37-Yr17-Sr38 cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. *Crop Science*, 43:1839-1847.
- Jannink, J. L.; Lorenz, A. J. and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomic Proteomic*, 9:166-177.
- Meuwissen, T. and Goddard, M. (2010). Accurate prediction of genetic values for complex traits by whole genome resequencing. *Genetics*, 185:623-631.
- Misztal, I. (2006). Challenges of application of marker assisted selection - a review. *Animal Science Papers and Reports*, 24 (1):5-10.
- Mullis, K. B.; Faloona, F.; Scharf, S. J.; Saiki, R. K.; Horn, G. T., and Erlich, H.A. (1986). Specific enzymatic amplification of ADN *in vitro*: The polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.*, 51:263-273.

- Neimann-Sorenson, A. and Robertson, A. (1961). The association between blood groups and several production characters in three Danish cattle breeds. *Acta Agric. Scand.*, 11:163-196.
- Olson, M.; Hood, L.; Cantor, D. and Botstein, D. (1989). A common language for physical mapping of the human genome. *Science*, 245:1434-1435
- Paterson, A. H.; Lander, E. S.; Hewitt, J. D.; Peterson, S.; Lincoln, S. E. and Tanksley, S. D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragmented length polymorphisms. *Nature*, 335:721-726.
- Sellner, E. M.; Kim, J. W.; McClure, M. C., *et al.* (2007). Applications of genomic information in livestock. *Journal of Animal Science*, 85:3148-3158.
- Sonesson, A. K. and Meuwissen, T. H. (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genet Sel Evol*, 41:37.
- Tautz, D. (1989). Hypervariability of simple sequence repeats as a general source for polymorphic ADN markers. *Nucleic Acids Res*, 17:6463-6471.
- Vos, P.; Hogers, R.; Bleeker, M., *et al.* (1995). AFLP - A new technique for DNA fingerprinting. *Nucleic acids research*. Volume: 23. Issue, 21:4407-4414.
- Weber, J. L. and May, P. E. (1989). Abundant class of human ADN polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet*, 44: 388-396.
- Williams, J. G .K.; Kubelik, A. R.; Livak, K. J.; Rafalski, J. A. and Tingey, S. V. *et al.* (1990). ADN polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18:6531-6535.
- Young N. A. (1999). cautiously optimistic vision for marker-assisted breeding. *Molecular Breeding*, 5:505-510.

8.2. Descripción de tecnologías. Técnicas de genotipado de SNP

- 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. *Nature*, 467 (7319):1061-73.
- Albert, T. J.; Molla, M. N.; Muzny, D. M. *et al.* (2007). Direct selection of human genomic loci by microarray hybridization. *Nat Methods*, 4:903-905.
- Borza, T.; Higgins, B.; Simpson, G. and Bowman, S. (2010). Integrating the markers Pan I and haemoglobin with the genetic linkage map of Atlantic cod (*Gadus morhua*). *BMC Res Notes*, 3:261.
- Campbell, P. J.; Stephens, P. J.; Pleasance, E. D. *et al.* (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet*, 40:722-729.
- Cánovas, A.; Rincon. G.; Islas-Trejo, A. *et al.* (2010). SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mamm Genome*. DOI 10.1007/s00335-010-9297-z .
- Fan, J. B.; Oliphant, A.; Shen, R. *et al* (2003). Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol*, 68:69-78.
- Farrar, J. S.; Reed, G. H. and Wittwer, C. T. (2009). High resolution melting curve analysis for molecular diagnostics. In: Patrinos GP, Anson W, editors. *Molecular Diagnostics*, 2nd Ed. Burlington: Elsevier.
- Fu, Y.; Springer, N. M.; Gerhardt *et al.* (2010). Repeat subtraction-mediated sequence capture from a complex genome. *Plant J.*, 62(5):898-909.
- Garvey, C. (2010). A decade and genome of change. *Genome Biol.* 2010;11(5):120.
- Gelfand, D. H.; Holland, P. M.; Saiki, R. K. and Watson, R. M. (1993). Homogeneous assay system using the nuclease activity of a nucleic acid polymerase. US Patent., 5.210.015.
- Goddard, M. E., i. and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet*, 10:381-391.
- Gunderson, K. L.; Steemers, F. J.; Lee, G.; Mendoza, L. G. and Chee, M. S. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Nat. Genet.*, 37:549-54.
- Hayes, B. J.; Bowman, P. J.; Chamberlain, A. J. and Goddard, M. E. (2009). Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci*, 92:433-443.
- Hayes, B. and Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome*, 53(11):876-83.
- Harhay, G. P.; Smith, T. P.; Alexander, L. J. *et al.* (2010) An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. *Genome Biol.* 20;11(10):R102.
- Heffner, E. L.; Sorrells, M. E. and Jannink, J.-L. (2009) Genomic selection for crop improvement. *Crop Sci*, 49:1-12.

- Hodges, E.; Xuan, Z.; Balijs, V. *et al.* (2007). Genome-wide in situ exon capture for selective resequencing. *Nat Genet*, 39:1522–1527.
- Huang, X.; Wei, X.; Sang, T. *et al.* (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genet*, 42:961–967.
- Jannink, J. L.; Lorenz, A. J. and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomic Proteomic*, 9:166–177.
- Kennedy, G. C.; Matsuzaki, H.; Dong, S. *et al.* (2003). Large-scale genotyping of complex ADN. *Nat. Biotechnol*, 21:1233–37
- Kutuyavin, I. V.; Afonina, I.A.; Mills, A. *et al.* (2000). 3'-Minor groove binder-ADN probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res.*, 28:655–661.
- Lai, J.; Li, R.; Xu, X. *et al.* (2010). Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genet.*, 42:1027–1030.
- Lee SH, van der; Werf, J. H.; Hayes, B. J.; Goddard, M. E. and Visscher, P. M. (2008). Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet*, 4: e1000231.
- Li, Y. D.; Chu, Z. Z.; Liu, X. G. *et al.* (2010). A Cost-effective High-resolution Melting Approach using the EvaGreen Dye for ADN Polymorphism Detection and Genotyping in Plants. *J Integr Plant Biol.*, 52(12):1036-42.
- Liew, M.; Pryor, R.; Palais, R. *et al.* (2004). Genotyping of single-nucleotide polymorphisms by HRM of small amplicons. *Clin Chem*, 50:1156–64.
- Mamanova, L.; Coffey, A.J.; Scott, C. E. *et al.* (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods*, 7(2):111-8.
- Mardis, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet*, 24:133-141.
- Matukumalli, L. K.; Lawley, C. T.; Schnabel, R. D. *et al.* (2009) Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* 4:e5350.
- Meuwissen, THE; Hayes, B. J. and Goddard, M. E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157:1819–1829.
- Muleo, R.; Colao, M. C.; Miano, D. *et al.* (2009). Mutation scanning and genotyping by high-resolution ADN melting analysis in olive germplasm. *Genome*, 52(3):252-60.
- Myles, S.; Chia, J. M.; Hurwitz, B. *et al.* (2010). Rapid genomic characterization of the genus vitis. *PLoS One*. 5(1):e8219.
- Nielsen, H. M.; Sonesson, A. K.; Yazdi, H. and Meuwissen THE (2009). Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. *Aquaculture*, 289:259–264.
- Nielsen, H. M.; Sonesson, A. K.; Meuwissen, T. H. (2010) Optimum contribution selection using traditional BLUP and genomic breeding values in aquaculture breeding schemes. *J Anim Sci*. Oct 29 doi:10.2527/jas.2009-2731.
- Porreca GJ, Zhang K, Li JB *et al.* (2007). Multiplex amplification of large sets of human exons. *Nat Methods*, 4:931–936.

- Ramos, A. M.; Crooijmans, R. P.; Affara, N. A. *et al.* (2009). Design of a high density SNP genotyping assay in the pig using SNP identified and characterized by next generation sequencing technology. *PLoS One.*, 4(8):e6524.
- Ririe, K. M.; Rasmussen, R. P.; Wittwer, C. T. (1997). Product differentiation by analysis of ADN melting curves during the polymerase chain reaction. *Anal Biochem*, 245:154-160.
- Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet*, 123: 218-223.
- Shendure, J. and H. Ji. (2008). Next-generation ADN sequencing. *Nature Biotech*, 26:1135-1145.
- Sonesson, A. K. and Meuwissen, T. H. (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genet Sel Evol*, 41:37
- Steemers, F. J.; Chang, W.; Lee, G. *et al.* (2006). Whole-genome genotyping with the single-base extension assay. *Nat. Methods*, 3:31-33.
- The Bovine HapMap Consortium. (2009) Genome wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 324(5926):528-532.
- The Bovine Genome Sequencing and Analysis Consortium. (2009) The genome sequence of taurine cattle: A window to ruminant biology and evolution. *Science*, 324(5926):522-8.
- Tewhey, R.; Warner, J. B.; Nakano, M. *et al.* (2009). PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol*, 27(11):1025-31.
- Thelwell, N.; Millington, S.; Solinas, A. *et al.* (2000). Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Res*, 28:3752-61.
- Turner, E. H.; Ng, S. B.; Nickerson, D. A. and Shendure, J. (2009). Methods for genomic partitioning. *Annu Rev Genomics Hum Genet*, 10:263-84.
- Tyagi, S.; Bratu, D. P.; Kramer, F. R. (1998). Multicolor molecular beacons for allele discrimination, *Nat. Biotechnol.*, 16:49-53.
- Van Tassell, C. P.; Smith, T. P. L.; Matukumalli, L. K. *et al.* (2008). Simultaneous SNP discovery and allele frequency estimation by high throughput sequencing of reduced representation genomic libraries. *Nature Methods*, 5:247-52.
- Wang, J.; Lin, M.; Crenshaw, A. *et al.* (2009). High-throughput single nucleotide polymorphism genotyping using nanofluidic Dynamic Arrays. *BMC Genomics*, 10:561.
- Wheeler, D. A.; Srinivasan, M.; Egholm, M. *et al.* (2008). The complete genome of an individual by massively parallel ADN sequencing. *Nature*, 452:872-876.
- Wittwer, C. T.; Reed, G. H.; Gundry, C. N. *et al.* (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem*, 49:853-60.
- Wittwer, C. T. (2009). High-resolution ADN melting analysis: advancements and limitations. *Hum Mutat*, 30:857-859.

8.3. Aplicación de la selección genética y genómica en ganadería

- Henderson, C. R. (1948). *Estimation of general, specific and maternal abilities in crosses among inbred lines of swine*. Iowa State University Press, Ames.
- FABRE TECHNOLOGY PLATFORM (2006). *Sustainable Farm Animal Breeding and Reproduction-A vision for 2025*.
- Harvenstein, G.; Ferket, P. R.; Scheideler, S. E. and Rives, D. V. (1994). Carcass composition and yield of 1991 vs 1957 broilers when fed "typical" 1957 and 1991 broiler diets. *Poultry Science*, 73:1795-1804.
- Flint, A. P. F. and Woolliams, J. A. (2008). Precision animal breeding. *Philosophical Transactions of the Royal Society B*, 363:573-590.
- Lynch, M.; Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer, Suntherland.
- Kim, K. S.; Larsen, N; Short, T; Plastow, G. and Rothschild, M. F. (2000). A missense variant of the porcine melanocortin-4 receptor (*MC4R*) gene is associated with fatness, growth, and feed intake traits. *Mammalian Genome*, 1:131-135.
- Georges, M. (2007). Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Annual Review of Genomics and Human Genetics*, 8:131-62.
- Haley, C. and de Koning, D. J. (2006). Genetical genomics in livestock: potentials and pitfalls. *Animal Genetics*, 37 Suppl 1:10-12.
- Goddard M. E.; Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10: 381-391.
- Rothschild, M.; Jacobson, C.; Vaske, D.; Tuggle, C.; Wang, L.; Short, T.; Eckardt, G.; Sasaki, S.; Vincent, A.; McLaren, D.; Southwood, O.; van der Steen, H.; Mileham, A. and Plastow, G. (1996). The estrogen receptor locus is associated with a major gene influencing litter size in pigs. *Proc. Natl. Acad. Sci. USA*, 93:201-205.
- Alfonso, L. (2005). Use of meta-analysis to combine candidate gene association studies: application to study the relationship between the *ESR PvuII* polymorphism and sow litter size. *Genetics, Selection, Evolution*, 37:417-435.
- Georges, M.; Nielsen, D.; Mackinnon, M.; Mishra, A.; Okimoto, R.; Pasquino, A. T.; Sargeant, L. S.; Soerensen, A.; Steele, M. R.; Zhao, X.; Womack, J. E. and Hoeschele, I. (1995). Mapping QTL controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*, 139:907-920.
- de Koning, D-J. (2006). Conflicting candidates for cattle QTL. *Trends in Genetics*, 22:301-305.
- Van Eenennaam, A. L.; Li, J.; Thallman, R. M.; Quaas, R. L.; Dikeman, M. E.; Gill, C. A.; Franke, D. E.; Thomas, M. G. (2007). Validation of commercial ADN tests for quantitative beef quality traits. *Journal of Animal Science*, 85:891-900.

- Toro, M. A. (2010). Future trends in Animal Breeding due to new genetic technologies (in press).
- Meuwissen, T. H. E.; Hayes, B. J. and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157:1819-1829.
- Gianola, D.; de los Campos G. (2008). Inferring genetic values for quantitative traits non-parametrically. *Genetics Research*, 90:525-540.
- González-Recio, O.; Gianola, D.; Long, N.; Weigel, K. A.; Rosa, G. J. M. and Avendaño, S. (2008). Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in broilers. *Genetics*, 178:2305-2313.

8.4. Aplicación de la selección genética y genómica en agricultura

- Aranzana, M. J.; El Kadri, A.; Howad, W. and Arús, P. (2010). Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC Genetics*, 11:69.
- Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the past 20 years. *Crop Sci*, 48:1649-1664.
- Buckler, E. S.; Holland, J. B.; Bradbury, P. J.; Acharya, C. B.; Brown, P. J.; Browne, C.; Ersoz, E.; Flint-Garcia, S.; Garcia, A.; Glaubitz, J. C. *et al.* (2010). The genetic architecture of maize flowering time. *Science*, 325:714-718.
- Caldwell, K. S.; Russell, J.; Langridge, P. and Powell, W. (2006). Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics*, 172:557-567.
- Cavanagh, C. and Morell, M. (2008). Genome resources for bread wheat. In: Appels, R.; Eastwood, R.; Lagudah, E.; Langridge, P.; Mackay, M.; McIntyre, L.; Sharp, P. (eds) Proceedings of 11th International Wheat Genet Symposium, Brisbane Australia, August 24–29, 2008. Sydney University Press. <http://hdl.handle.net/2123/3245>.
- Darvasi, A. and Soller, M. (1995). Advanced intercross lines: an experimental population for fine genetic mapping. *Genetics*, 141:1199-1207.
- Dirlwanger, E.; Graziano, E.; Joobeur, T.; Garriga-Calderé, F.; Cosson, P.; Howad, W. and Arús, P. (2004). Comparative mapping and marker-assisted selection in *Rosaceae* fruit crops. *Proc Nat Acad Sci USA*, 101:9891-9896.
- Eduardo, I.; Arús, P. and Monforte, A. J. (2005). Development of a genomic library of near isogenic lines (NILs) in melon (*Cucumis melo L.*) from the exotic accession PI161375. *Theor Appl Genet*, 112:139-148.
- Goddard, M. E. and Hayes, B. J. (2007). Genomic selection. *J Anim Breed Genet*, 124:323-330.
- Gupta, P. K.; Langridge, P. and Mir, R. R. (2010). Marker-assisted wheat breeding: present status and future possibilities. *Mol Breed*, 26:145-161.
- Kover, P. X.; Valdar, W.; Trakalo, J.; Scarcelli, N.; Ehrenreich, I. M.; Purugganan, M. D.; Durrant, C. and Mott, R. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet* 5:e1000551.
- McMullen, M. D.; Kresovich, S.; Villeda, H. S.; Bradbury, P.; Li, H.; Sun, Q.; Flint-Garcia, S.; Thornsberry, J.; Acharya, C.; Bottoms, C. *et al.* (2009). Genetic properties of the maize nested association mapping population. *Science*, 325:737-740.
- Medina-Filho, H. (1980). Linkage of Aps-1, Mi and other markers on chromosome 6. *Rep Tomato Genet Cooperative*, 30:26–28.
- Michelmore, R. W.; Paran, I. and Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci* 88:1212-1217.
- Pitrat, M. (2006). Gene list for melon. Cucurbit Genetics Cooperative Report 28-29: 142-163.

- Rafalski, J. A. (2010). Association genetics in crop improvement. *Curr Opin Plant Biol*, 13:178-180.
- Ribaut, J.-M., de Vicente, M. C. and Delannay, X. (2010). Molecular breeding in developing countries :challenges and prespectives. *Curr Opin Plant Biol*, 13:213-218.
- Sageret, A. (1826). Considérations sur la production des hybrides, des variantes et des variétés en général, et sur celles de la famille des Cucurbitacées en particulier. *Annales des Sciences Naturelles*, 8:294-314.
- Sterck, L.; Rombauts, S.; Vandepoele, K.; Rouzé, P.; Van de Peer, Y. (2007). How many genes are there in plants(... and why are they there)? *Curr Opin Plant Biol*, 10:199-203.
- Tanksley, S. D. and Orton, T. J. (1983). Isozymes in Plant Genetics and Breeding: Parts A and B. Elsevier Sc. Pub.
- Yu, J. and Buckler, E. S. (2006). Genetic association mapping and genome organization in maize. *Curr Opin Biotech*, 17:155-160.

8.5. Aplicación de la selección genética y genómica en silvicultura

- Brondani, R. P. V.; Williams, E. R.; Brondani, C.; Grattapaglia, D. (2006). A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. *BMC Plant Biology*, 6:20.
- Butcher, P.; Southerton, S. (2007). MAS in Forestry Species. In: Marker-Assisted Selection (MAS) in Crops, Livestock, Forestry and Fish: Current Status and the Way Forward. FAO pp. 283-305.
- Byrne, M.; Murrell, J. C.; Owen, J. V.; Kriedemann, P.; Williams, E. R.; Moran, G. F. (1997a). Identification and mode of action of quantitative trait *loci* affecting seedling height and leaf area in *Eucalyptus nitens*. *Theoretical and Applied Genetics*, 94:674–681.
- Byrne, M.; Murrell, J. C.; Owen, J. V.; Williams, E. R.; Moran, G. F. (1997b). Mapping of quantitative trait *loci* influencing frost tolerance in *Eucalyptus nitens*. *Theoretical and Applied Genetics*, 95:975–979.
- Casasoli, M.; Derory, J.; Morera-Dutrey, C.; Brendel, O.; Porth, I. *et al.* (2006). Comparison of quantitative trait loci for adaptive traits between oak and chestnut based on an expressed sequence tag consensus map. *Genetics*, 172 533–546.
- Cato, S.; Pot, D.; Kumar, S.; Douglas, J.; Gardner, R. and Wilcox, P. (2006). In: *Abstracts of Plant and Animal Genomes XIVth Conference*, San Diego, CA (USA), January, 14-18.
- De Aguiar, M. S.; Ferreira, D. F.; Aguiar, A. M.; Bison, O.; Rezende, GDSP and Grattapaglia, D. 2007. Potencial de híbridos entre clones-elite de eucalipto por meio de marcadores microssatélites. *Pesquisa Agropecuaria Brasileira*, 42: 1007–1012.
- Freeman, J.; Whittock, S.; Potts, B. and Vaillancourt, R. (2009). QTL influencing growth and wood properties in *Eucalyptus globulus*. *Tree Genet Genom*, 5:713–722.
- Gaiotto, F. A.; Bramucci, M. and Grattapaglia, D. (1997). Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* s.t. Blake with dominant RAPD and AFLP markers. *Theoretical and Applied Genetics*, 95:842–849.
- Grattapaglia, D.; Bertolucci, F. L. G.; Penchel, R. and Sederoff, R. (1996). Genetic mapping of quantitative trait *loci* controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics*, 144:1205–1214.
- Grattapaglia, D.; Bertolucci, F. L. G. and Sederoff, R. (1995). Genetic mapping of QTL controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Theoretical and Applied Genetics*, 90: 933–947.
- Grattapaglia, D. and Kirst; M. *Eucalyptus* applied genomics: from gene sequences to breeding tools. *New Phytologist*. 2008;179:911–929.
- Grattapaglia, D. (2008). Genomics of *Eucalyptus*, a global tree for energy, paper, and wood. In: MooreP, MingR, eds. *Genomics of tropical crop plants*. New York, NY, USA: Springer, 259–298.

- Henery, M.L.; Moran, G. F.; Wallis, I. R. and Foley, W. J. (2007). Identification of quantitative trait loci influencing foliar concentrations of terpenes and formylated phloroglucinol compounds in *Eucalyptus nitens*. *New Phytologist*, 176:82–95.
- Ikemori, Y. K.; Penchel, R. M. and Bertolucci, F. L. G. (1994). Integrating biotechnology into *Eucalyptus* breeding, In: *Proceedings of the International Symposium on Wood Biotechnology*. Tokyo, Japan: TAPPI, Japan Wood Research Society and Nippon Paper Industries, 79–84.
- Keil, M. and Griffin, A. R. (1994). Use of random amplified polymorphic ADN (RAPD) markers in the discrimination and verification of genotypes in *Eucalyptus*. *Theoretical and Applied Genetics*, 89:442–450.
- Kirst, M.; Myburg, A. A.; De Leon, J. P.; Kirst, M. E.; Scott, J. and Sederoff, R. (2004). Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cADN microarray data in an interspecific backcross of *Eucalyptus*. *Plant Physiology*, 135:2368–2378.
- Lezar, S.; Myburg, A. A.; Berger, D. K.; Wingfield, M. J. and Wingfield, B. D. (2004). Development and assessment of microarray-based ADN fingerprinting in *Eucalyptus grandis*. *Theoretical and Applied Genetics*, 109:1329–1336.
- Marques, C. M.; Carocha, V. J.; Pereira de Sá, A. R.; Oliveira, M. R.; Pires, A. M.; Sederoff, R. R. and Borralho, N. M. G. (2005). Verification of QTL linked markers for propagation traits in *Eucalyptus*. *Tree Genetics and Genomes*, 1:103–108.
- Marques, C. M.; Vasquez-Kool, J.; Carocha, V. J.; Ferreira, J. G.; O'Malley, D. M.; Liu, B. H. and Sederoff, R. R. (1999). Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*. *Theoretical and Applied Genetics*, 99: 936–946.
- Missiaggia, A. A.; Piacuzzi, A. L. and Grattapaglia, D. (2005). Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. *Tree Genetics and Genomes*, 1:79–84.
- Myburg, A. A.; Potts, B.; Marques, C. M.; Kirst, M.; Gion, J. M.; Grattapaglia, D. and Grima-Pettenati, J. (2007). *Eucalyptus*. In: KoleC, ed. *Genome mapping and molecular breeding in plants*, Vol. 7: Forest trees. New York, NY, USA: Springer, 115–160.
- Ottewell, K. M.; Donnellan, S. C.; Moran, G. F. and Paton, D. C. (2005). Multiplexed microsatellite markers for the genetic analysis of *Eucalyptus leucoxylon*, *Myrtaceae* and their utility for ecological and breeding studies in other *Eucalyptus* species. *Journal of Heredity*, 96:445–451.
- Poke, F. S.; Vaillancourt, R. E.; Potts, B. M. and Reid, J. B. (2005). Genomic research in *Eucalyptus*. *Genetica* 125:79–101.
- Sansaloni, C. P.; Petroli, C. D.; Carling, J.; Hudson, C. J.; Steane, D. A.; Myburg, A. A.; Grattapaglia, D.; Vaillancourt, R. E. and Kilian, A. (2010). A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods*, 30;6:16.
- Shepherd, M.; Chaparro, J. X. and Teasdale, R. (1999). Genetic mapping of monoterpenes composition in an interspecific eucalypt hybrid. *Theoretical and Applied Genetics*, 99:1207–1215.

- Telfer, E.; Nelson, C. D.; Echt, C. and Wilcox, P. (2006). Comparative Mapping in *Pinus radiata* and *P. taeda* Reveals Co-Location of Wood Density-Related QTL. In: Plant & Animal Genome XIV. The International Conference on the Status of Plant and Animal Genome Research, Final Program and Abstracts Guide, p.34. January 14-19, 2006, San Diego, California, USA.
- Thamarus KA, Groom K, Bradley A, Raymond CA, Schimleck LR, Williams ER, Moran GF. 2004. Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of *Eucalyptus globulus*. *Theoretical and Applied Genetics*, 109: 856–864.
- Thumma, B. R., B. A. Matheson, *et al.* (2009). "Identification of a Cis-Acting Regulatory Polymorphism in a Eucalypt COBRA-Like Gene Affecting Cellulose Content." *Genetics*, 183(3): 1153-1164.
- Thumma, B. R., M. R. Nolan, R. Evans y G. F. Moran, 2005. Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus spp.* *Genetics*, 171 1257–1265.
- Thumma, B. R., S. G. Southerton, JC Bell, JV Owen, ML Henery y GF Moran. 2010. "Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens*." *Tree Genetics & Genomes*, 6(2): 305-317.
- Vaillancourt, R. E.; Potts, B. M.; Manson, A.; Eldridge, T. and Reid, J. B. (1995b). Using RAPD's to detect QTL in an interspecific F2 hybrid of *Eucalyptus*. In: PottsBM, BorralhoNMG, ReidJB, CromerRN, TibbitsWN, RaymondCA, eds. Proceedings of the CRC/IUFRO Conference Eucalypt plantations, improving fibre yield and quality. Hobart, Australia: CRC for Temperate Hardwood Forestry, 430–433.
- Verhaegen, D.; Plomion, C.; Gion, J. M.; Poitel, M.; Costa, P. and Kremer, A. (1997). Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers. 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theoretical and Applied Genetics*, 95:597–608.
- Wilcox, Philip L.; Echt, Craig E.; Burdon, Rowland, D. (2007). Gene-assisted selection: applications of association genetics for forest tree breeding. In: *Associated Mapping in Plants*. 211-247.

8.6. Aplicación de la selección genética y genómica en acuicultura

- Andersson, L. and Georges, M. (2004). Domestic-animal genomics: deciphering the genetics of complex traits. *Nature Reviews Genetics*, 5:202-212.
- APROMAR. 2009. *La Acuicultura Marina de Peces en España*. Asociación Empresarial de Productores de Cultivos Marinos.
- Bouza, C. and Martínez, P. (2007). Mapas genéticos en Acuicultura. En: *Genética y Genómica en Acuicultura* (Ed.: J. Espinosa; Coord.: P. Martínez y A. Figueras). Publicaciones Consejo Superior de Investigaciones Científicas. Madrid. ISBN: 978-84-00-08553-7: 309-357.
- Bouza, C.; Hermida, M.; Pardo, B. G.; Fernández, C.; Castro, J.; Fortes, G.; Sánchez, L.; Presa, P.; Pérez, M.; Sanjuán, A.; Comesaña, S.; Álvarez, J. A.; Calaza, M.; Cal, R.; Piferrer, F. and Martínez, P. (2007). A microsatellite genetic map in the turbot (*Scophthalmus maximus*). *Genetics*, 177:2457-2467
- Canario, A. V. M.; Bargelloni, L.; Volckaert, F.; Houston, R. D.; Massault, C.; Guiguen, Y. (2008). Genomics toolbox for farmed fish. *Reviews in Fisheries Science*, 16:3-15.
- Falconer, D. S. and Mackay, T. F. C. (2001). *Introduction to Quantitative Genetics*. Longman. London.
- Gjedrem, T. (2005). *Selection and Breeding Programs in Aquaculture*. Springer, Dordrecht.
- Hulata, G. (2001). Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. *Genetica*, 111:155-173.
- Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics*, 35:303-339.
- Martínez, P. (2005). Aplicaciones de la Genética para la mejora en Acuicultura. *Boletín del Instituto Español de Oceanografía* 21:225-238.
- Martínez, P.; Bouza, C.; Hermida, M.; Fernández, J.; Toro, M. A.; Vera, M.; Pardo, B.; Millán, A.; Fernández, C.; Vilas, R.; Viñas, A.; Sánchez, L.; Felip, A.; Piferrer, F.; Ferreiro, I. and Cabaleiro, S. (2009). Identification of the major sex-determining region of turbot (*Scophthalmus maximus*). *Genetics*, 183:1443-1452.
- Matsuda, M.; Nagahama, Y.; Shinomiya, A.; Sato, T.; Matsuda, C.; Kobayashi, T.; Morrey, C. E.; Shibata, N.; Asakawa, S.; Shimizu, N.; Hori, H.; Hamaguchi, S.; Sakaizumi, M. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, 417: 559-63.
- Millán, A.; Gómez-Tato, A.; Fernández, C.; Pardo, B. G.; Álvarez-Dios, J. A.; Calaza, M.; Bouza, C.; Vázquez, M.; Cabaleiro, S. and Martínez, P. (2010). Design and performance of a turbot (*Scophthalmus maximus*) oligo-microarray based on ESTs from immune tissues. *Marine Biotechnology*, 12: 452-465.
- Pardo, B. G.; Fernández, C.; Millán, A.; Bouza, C.; Vázquez-López, A.; Vera, M.; Álvarez-Dios, J. A.; Calaza, M.; Gómez-Tato, A.; Vázquez, M.; Cabaleiro, S.; Margariños, B.; Lemos, M. L.; Leiro, J. M. and Martínez, P. (2008). Expressed sequence tags (ESTs) from immune tissues of turbot (*Scophthalmus maximus*) challenged with pathogens. *BMC Veterinary Research*, 4: 37.

- Penman, D. J. and Piferrer, F. (2008). Fish Gonadogenesis. Part I: Genetic and Environmental Mechanisms of Sex Determination. *Reviews in Fisheries Science*, 16:16-34.
- Pérez-Enciso, M. and Toro, M. A. (2007). Localización de genes y selección mediante marcadores moleculares. En: *Genética y Genómica en Acuicultura* (Ed.: J. Espinosa; Coord.: P. Martínez y A. Figueras). Publicaciones Consejo Superior de Investigaciones Científicas. Madrid. ISBN: 978-84-00-08553-7. Pp: 359-400.
- Piferrer, F. and Guiguen, Y. (2008). Fish Gonadogenesis. Part II: Molecular Biology and Genomics of Sex Differentiation. *Reviews in Fisheries Science*, 16:35-55.
- Piferrer, F.; Beaumont, A.; Falguiere, J. C.; Flajshans, M.; Haffray, P.; Colombo, L. (2009). Polyploid fish and shellfish: Production, biology and applications to aquaculture for performance improvement and genetic containment. *Aquaculture*, 293: 125-156.

Genoma España



Pedro Teixeira, 8 - planta 2 • 28020 Madrid
Teléfono: 91 449 12 50 • Fax: 91 571 54 89
www.gen-es.org

• • • PATRONOS • • •



MINISTERIO
DE CIENCIA
E INNOVACIÓN

MINISTERIO
DE SANIDAD, POLÍTICA SOCIAL
E IGUALDAD

MINISTERIO
DE INDUSTRIA, TURISMO
Y COMERCIO

MINISTERIO
DE MEDIO AMBIENTE
Y MEDIO RURAL Y MARINO

