

# Marcadores moleculares: una herramienta para explorar la diversidad genética

## 1 Introducción

Los marcadores de ADN son útiles tanto en la investigación básica (p. ej., análisis filogenético y búsqueda de genes útiles) como en la aplicada (p. ej., selección asistida por marcador, pruebas de paternidad y trazabilidad de los alimentos). Esta sección se centra sobre todo en su aplicación a la caracterización de la diversidad de los recursos zoológicos, y en la búsqueda de variantes funcionales de determinados genes. Hay que destacar que el ARN y las proteínas también contienen información clave, y que por tanto merecen un estudio paralelo; su papel en la búsqueda de variantes funcionales se explora asimismo más abajo.

La diversidad entre organismos es consecuencia de las diferencias en las secuencias de ADN y de los efectos ambientales. La variación genética es notable, y cada individuo de una especie, a excepción de los gemelos monocigóticos, posee una secuencia de ADN única. Las variaciones en el ADN son mutaciones resultantes de la sustitución de un solo nucleótido (polimorfismos de un solo nucleótido – SNP), inserción o deleción de fragmentos de ADN de diversas longitudes (desde uno a varios miles de nucleótidos), o duplicación o inversión de fragmentos de ADN. Las variaciones del ADN se clasifican como «neutras» cuando no originan cambios en los caracteres metabólicos o fenotípicos, y por consiguiente no están sometidas a selección positiva, negativa o de reequilibrio; en caso contrario, se denominan «funcionales». Las mutaciones en nucleótidos clave de una secuencia codificadora pueden alterar la composición de

aminoácidos de una proteína, y conducir a nuevas variantes funcionales. Dichas variantes pueden tener una mayor o menor eficiencia metabólica comparadas con el «tipo nativo» original, pueden perder completamente su funcionalidad, o incluso adquirir una función nueva. Las mutaciones en las regiones reguladoras pueden afectar a los niveles y las pautas de expresión génica; por ejemplo, activar o desactivar genes, o sobreexpresar o infraexpresar proteínas en tejidos concretos en distintos estadios del desarrollo o en distintos estados fisiológicos.

Aunque el análisis de tipos únicos de biomoléculas ha resultado extremadamente útil para la comprensión de los fenómenos biológicos, la investigación paralela y a gran escala del ADN, ARN y proteínas ha abierto nuevas perspectivas para la interpretación y el modelado de la complejidad de los organismos vivos. Están surgiendo nuevas disciplinas científicas con el sufijo «-ómica». En dichos campos, los avances recientes en la preparación, identificación y secuenciado del ADN, ARN y proteínas, así como el almacenamiento de datos a gran escala y su análisis, están creando una revolución en nuestros conceptos. Está surgiendo una visión global e integrada de un conjunto completo de moléculas biológicas implicadas en procesos biológicos complejos. La genómica estructural, la transcriptómica y proteómica han ido seguidas de la metabolómica y la interactómica, entre otras, y, a un nivel de complejidad aún mayor, la biología de sistemas (Hood *et al.*, 2004; Recuadro 71).

## PARTE 4

**Recuadro 70**  
**ADN, ARN y proteína**

El ADN (ácido desoxirribonucleico) está organizado en pares de cromosomas, siendo cada uno heredado de uno de los progenitores. Cada gen de un individuo, por tanto, posee dos copias, llamadas alelos, uno en cada cromosoma de un par. En los mamíferos, los genes están diseminados a lo largo de los cromosomas, y separados por secuencias de ADN largas y normalmente repetitivas. Los genes están formados por secuencias codificadoras (exones) separadas por intrones. Estos últimos no llevan información para codificar proteínas, pero en ocasiones desempeñan un papel en la regulación de la expresión génica. Las instrucciones codificadas por los genes se activan mediante dos procesos. El primero es la transcripción (copia) de información genética a otro tipo de ácido nucleico, el ARN (ácido ribonucleico). Tanto los exones como los intrones se transcriben a una molécula de ARN mensajero primario (mARN). A continuación, dicha molécula es modificada, en un proceso que conlleva la eliminación de los intrones, la unión de los exones, y la adición de características únicas a cada extremo del mARN. Se crea así una molécula de ARN madura, que se transporta entonces a estructuras conocidas como ribosomas, localizadas en el citoplasma celular. Los ribosomas están formados por ARN ribosómico (rARN) y proteínas, y proporcionan sitios para el segundo proceso

– traducción de la información genética, previamente copiada al mARN, a un polipéptido (una proteína entera o una de las cadenas de un complejo proteico). La molécula de mARN se lee o traduce a razón de tres nucleótidos (un codón) por vez. La complementariedad entre el codón de mARN y el anti-codón de una molécula de ARN de transferencia (tARN), que lleva el aminoácido correspondiente al ribosoma, garantiza que el polipéptido recién formado contenga la secuencia específica de aminoácidos necesarios.

No todos los genes se traducen en proteínas; algunos expresan su función (como el rARN y tARN implicados en la traducción). Se han descubierto recientemente nuevas funciones del ARN en el proceso de modificación del mARN y en la regulación de la expresión génica (Storz *et al.*, 2005; Aravin y Tuschl, 2005; Wienholds y Plasterk 2005). Por ejemplo, los ARN no codificadores parecen ser elementos clave en diversos procesos reguladores (Bertone *et al.*, 2004; Clop *et al.*, 2006). Así pues, existen tres tipos de moléculas disponibles para el estudio de las características genéticas a nivel celular, tisular y de todo el organismo: el ADN que contiene las instrucciones codificadas; el ARN que transfiere las instrucciones a la «fábrica» celular; y las proteínas que se construyen según las instrucciones y que crean células y organismos funcionales.

La investigación de la complejidad biológica es una nueva frontera que requiere tecnología molecular ultrarrápida, altas velocidades de computación y memoria, nuevos enfoques en el análisis de datos, y la integración de pericia interdisciplinaria (Recuadro 72).

**2 El papel de las tecnologías moleculares en la caracterización**

La información sobre la diversidad genética es esencial para optimizar tanto las estrategias de conservación de los recursos zoogenéticos como

**Recuadro 71**  
**Las nuevas disciplinas científicas, las «-ómicas»**

La genómica cartografía los genes y las variaciones genéticas de distintos individuos y grupos. Nos informa de la traducción de la información genética en funciones metabólicas y caracteres fenotípicos. Revela procesos biológicos y sus interacciones con factores medioambientales. La genómica consiste en combinar un conjunto de tecnologías ultrarrápidas, como la proteómica y la metabolómica, con las técnicas de bioinformática que posibilitan el procesado, análisis e integración de grandes volúmenes de datos.

### Recuadro 72 Avances recientes en biología molecular

Los actuales avances revolucionarios en el campo de la biología molecular con aplicaciones en la cría de ganado y la conservación de la diversidad genética incluyen:

1. establecimiento de la secuencia genómica completa de las especies agropecuarias más importantes;
2. desarrollo de tecnología para medir polimorfismos en loci diseminados por todo el genoma (p. ej., métodos para detectar SNP); y
3. desarrollo de tecnología de microarrays para medir la transcripción génica a gran escala.

La información obtenida mediante el secuenciado del genoma entero (ya conseguido para gallinas y casi terminado para cerdos y bovinos), integrada con la tecnología SNP, acelerará la búsqueda de genes. El cartografiado de los loci de caracteres cuantitativos (QTL) para identificar regiones cromosómicas que influyen en un determinado carácter diana, la presencia de genes candidatos localizados en la misma región, y la investigación de sus patrones de expresión (mediante los análisis por microarrays y proteómicos), así como su función en distintas especies, se combinarán para identificar genes clave y desvelar la complejidad de la regulación fisiológica de los caracteres diana.

\_\_\_\_\_  
Véase más adelante una descripción más detallada de dichos avances.

las de utilización. Dado que los recursos para la conservación son limitados, suele ser necesaria una priorización. Las nuevas herramientas moleculares permitirán la identificación de genes implicados en un conjunto de caracteres, incluyendo los caracteres adaptativos, así como los polimorfismos que causan la variación genética funcional (QTN – Nucleótidos de Caracteres Cuantitativos). Sin embargo, aún no disponemos de conocimientos suficientes para priorizar las decisiones de conservación sobre la base de la

diversidad molecular funcional, y se requieren medidas alternativas. La caracterización fenotípica proporciona una estimación rudimentaria del promedio de variantes funcionales de genes de los que son portadores un individuo o una población. Sin embargo, la mayoría de fenotipos de la mayor parte de especies agropecuarias no están documentados.

**Primer papel.** En ausencia de datos de QTN y fenotípicos fiables, o para complementar los ya existentes, el modo más rápido y rentable de medir la diversidad genética es mediante el análisis de polimorfismos utilizando marcadores genéticos moleculares anónimos. Los marcadores anónimos pueden proporcionar información indirecta sobre genes funcionales para caracteres importantes, partiendo del supuesto de que las poblaciones singulares que han tenido una determinada historia evolutiva en los marcadores neutros (ya sea debido a un antiguo aislamiento o a una domesticación independiente) es probable que sean portadoras de variantes únicas de las variaciones funcionales. Las técnicas moleculares también han resultado útiles para investigar el origen y la domesticación de especies agropecuarias, y sus migraciones posteriores, así como para dar información sobre relaciones evolutivas (árboles filogenéticos) e identificar áreas geográficas de mezcla entre poblaciones de distintos orígenes genéticos. El subcapítulo 3.1 describe las técnicas moleculares para evaluar la diversidad genética dentro de una raza y entre distintas razas.

**Segundo papel.** El tamaño efectivo de la población ( $N_e$ ) es un índice que estima el número efectivo de animales de una población que se reproducen y aportan genes a la generación siguiente.  $N_e$  está estrechamente vinculado al nivel de endogamia y de deriva genética de una población, y es por tanto, un indicador crucial para evaluar el estado de peligro de una determinada población (véanse las Secciones A y F). Los enfoques tradicionales para obtener valores fiables de  $N_e$  en poblaciones de razas se basan en los datos del pedigrí o en los censos.

## PARTE 4

Los datos necesarios sobre variabilidad del éxito reproductivo e intervalos entre generaciones no suelen estar disponibles de modo fiable en los países en desarrollo. Por consiguiente, los enfoques moleculares pueden ser una alternativa prometedora (véase el subcapítulo 3.2 para más detalles).

**Tercer papel.** Una prioridad máxima en la gestión de recursos zoogenéticos es la conservación de razas con caracteres singulares o únicos. Entre estos, la capacidad de vivir y producir en condiciones desfavorables, y la resistencia a las enfermedades infecciosas son de capital importancia, sobre todo en los países en desarrollo. Los caracteres complejos, como la adaptación y la resistencia a los patógenos, no se visualizan o miden fácilmente. Se pueden investigar en experimentos en los que se somete a los animales a condiciones ambientales específicas o se les inocula el patógeno en estudio. Ahora bien, dichos experimentos son difíciles y caros de realizar, y generan dudas sobre la protección y bienestar de los animales. Esta es la razón por la que los investigadores tienen tanto interés en identificar los genes que controlan caracteres complejos. Dichos genes pueden estudiarse mediante una gama de enfoques distintos. Las herramientas que se están desarrollando para estudiar la variación funcional se describen en el subcapítulo 3.3.

### 3 Visión general de las técnicas moleculares

Esta sección describe las técnicas moleculares más importantes que se están utilizando y desarrollando actualmente para evaluar la diversidad genética y para estudiar la variación funcional. El Recuadro 73 describe cómo se extrae el ADN y ARN del material biológico y su preparación para el análisis. Los atributos de los marcadores moleculares más utilizados se describen en el Recuadro 74, y la toma de muestras (aspecto sumamente importante de los estudios moleculares) en el Recuadro 75.

Los polimorfismos proteicos fueron los primeros marcadores utilizados en estudios genéticos de ganado. Sin embargo, el número de loci polimórficos que se pueden analizar, y el nivel de polimorfismos observados en dichos loci suelen ser bajos, lo cual limita mucho su aplicación a estudios de diversidad genética. Con el desarrollo de nuevas tecnologías, los polimorfismos del ADN han pasado a ser los marcadores de elección en las encuestas de variación genética basadas en datos moleculares (Recuadro 74).

#### 3.1 Técnicas que utilizan marcadores de ADN para evaluar la diversidad genética

##### *Marcadores de ADN nuclear*

Existe un conjunto de marcadores ya disponibles para detectar polimorfismos del ADN nuclear. En los estudios de diversidad genética, los marcadores más utilizados son los microsatélites.

##### *Microsatélites*

Actualmente, los microsatélites (Recuadro 74) son marcadores más populares en los estudios de caracterización genética del ganado (Sunnucks, 2001). Su alta tasa de mutación y naturaleza codominante permiten la estimación de la diversidad genética dentro y entre razas, así como la mezcla genética entre razas incluso si están estrechamente emparentadas.

Hay cierta polémica respecto a la elección de un modelo de mutación –el modelo de alelo continuo o infinito o el modelo de mutación discreto o por pasos (Goldstein *et al.*, 1995)– para el análisis de los datos de microsatélites. De todos modos, los estudios de simulación han demostrado que el modelo de mutación de alelo infinito suele ser generalmente válido para la evaluación de la diversidad dentro de una especie (Takezaki y Nei, 1996).

El número medio de alelos (MNA) por población, y la heterocigosidad observada y esperada ( $H_o$  y  $H_e$ ), son los parámetros más usuales en la evaluación de la diversidad intrarracial. Los parámetros más simples para evaluar la diversidad

### Recuadro 73 Extracción y multiplicación del ADN y ARN

El primer paso en el análisis de ADN, ARN y proteínas es su extracción y purificación a partir de las muestras biológicas. Existen varios protocolos y equipamientos comerciales. Las estrategias aplicadas dependen del material fuente y de la molécula diana. Por ejemplo, la extracción de ADN de sangre entera o leucocitos es relativamente fácil, en tanto que su extracción de alimentos procesados es más difícil. La extracción de ARN de tejido pancreático es difícil debido a su rápida degradación en dicho órgano. La pureza del ADN, ARN y proteínas es a menudo un factor clave que no se cuida lo suficiente en la obtención de resultados fiables.

Tras aislar el ADN (o el ARN) de las células, el siguiente paso es obtener miles o millones de copias de un gen determinado o fragmento de ADN. La multiplicación de fragmentos de ADN se puede encomendar a la acción de microorganismos, generalmente *E. coli*, o puede realizarse *in vitro* utilizando la reacción en cadena de la polimerasa (PCR). Esta técnica, por la que su creador, Cary Mullis, obtuvo el Premio Nobel, amplifica exponencialmente cualquier segmento de ADN de secuencia conocida. El componente clave de una reacción PCR es la ADN polimerasa aislada de *Thermus aquaticus*, un microorganismo adaptado a vivir y multiplicarse a muy alta temperatura. Esta polimerasa termoestable (llamada Taq- por *Thermus aquaticus*) permite la replicación de cadenas en ciclos y produce un crecimiento geométrico del número de copias del ADN diana. Un ciclo PCR incluye tres pasos: i) desnaturalización del ADN a 90-95 °C para separar el ADN en dos cadenas que servirán de plantilla; ii) hibridación de una pareja de oligonucleótidos de cadena única corta (cebadores) complementarios de las regiones diana que flanquean por ambos extremos el fragmento de interés, a 45-65 °C; iii) extensión o elongación de las cadenas recién sintetizadas de ADN iniciada por los cebadores y facilitada por la Taq-polimerasa, a 72 °C. Dicho ciclo se puede repetir, normalmente unas 25 a 45 veces, para permitir la amplificación de suficientes amplicones (un fragmento de un gen o de ADN sintetizado usando la PCR) para que se puedan detectar.

interracial son los índices de diferenciación genética o de fijación. Se han propuesto varios estimadores ( $F_{ST}$  y  $G_{ST}$ ), el más ampliamente utilizado de los cuales es  $F_{ST}$  (Weir y Basten, 1990), que miden el grado de diferenciación genética de las subpoblaciones calculando las varianzas estandarizadas de frecuencias alélicas entre poblaciones. Se puede calcular la significación estadística para los valores de  $F_{ST}$  entre pares de poblaciones (Weir y Cockerham, 1984) para comprobar la hipótesis nula de una falta de diferenciación genética entre poblaciones, y por tanto, la división de la diversidad genética (p. ej., Mburu *et al.*, 2003). Puede realizarse un análisis jerárquico de la varianza molecular (AMOVA) (Excoffier *et al.*, 1992) para evaluar la distribución de diversidad dentro y entre grupos de razas.

Los datos de microsatélites se usan también frecuentemente para evaluar relaciones genéticas entre poblaciones e individuos mediante el cálculo de las distancias genéticas (p. ej., Beja-Pereira *et al.*, 2003; Ibeagha-Awemu *et al.*, 2004; Joshi *et al.*, 2004; Sodhi *et al.*, 2005; Tapio *et al.*, 2005). La medida más habitual de la distancia genética es la distancia genética estándar de Nei ( $D_N$ ) (Nei, 1972). Sin embargo, para poblaciones estrechamente emparentadas, en las que la deriva genética es el principal factor de diferenciación genética, como suele ocurrir en las razas agropecuarias, particularmente en el mundo en desarrollo, se recomienda la distancia Cavalli-Sforza modificada ( $D_A$ ) (Nei *et al.*, 1983). Se suele visualizar la relación genética entre razas mediante la reconstrucción de una filogenia, utilizando habitualmente el método de adyacencia (N-J) (Saitou y Nei, 1987). No obstante, un inconveniente importante de la reconstrucción de árboles filogenéticos es que presupone que la evolución de los linajes no es reticular, es decir, que los linajes pueden divergir, pero nunca ser resultado de cruces entre linajes. Este supuesto rara vez es aplicable al ganado, ya que a menudo las nuevas razas se originan por cruce entre dos o más razas ancestrales. La visualización de la evolución de las razas obtenida por reconstrucción filogenética debe interpretarse, pues, con cautela.

## PARTE 4

### Recuadro 74 Marcadores habituales de ADN

Los polimorfismos por restricción de la longitud de los fragmentos (RFLP) se identifican usando enzimas de restricción que parten el ADN únicamente en «puntos o sitios de restricción» precisos (p. ej., EcoRI corta en el sitio definido por la secuencia palindrómica GAATTC). Actualmente, el uso más frecuente de los RFLP es en combinación con la PCR (PCR-RFLP), para detectar alelos que difieren en secuencia en un sitio de restricción concreto. Primero se amplifica un fragmento de gen con la PCR, y luego se expone a un enzima de restricción específico que corta solamente una de las formas alélicas. Los amplicones digeridos suelen resolverse mediante electroforesis.

Los microsatélites o SSR (Repeticiones de Secuencia Única) o STR (Repeticiones Simples en Tándem) consisten en un tramo de ADN de unos cuantos nucleótidos de longitud – de 2 a 6 pares de bases (bp) – que se repiten varias veces en tándem (p. ej., CACACACACACACA). Están diseminados por todo el genoma de los eucariotas. Los microsatélites son de un tamaño relativamente pequeño y, por consiguiente, pueden ser fácilmente amplificados con la PCR usando ADN extraído de fuentes diversas, como la sangre, el pelo, la piel, e incluso las heces. Los polimorfismos se pueden visualizar en un gel secuenciador, y la disponibilidad de secuenciadores automáticos de ADN permite un análisis ultrarrápido de un gran número de muestras (Goldstein y Schlötterer, 1999; Jarne y Lagoda, 1996). Los microsatélites son hipervariables; muestran a menudo decenas de alelos en un locus que difieren entre sí en el número de repeticiones. Siguen siendo los marcadores de elección para estudios de diversidad, para análisis de parentesco y para el cartografiado de Loci de Caracteres Cuantitativos (QTL), pero esto podría cambiar en el futuro próximo con el desarrollo de métodos baratos para el análisis

de los SNP. La FAO ha publicado recomendaciones para conjuntos de loci de microsatélites a utilizar en estudios de diversidad de las especies agropecuarias más importantes, que fueron desarrollados por el Grupo Asesor sobre Diversidad Genética Animal de la ISAG-FAO (véase la biblioteca DAD-IS en <http://www.fao.org/dad-is/>).

Los minisatélites comparten las mismas características que los microsatélites, pero la longitud de las repeticiones es de entre diez y algunos centenares de pares de bases. Los micro y minisatélites también se denominan polimorfismos VNTR (Número Variable de Repeticiones en Tándem).

Los polimorfismos por ampliación de la longitud del fragmento (AFPL) son una técnica de identificación del ADN que detecta fragmentos de restricción de ADN mediante amplificación con PCR.

Los STS (Sitios con Marca de Secuencia) son secuencias de ADN que solo se dan una vez en un genoma, en una posición conocida. No tienen por qué ser polimórficos y se utilizan para construir mapas físicos.

Los SNP son variaciones en nucleótidos únicos que no cambian la longitud total de la secuencia de ADN en la región. Existen SNP en todo el genoma. Son muy abundantes en el genoma humano, a razón de un SNP por cada 1 000 pares de bases (Sachinandam *et al.*, 2001). La mayoría de SNP se localizan en las regiones no codificantes, y no tienen un impacto directo en el fenotipo de un individuo. No obstante, algunos introducen mutaciones en secuencias expresadas o en regiones que influyen en la expresión génica (promotores, potenciadores), y pueden inducir cambios en la estructura o regulación de las proteínas. Dichos SNP tienen el potencial de detectar la variación genética funcional.

### Recuadro 75 Muestreo del material genético

La recogida de muestras es el primer y más importante paso en cualquier estudio de diversidad. A poder ser, las muestras no deberían estar relacionadas y sí ser representativas de las poblaciones en estudio. Por lo general, la toma de muestras de 30 a 50 individuos bien escogidos por raza se considera suficiente para dar una pista sobre los caracteres distintivos de la raza y la diversidad intraracial, siempre y cuando se analice un número adecuado de marcadores independientes (unos 20–30 microsatélites; Nei y Roychoudhury, 1974; Nei, 1978). Sin embargo, las cifras reales pueden variar de un caso a otro, y pueden incluso ser inferiores en el caso de una población local muy endogámica, y superiores en una población muy diseminada dividida en ecotipos distintos.

La elección de muestras no relacionadas es bastante sencilla en una raza bien definida, ya que puede basarse en el libro del rebaño o en el registro del pedigrí. En cambio, puede ser más difícil en una población semi-asilvestrada de la que no existen registros escritos. En este caso, es muy recomendable el uso de un criterio geográfico, es decir, tomar muestras de un único o de muy pocos animales (no relacionados) por rebaño a partir de los rebaños diseminados en una amplia área geográfica. El registro de las coordenadas geográficas, y la fotodocumentación de los lugares de toma de muestras, de animales y de rebaños resulta valiosísimo –para documentar cruces en caso de valores atípicos inesperados, o para identificar pautas geográficas interesantes de diversidad genética–. Un conjunto bien escogido de muestras es un recurso valioso y duradero, que se puede usar para producir resultados significativos incluso con tecnología básica. Por el contrario, una muestra sesgada producirá resultados distorsionados o difíciles de entender aunque se utilicen las herramientas moleculares más avanzadas.

El análisis multivariante, y más recientemente, los conglomerados bayesianos se han utilizado para el estudio de mezclas de datos de microsatélites de distintas poblaciones (Pritchard *et al.*, 2000). Probablemente, el estudio más completo de este tipo en ganado es el llevado a cabo en África en ganado bovino (Hanotte *et al.*, 2002), que ha revelado las rúbricas genéticas de los orígenes, movimientos secundarios, y diferenciación del pastoreo del ganado bovino africano.

Los datos genéticos moleculares, en conjunción con, y complementados por, datos de otras fuentes, como la evidencia arqueológica y los registros escritos, proporcionan información útil sobre los orígenes, movimientos posteriores, y cambios de la diversidad genética en las especies agropecuarias. El cartografiado del origen de la diversidad genética actual podría permitir hacer inferencias sobre dónde puede hallarse variación genética funcional dentro de una especie de la que solo existen datos limitados sobre la variación fenotípica.

El análisis combinado de los datos de microsatélites obtenidos en estudios separados es muy deseable, pero rara vez ha resultado posible. Ello se debe a que la mayoría de estudios de genética poblacional que usan marcadores de ADN se limitan a un pequeño número de razas, a menudo de un único país (Baumung *et al.*, 2004). Con frecuencia se utilizan distintos subconjuntos de los marcadores recomendados por la FAO, y no se genotipan muestras estándar en distintos proyectos. La aplicación de sistemas distintos de genotipaje de microsatélites causa variación entre estudios respecto al tamaño estimado de alelos en los mismos loci. Para promover el uso de marcadores comunes, la FAO propone ahora una lista<sup>3</sup> actualizada y jerarquizada de loci de microsatélites para las especies agropecuarias más importantes. La FAO recomienda el uso de marcadores en el mismo orden jerárquico, para maximizar el número de marcadores que se solapan en distintas investigaciones. Para algunas especies existe ADN de animales estándar.

<sup>3</sup> Las listas y guías se pueden encontrar en la biblioteca de DAD-IS (<http://www.fao.org/dad-is>).

## PARTE 4

Por ejemplo, se han distribuido muestras del ADN estándar de oveja y cabra utilizado en el proyecto Econogene de la Unión Europea (UE) a otros proyectos a gran escala en Asia y África, que se pueden solicitar a través del Sitio Web de Econogene (<http://www.econogene.eu>).

Solo existen unos cuantos ejemplos de análisis a gran escala de la diversidad genética de las especies agropecuarias. Hillel *et al.* (2003) y SanCristobal *et al.* (2006a) investigaron, respectivamente, la diversidad en gallinas y cerdos en toda Europa; Hanotte *et al.* (2002) obtuvieron datos de ganado bovino a escala de casi todo el continente africano; Tapio *et al.* (2005) evaluaron la diversidad en ovejas a gran escala regional en los países del norte de Europa; y Cañon *et al.* (2006) estudiaron la diversidad en cabras de Europa y Cercano y Medio Oriente. No obstante, para la mayoría de especies, falta aún una revisión completa. La estrecha coordinación actual entre proyectos a gran escala promete ofrecer una estimación mundial de la diversidad genética en un futuro próximo para algunas especies como ovejas y cabras. En el interín, se están desarrollando nuevos métodos de análisis de datos que permitirán el metanálisis de conjuntos de datos que incluyen solo unas cuantas razas y ninguno, o muy pocos, marcadores en común (Freeman *et al.*, 2006). Esta perspectiva global de la diversidad agropecuaria será extremadamente valiosa para reconstruir el origen y la historia de las poblaciones de animales domésticos, e, indirectamente, de las poblaciones humanas. Revelará asimismo «puntos calientes» de diversidad genética regional y local que podrán ser objeto de actuaciones de conservación.

**SNP**

Los SNP (Recuadro 74) se utilizan como alternativa a los microsatélites en los estudios de diversidad genética. Existen diversas tecnologías para detectar y tipar los marcadores SNP (véase la revisión de Syvänen, 2001). Al ser marcadores bialélicos, los SNP poseen un contenido de información bastante bajo, y hay que usar muchos para llegar al nivel de información obtenido a partir de un panel estándar de 30 loci

de microsatélites. No obstante, las tecnologías moleculares están en constante evolución, aumentando la automatización y disminuyendo el costo del tipaje de SNP. Probablemente, en el futuro próximo, ello permitirá el análisis en paralelo de un gran número de marcadores a un costo inferior. Con esta perspectiva, están en marcha proyectos a gran escala en varias especies agropecuarias para identificar millones (p. ej., Wong *et al.*, 2004) y validar varios miles de SNP, e identificar bloques de haplotipos en el genoma. Al igual que la información sobre secuencias, los SNP permitirán una comparación directa y análisis conjunto de diferentes experimentos.

Los SNP parecen ser marcadores atractivos para aplicarlos en el futuro a estudios de diversidad genética ya que se pueden usar fácilmente en la evaluación de la variación funcional o neutra. Sin embargo, la fase preliminar del descubrimiento o selección de los SNP a partir de bases de datos es crítica. Los SNP se pueden generar mediante varios protocolos experimentales, ya sea secuenciado, polimorfismo conformacional de cadena única (SSCP) o cromatografía líquida desnaturalizadora de alto rendimiento (DHPLC), o *in silico*, alineando y comparando secuencias múltiples de la misma región a partir de bases de datos públicas del genoma y de secuencias expresadas (EST). Si los datos no se han obtenido aleatoriamente, no se pueden aplicar los estimadores estándar de los parámetros de la genética poblacional. Un ejemplo frecuente es el que se da cuando los SNP identificados inicialmente en una muestra pequeña (panel) de individuos se tipan en una muestra mayor de cromosomas. Como, de los SNP, se toman muestras preferencialmente a frecuencias intermedias, dicho protocolo sesga la distribución de frecuencias alélicas cuando se comparan con lo esperado de una muestra aleatoria. Los SNP son prometedores para aplicaciones futuras en análisis de genética poblacional; sin embargo, aún están por desarrollar los métodos estadísticos que de manera explícita cubran todos los métodos de descubrimiento de los SNP (Nielsen y Signorovitch, 2003; Clark *et al.*, 2005).



**AFLP**

Los AFLP son marcadores bialélicos dominantes (Vos *et al.*, 1995). Las variaciones en muchos loci se pueden detectar simultáneamente en un microarray para detectar variaciones en nucleótidos únicos en regiones genómicas desconocidas, en las que puede hallarse frecuentemente una mutación en genes funcionales indeterminados. Un inconveniente, sin embargo, es que muestra un modo dominante de transmisión; ello reduce su capacidad en análisis genéticos poblacionales de la diversidad dentro de una raza, y en la endogamia. Por otra parte, los perfiles AFLP dan mucha información relativa a la relación entre razas (Ajmone-Marsan *et al.*, 2002; Negrini *et al.*, 2006; De Marchi *et al.*, 2006; SanCristobal *et al.*, 2006b) y especies emparentadas (Buntjer *et al.*, 2002).

**Marcadores de ADN mitocondrial**

Los polimorfismos del ADN mitocondrial (mtADN) se han utilizado ampliamente en análisis filogenéticos y de diversidad genética. El mtADN haploide, contenido por las mitocondrias en el citoplasma celular, presenta un modo materno de transmisión (los individuos heredan el mtADN de sus madres y no de los sementales), así como una alta tasa de mutación; no se recombina. Dichas características permiten a los biólogos reconstruir relaciones evolutivas dentro de una especie y entre distintas especies valorando las pautas de mutación del mtADN. Los marcadores de mtADN pueden también permitir la rápida detección de hibridación entre especies o subespecies de ganado (p. ej., Nijman *et al.*, 2003).

Los polimorfismos en la secuencia de la región hipervariable del bucle-D o región de control del mtADN han contribuido en gran medida a identificar los progenitores salvajes de las especies domésticas, establecer pautas geográficas de diversidad genética, y a entender la domesticación del ganado (véase la revisión de Bruford *et al.*, 2003). Por ejemplo, se ha demostrado recientemente que el origen del ganado bovino europeo moderno tuvo lugar en Oriente Medio (Troy *et al.*, 2001). El estudio

identificó cuatro estirpes maternas de *Bos taurus* y demostró asimismo la pérdida de variabilidad genética bovina durante la migración neolítica humana a partir del Creciente Fértil. Del mismo modo, se han observado en cabras múltiples orígenes maternos, con tres estirpes de mtADN (Luikart *et al.*, 2001), siendo Asia y el Creciente Fértil los posibles centros de origen. Recientemente, se descubrió un tercer linaje de mtADN en ovejas chinas nativas (Guo *et al.*, 2005), un cuarto en cabras chinas nativas (Chen *et al.*, 2005), y un quinto en bovinos chinos (Lai *et al.*, 2006). En gallinas asiáticas, se han descubierto nueve clados diferentes de mtADN (Liu *et al.*, 2006), lo cual sugiere orígenes múltiples en el Asia meridional y sudoriental. Todos estos resultados indican que nuestro conocimiento actual de la domesticación y diversidad genética del ganado dista de ser completo. Para mayor información sobre los orígenes de las especies agropecuarias domésticas, véase la Parte 1 –Sección A.

**3.2 Marcadores utilizados para calcular el tamaño efectivo de una población**

Hill (1981) sugirió utilizar el desequilibrio de la fase gamética en los polimorfismos del ADN para calcular el tamaño efectivo de una población ( $N_e$ ). Dicho cálculo se puede basar en genotipos para marcadores ligados (microsatélites o SNP). La correlación esperada de frecuencias alélicas en loci ligados es función de  $N_e$  y de la tasa de recombinación.  $N_e$  puede, por tanto, calcularse a partir del desequilibrio observado. Hayes *et al.* (2003) sugirió un enfoque similar, basado en la homocigosis cromosómica segmentaria, que, además, tiene el potencial de calcular  $N_e$  de generaciones anteriores, y por tanto, posibilita decidir si una población existente aumentaba o disminuía de tamaño en el pasado. El estudio demostró, con conjuntos de datos de ejemplo, que la raza bovina Holstein-frisona sufrió una reducción notable de su  $N_e$  en el pasado, en tanto que el tamaño efectivo de la población humana estaba aumentando, lo cual está de acuerdo con los estudios de censo y de pedigrí.

## PARTE 4

### 3.3 Herramientas moleculares para estudiar la variación funcional

#### **Enfoques basados en la posición: cartografiado de loci para caracteres cuantitativos (QTL)**

Los marcadores genéticos se comportan como caracteres mendelianos; dicho de otro modo, siguen las leyes de segregación y distribución independiente que Mendel fue el primero en describir. Dos genes que se encuentren en el mismo cromosoma están físicamente ligados y tienden a heredarse juntos. Durante la meiosis, la recombinación entre cromosomas homólogos puede romper este ligamiento. La frecuencia de recombinación entre dos genes situados en el mismo cromosoma depende de la distancia que los separa. La tasa de recombinación entre marcadores es, por tanto, indicativa de su grado de ligamiento: cuanto más baja es la tasa de recombinación, más cerca estarán los marcadores. La construcción de mapas genéticos aprovecha esta característica para inferir el orden probable de los marcadores y la distancia entre ellos.

Los estudios de cartografiado se suelen realizar tras la co-segregación de marcadores polimórficos en poblaciones experimentales estructuradas (es decir, F2 o retrocruzamiento) o en poblaciones ya existentes sometidas a programas de selección (familias de hermanos completos o parciales). Para la mayoría de especies agropecuarias se dispone de mapas genéticos de densidad media o alta para algunos centenares o millares de marcadores.

Para identificar un QTL para un carácter concreto, una familia que segregue el carácter se genotipa con un conjunto de marcadores moleculares regularmente repartidos por el genoma (Recuadro 76). Existen varios métodos estadísticos para inferir la presencia de un QTL significativo en un determinado intervalo del marcador, pero todos se basan en el hecho de que las familias poseen un alto grado de desequilibrio de ligamiento, es decir, que grandes segmentos cromosómicos se transmiten de los padres a la progenie sin recombinación alguna.

El resultado de un experimento de cartografiado QTL es la identificación de una región cromosómica, que a menudo ocupa medio cromosoma, en la que se detecta un efecto significativo para el carácter diana. La investigación moderna utiliza activamente el cartografiado para identificar la influencia de QTL en los caracteres adaptativos. En gallinas, ejemplos de dichos caracteres incluyen la resistencia a la colonización y excreción de *Salmonella* (Tilquin *et al.*, 2005), y la susceptibilidad a presentar el síndrome de hipertensión pulmonar (Rabie *et al.*, 2005); y en ganado bovino, la tripanotolerancia (Hanotte *et al.*, 2002).

La fase de cartografiado QTL va seguida generalmente por un refinamiento de la posición del QTL en el mapa (cartografiado QTL preciso). Para llevar a cabo esta tarea, se analizan otros marcadores, y especialmente, todos los casos de recombinación adicional en el área diana. Recientemente se ha diseñado y aplicado un astuto enfoque al cartografiado preciso de una región cromosómica en BTA14, que presenta un QTL significativo para el porcentaje de grasa en leche y otros caracteres (Farnir *et al.*, 2002). Este enfoque explota la recombinación histórica en generaciones anteriores para restringir la posición en el mapa a una región relativamente pequeña, de 3,8 cM (centimorgan), tamaño que ha permitido el clonaje posicional del gen (DGAT1) (Grisart *et al.*, 2002).

Concluido el cartografiado preciso, los genes que determinan el carácter de interés se pueden buscar entre los genes situados en las regiones identificadas. Los genes candidatos se pueden buscar en la misma especie (p. ej., cuando existe un mapa rico en etiquetas de secuencia expresada, o cuando el genoma está totalmente secuenciado), o en regiones ortólogas de un organismo modelo para el cual se dispone de información genómica completa.

Ocasionalmente, la información clave relativa a una función génica proviene de una fuente inesperada. Esto ocurrió con el gen de la miostatina, cuya función se descubrió

inicialmente en ratones y más tarde se localizó en ganado bovino en la región cromosómica donde previamente se había cartografiado el gen de doble musculatura (McPherron y Lee, 1997).

Queda claro que identificar el gen responsable (genes de caracteres cuantitativos – QTG) y la mutación funcional (QTN) de un carácter complejo sigue siendo una tarea ímproba, y se requieren diversos enfoques para disminuir el número de genes candidatos posicionales. A este respecto, disponer de información sobre la función génica es fundamental. Sin embargo, desconocemos la posible función o funciones de la mayoría de genes identificados mediante secuenciado del genoma y del cADN (ADN complementario). Por eso, la investigación de pautas de expresión génica puede proporcionar información útil, en combinación con el enfoque posicional antes descrito, para identificar genes candidatos para caracteres complejos. A este enfoque combinado se le conoce como genómica genética (Haley y de Koning, 2006). En la sección siguiente se describen nuevos avances en la investigación de pautas de expresión génica.

Actualmente se están investigando enfoques alternativos para detectar genes adaptativos utilizando marcadores genéticos (Recuadro 77). Se hallan en etapa experimental, y solo la investigación ulterior permitirá evaluar su eficacia.

El objetivo final del cartografiado QTL es identificar el QTG, hasta llegar al QTN. Aunque a fecha de hoy existen pocos ejemplos en ganado, este es el tipo de mutaciones que podrían tener un impacto directo en la cría basada en marcadores y en las decisiones de conservación. Deberán desarrollarse modelos de conservación que incluyan los caracteres funcionales y las mutaciones, ya que en un futuro próximo se descubrirá un número creciente de QTG y QTN.

#### **Investigación de pautas de expresión génica**

En el pasado, la expresión de caracteres específicos, como la adaptación y la resistencia, solo podían medirse a nivel fenotípico. Actualmente, el transcriptoma (el conjunto de

#### **Recuadro 76 Cartografiado de QTL**

Si existe un QTL para un carácter diana, el alelo variante más- y menos- del gen responsable desconocido (Q y q) se co-segregará con los alelos en un marcador M1 cercano (M1 y m1) que podremos genotipar en el laboratorio. Supongamos que M1 se co-segrega con Q, y m1 con q, es decir, que M1 y Q están cerca en el mismo cromosoma, y que m1 y q están en el cromosoma homólogo (M1Q y m1q).

Supongamos también que genotipamos una población F2 obtenida por apareamiento de individuos heterocigotos F1. Tras el genotipaje, las progenies F2 se agrupan sobre la base de su genotipo marcador (M1M1 y m1m1; M2M2 y m2m2; ... MnMn y mnmn), y luego comparamos el fenotipo medio de los grupos. Si no existe ningún QTL ligado a un marcador dado (p. ej., M2), entonces no detectaremos una diferencia significativa entre el valor fenotípico medio de las progenies M2M2 y m2m2 para el carácter diana. Por el contrario, cuando las progenies se agrupan por genotipo en el marcador M1, entonces el grupo M1M1 será en su mayor parte QQ en el QTL, en tanto que el grupo m1m1 será principalmente qq. En este caso, se observa una diferencia significativa entre las medias de la progenie, y por lo tanto, detectamos la presencia de un QTL. En especies como aves y cerdos en las que se suelen entrecruzar comercialmente líneas y razas, el estudio puede emprenderse en poblaciones experimentales (F2, BC), en tanto que en rumiantes se suelen utilizar dos pedigrís de generación (diseño hijas – DD) o tres (diseño nietas – GDD). En DD la segregación de marcadores heterocigóticos en un semental (generación I) se sigue en las hijas (generación II) cuyos datos fenotípicos se están recogiendo. En GDD, la segregación de marcadores heterocigóticos en un semental abuelo (generación I) se sigue en sus hijos medio-hermanos (generación II), cuyo fenotipo se infiere a partir del de las nietas (generación III).

todos los transcriptos en una célula o tejido), y el proteoma (el conjunto de todas las proteínas), se pueden investigar directamente mediante

## PARTE 4

técnicas ultrarrápidas, como la visualización diferencial (DD) (Liang y Pardee, 1992), cADN-AFLP (Bachem *et al.*, 1996), el análisis seriado de la expresión génica (SAGE) (Velculescu *et al.*, 1995; 2000), la espectrometría de masas, y los microarrays de proteínas y ADN. Dichas técnicas representan un gran avance en el análisis del ARN y las proteínas, permitiendo el análisis en paralelo de prácticamente todos los genes expresados en un tejido en un momento dado. Así pues, las técnicas contribuyen a descodificar las redes que probablemente subyacen en muchos caracteres complejos.

Las tecnologías «-ómicas» se suelen comparar a iluminar la totalidad de un fresco de Michelangelo en vez de hacerlo con una antorcha que solo permite ver una parte a la vez. Esta visión global permite entender el significado de la representación y apreciar su belleza. En realidad, la potencia de estas técnicas es paralela a la dificultad y el costo de aplicarlas y de procesar los datos obtenidos. Aislar muestras homogéneas de células es bastante difícil, pero es una condición previa importante en muchos estudios de perfiles de expresión génica. El gran número de análisis en paralelo redundan en un menor costo por análisis, pero en un mayor costo por experimento. El equipo es caro, y se precisa una pericia técnica considerable en todas las fases experimentales. Aparte de la dificultad general de analizar el ARN y compararlo con el ADN. El ARN es muy sensible a la degradación, y debe extraerse con mucho cuidado de los tejidos que muestran un metabolismo muy activo. De hecho, la conservación y manipulación de la muestra es una de las claves del éxito en experimentos de análisis del ARN. La aplicación de nanotecnologías al análisis de moléculas biológicas está abriendo perspectivas prometedoras en la resolución en estos problemas (Sauer *et al.*, 2005).

Otro problema es el tratamiento de los datos. Los conjuntos de datos moleculares, como los perfiles de expresión génica, se pueden generar en un tiempo relativamente corto. Sin embargo, es fundamental homologar datos entre laboratorios para poder analizar coherentemente los diversos

conjuntos de datos biológicos. Es esencial llegar a acuerdos sobre la estandarización, así como crear bases de datos interconectadas, para un análisis eficiente de las redes moleculares.

### *Perfiles de transcriptos*

En esta sección se describen brevemente las técnicas SAGE y de microarrays. Se pueden hallar descripciones de otras técnicas en revisiones recientes (p. ej. Donson *et al.*, 2002). SAGE genera perfiles completos de expresión de tejidos o líneas celulares. Comprende la construcción de bibliotecas de mARN total que permiten un análisis cuantitativo de todos los transcriptos expresados o inactivados en determinados pasos de una activación celular. Se basa en tres principios: i) una etiqueta de secuencia corta (9–14 pares de bases) obtenida de una región definida dentro de cada transcripto de mARN que contiene suficiente información para identificar de manera inequívoca un transcripto específico; ii) las etiquetas de secuencia se pueden unir entre sí para formar largas moléculas de ADN (concatámeros), que se pueden clonar y secuenciar – el secuenciado de los clones de concatámeros conduce a una rápida identificación de numerosas etiquetas individuales; iii) el nivel de expresión se cuantifica en base al número de veces que se observa una etiqueta determinada.

Los microarrays pueden utilizarse para comparar, en un único experimento, los niveles de expresión de mARN de varios millares de genes entre dos sistemas biológicos, por ejemplo, entre animales en un ambiente normal y animales en un ambiente desfavorable. Las técnicas de microarray pueden contribuir a una mejor comprensión de las pautas temporales y espaciales de expresión génica en respuesta a una amplia gama de factores a los que el organismo está expuesto.

Se depositan volúmenes muy pequeños de una solución de ADN sobre un portaobjetos de un material no poroso como el vidrio, creando manchitas de unas 100 a 150  $\mu\text{m}$  de diámetro. Actualmente se pueden depositar mediante un robot unos 50 000 ADN complementarios (cADN) en un portaobjetos de microscopio. Los

microarrays de ADN contienen varios centenares de genes conocidos, y algunos millares de genes desconocidos. El microarray se cubre con manchitas de fragmentos de cADN o con oligonucleótidos prefabricados. Esta última opción tiene la ventaja de una mayor especificidad y reproducibilidad, pero sólo se puede diseñar si se conoce la secuencia. El uso de los microarrays se basa en el principio de «hibridación», es decir, se exponen entre sí dos secuencias de ADN de cadena única, o una de ADN y otra de ARN, y se mide la cantidad de molécula de doble cadena que se forma. La expresión del mRNA se puede medir cualitativa y cuantitativamente. Indica actividad génica en un tejido, y suele estar directamente relacionada con la producción de proteína inducida por el mRNA.

Los perfiles de expresión génica contribuyen a la comprensión de los mecanismos biológicos, y por tanto, facilitan la identificación de genes candidatos. Por ejemplo, el grupo de genes implicado en la expresión de la tripanotolerancia en ganado bovino se ha caracterizado mediante SAGE (Berthier et al., 2003), y por análisis de cADN en microarray (Hill et al., 2005). La investigación en paralelo de la expresión de muchos genes puede permitir la identificación de genes maestros responsables de caracteres fenotípicos que no se pueden detectar mediante análisis de expresión diferencial. Dichos genes maestros pueden, por ejemplo, presentar diferentes alelos que se expresan todos al mismo nivel, promoviendo la expresión de genes a niveles inferiores con distinta eficiencia. En este caso, el gen maestro se puede buscar ya sea aprovechando los conocimientos actuales sobre vías metabólicas, o bien a través de un QTL de expresión (eQTL) (Lan et al., 2006). En dicho enfoque, el nivel de expresión de los genes de nivel inferior se mide en una población segregante. La cantidad de transcrito de cada gen se trata como un carácter fenotípico, los QTL que influyen en la expresión génica pueden buscarse utilizando las metodologías antes descritas. Cabe señalar que el análisis de datos para la detección de QTL dista aún mucho de ser fácil. Esto es también aplicable a las técnicas de

perfiles de transcritos, debido a las numerosas señales falsas que se producen.

#### *Perfiles proteicos*

El estudio sistemático de las estructuras proteicas, las modificaciones post-traslacionales, los perfiles proteicos, las interacciones entre proteína-proteína, proteína-ácido nucleico, y proteína-molécula pequeña, así como la expresión espacial y temporal de las proteínas en las células eucariotas, son cruciales para comprender fenómenos biológicos complejos. Las proteínas son esenciales para la estructura de las células vivas y sus funciones.

La estructura de una proteína se puede revelar mediante difracción de rayos X o por espectroscopía de resonancia magnética nuclear. La primera requiere una gran cantidad de proteína cristalina, y esto suele ser restrictivo. Para comprender la función proteica y las interacciones proteína-proteína a nivel molecular, sería útil determinar la estructura de todas las proteínas en una célula u organismo. Hasta la fecha, sin embargo, esto no se ha conseguido. Es interesante observar que el número de variantes proteicas diferentes surgidas de la síntesis proteica (empalme alternativo y/o modificaciones post-traslacionales) es significativamente mayor que el número de genes en un genoma.

La espectrometría de masas (una técnica analítica para determinar la masa molecular) en combinación con técnicas de separación cromatográficas o electroforéticas, es actualmente el método de elección para identificar proteínas endógenas en las células, para caracterizar las modificaciones post-traslacionales y determinar la abundancia proteica (Zhu et al., 2003). La electroforesis bidimensional en gel es única con respecto al gran número de proteínas (>10 000) que pueden separarse y visualizarse en un solo experimento. Las manchas de las proteínas se recortan del gel, se someten a digestión proteolítica, y se identifican las proteínas usando la espectrometría de masas (Aebersold y Mann, 2003). Sin embargo,

## PARTE 4

ha resultado difícil estandarizar y automatizar la electroforesis bidimensional en gel, y el uso de los patrones proteicos resultantes como mapas de referencia proteómica solo ha tenido éxito en unos cuantos casos. Una técnica complementaria, la cromatografía líquida, es más fácil de automatizar, y se puede acoplar directamente al espectrómetro de masas. Los métodos proteómicos por afinidad, basados en microarrays, son un enfoque alternativo a los perfiles proteicos (Lueking *et al.*, 2003), y también se pueden utilizar para detectar interacciones proteína-proteína. Dicha información es esencial para el modelado algorítmico de las vías biológicas. No obstante, la especificidad del enlace sigue siendo un problema en la aplicación de los microarrays de proteínas, ya que la reactividad cruzada no se puede predecir con exactitud. Existen enfoques alternativos para detectar las interacciones proteína-proteína, como el sistema de dos híbridos (Fields y Song, 1989). Sin embargo, ninguno de los métodos actualmente utilizados permite la detección cuantitativa de las proteínas de enlace, y no queda claro hasta qué punto las interacciones observadas representan las interacciones proteína-proteína.

También se han desarrollado métodos basados en arrays o matrices para detectar la interacción ADN-proteína *in vitro* e *in vivo* (véase la de Sauer *et al.*, 2005), e identificar el enlace de proteínas desconocidas con secuencias génicas reguladoras. Los microarrays de ADN se emplean eficazmente para el cribado de extractos nucleares en busca de complejos que enlacen ADN, en tanto que los microarrays proteicos se utilizan básicamente para identificar proteínas desconocidas que enlazan ADN a nivel de todo el proteoma. En el futuro, estas dos técnicas permitirán conocer los detalles del funcionamiento de las redes reguladoras transcripcionales.

Muchos métodos de predecir la función de una proteína se basan en su homología con otras proteínas y su localización intracelular. Predecir las funciones proteicas es bastante complicado, y también exige técnicas para detectar las interacciones proteína-proteína, y detectar

asimismo el enlace de las proteínas con otras moléculas, puesto que las proteínas realizan sus funciones en dichos procesos de enlace.

### 4 El papel de la bioinformática

El desarrollo de tecnologías ultrarrápidas sería inútil sin la capacidad de analizar el volumen de datos biológicos, que crece exponencialmente. Los datos se almacenan en bases de datos electrónicas (Recuadro 78) asociadas a aplicaciones informáticas específicas diseñadas para permitir la actualización de los datos, su interrogación y recuperación. La información debe ser fácilmente accesible, flexible a la interrogación, para permitir la recuperación de información que se utilizará para desentrañar las vías metabólicas y el papel de los genes y proteínas implicadas.

La bioinformática es crucial para combinar información de diversas fuentes y generar nuevo conocimiento a partir de los datos existentes. Tiene, además, el potencial de simular la estructura, función y dinámica de los sistemas moleculares, y es por tanto útil en la formulación de hipótesis que orienten el trabajo experimental.

### 5 Conclusiones

La caracterización molecular puede desempeñar un papel en dilucidar la historia, y estimar la diversidad, caracteres distintivos y estructura poblacional de los recursos zoogenéticos. Puede también servir de ayuda en el manejo genético de pequeñas poblaciones, para evitar una endogamia excesiva. Se han descrito diversas investigaciones sobre diversidad entre y dentro de poblaciones – algunas a escala bastante grande. Sin embargo, estos estudios están fragmentados y son difíciles de comparar e integrar. Además, aún está por hacer una encuesta mundial completa de las especies relevantes. Por tanto, es de importancia estratégica desarrollar métodos para combinar los conjuntos de datos existentes, que se solapan parcialmente, y garantizar el

## Recuadro 77 El enfoque de la genómica poblacional

Recientemente se ha propuesto un enfoque alternativo a la identificación de regiones genómicas portadoras de genes de interés. Consiste en la detección de «rúbricas de selección» mediante un enfoque de «genómica poblacional» (Black *et al.*, 2001; Luikart *et al.*, 2003). Los tres principios básicos de la genómica poblacional aplicada al cartografiado QTL son:

1. que los loci neutros de todo el genoma se verán similarmente afectados por la deriva genética, la demografía, y la historia evolutiva de las poblaciones;
2. que los loci bajo selección se comportan a menudo de manera distinta, y, por tanto, revelan pautas de variación con valores atípicos, pérdida de diversidad (habría aumento de la misma si los loci se hallaran bajo una selección equilibrada), desequilibrio de ligamiento, y aumentos/disminuciones de los índices *Gst/Fst*; y
3. que merced a los efectos de ligamientos oportunistas, la selección afecta también a los marcadores ligados, permitiendo la detección de una «rúbrica de selección» (efectos de valores atípicos), que a menudo puede detectarse genotipando un gran número de marcadores en un cromosoma e identificando bloques de valores atípicos. Este enfoque utiliza datos fenotípicos a nivel de raza (o de subpoblaciones dentro de una raza) más que a nivel individual, y por lo tanto complementa bien el cartografiado QTL clásico dentro de los pedigrees.

El enfoque de la genómica poblacional permite asimismo identificar genes sometidos a una fuerte

presión de selección que han quedado fijados finalmente en las razas, y en concreto, los genes implicados en la adaptación a ambientes extremos, resistencia a las enfermedades, etc. Muchos de estos caracteres, que son de gran importancia para la sostenibilidad de la cría de ganado, son difíciles o imposibles de investigar mediante el cartografiado QTL clásico o los estudios de asociación. El potencial de la genómica poblacional se ha investigado recientemente desde un punto de vista teórico (Beaumont y Balding, 2004; Bamshad y Wooding, 2003), y mediante trabajo experimental con diferentes tipos de marcadores en poblaciones naturales (AFLP: Campbell y Bernatchez, 2004; microsatélites: Kayser *et al.*, 2003; SNP: Akey *et al.*, 2002). El enfoque se ha aplicado recientemente en el proyecto Econogene (<http://lasig.epfl.ch/projets/econogene>). En análisis preliminares, tres SNP en los genes de MYH1 (miosina 1), MEG3 (calipigia) y CTSB (catepsina B) en ovejas han mostrado un comportamiento atípico significativo (Pariset *et al.*, 2006).

En el seno del mismo proyecto, se ha diseñado un enfoque novedoso, basado en el Método de Análisis Espacial (SAM), para detectar rúbricas de selección natural dentro del genoma de animales domésticos y salvajes (Joost, 2006). Los resultados preliminares obtenidos con dicho método concuerdan con los obtenidos por aplicación de modelos teóricos de la genética poblacional, como los desarrollados por Beaumont y Balding (2004). El SAM va más allá de los enfoques clásicos, puesto que está concebido para identificar parámetros ambientales asociados con marcadores seleccionados.

suministro de muestras y marcadores estándar para su uso futuro como referencias mundiales. Facilitaría la puesta en práctica de una encuesta mundial disponer de una red de instalaciones que recogieran plasma germinal autóctono, y lo pusieran a disposición de la comunidad científica bajo una reglamentación apropiada.

Las tecnologías de marcadores evolucionan, y es probable que los microsatélites se complementen cada vez más con SNP. Dichos marcadores son muy prometedores debido a su abundancia en el genoma, y por su adecuación a la automatización en producción y puntuación. No obstante, queda por aclarar a fondo la eficiencia de los SNP en

## PARTE 4

la investigación de la diversidad en especies animales. El tema debería estudiarse con cierta distancia crítica para evitar la producción de resultados sesgados.

También evolucionan los métodos de análisis de datos. Los nuevos métodos permiten el estudio de la diversidad sin supuestos a priori respecto a la estructura de la población investigada; la exploración de diversidad para identificar genes adaptativos (p. ej. utilizando genómica poblacional, véase el Recuadro 77); y la integración de información de fuentes diversas, incluyendo parámetros socioeconómicos y medioambientales, para establecer prioridades de conservación (véase la Sección F). La adopción de una estrategia de toma de muestras correcta y la recogida sistemática de datos fenotípicos y ambientales, siguen siendo factores clave para explotar el pleno potencial de las nuevas tecnologías y enfoques.

Además de la variación neutra, la investigación busca activamente genes que influyen en caracteres clave. La resistencia a enfermedades, eficiencia productiva y calidad del producto se hallan entre los caracteres que reciben una alta prioridad. Para ello se usan diversas estrategias y las nuevas tecnologías «-ómicas» ultrarrápidas. La identificación de QTN ofrece nuevas oportunidades y retos para la gestión de recursos zoogenéticos. La información sobre la diversidad adaptativa complementa la existente sobre diversidad genética neutral y fenotípica, y se puede integrar en las herramientas de toma de decisiones sobre gestión y conservación de los recursos zoogenéticos. La identificación de alelos únicos o combinaciones de alelos para caracteres adaptativos en poblaciones concretas puede reforzar la justificación para su conservación y utilización. La selección basada en los genes tiene además el potencial de disminuir la brecha de eficiencia selectiva que se observa hoy entre grandes poblaciones criadas en sistemas de producción industriales, y las pequeñas poblaciones locales, donde los sistemas de evaluación genética poblacional y planes de cría no

## Recuadro 78

## Bases de datos de moléculas biológicas

Existen diversas bases de datos que recogen información sobre moléculas biológicas:

**Bases de datos de secuencias de ADN:**

- European Molecular Biology Lab (EMBL): <http://www.ebi.ac.uk/embl/index.html>
- GenBank: <http://www.ncbi.nlm.nih.gov/>
- DNA Data Bank of Japan (DDBJ): <http://www.ddbj.nig.ac.jp>

**Bases de datos de proteínas:**

- SWISS-PROT: <http://www.expasy.ch/sprot/sprot-top.html>
- Protein Information Resource (PIR): <http://pir.georgetown.edu/pirwww/>
- Protein Data Bank (PDB): <http://www.rcsb.org/pdb/>

**Sitios de identificación génica o Bio-Portal**

- GenomeWeb: <http://www.hgmp.mrc.ac.uk/GenomeWeb/nuc-geneid.html>
- BCM Search Launcher: <http://searchlauncher.bcm.tmc.edu/>
- MOLBIOL: <http://www.molbiol.net/>
- Pedro's BioMolecular Research tools: [http://www.biophys.uni-duesseldorf.de/BioNet/Pedro/research\\_tools.html](http://www.biophys.uni-duesseldorf.de/BioNet/Pedro/research_tools.html)
- ExPASy Molecular Biology Server: <http://www.expasy.ch/>

**Bases de datos de interés particular sobre animales domésticos:**

- <http://locus.jouy.inra.fr/cgi-bin/bovmap/intro.pl>
- <http://www.cgd.csiro.au/cgd.html>
- <http://www.ri.bbsrc.ac.uk/cgi-bin/arkdb/browsers/>
- <http://www.marc.usda.gov/genome/genome.html>
- <http://www.ncbi.nlm.nih.gov/genome/guide/pig/>
- <http://www.ensembl.org/index.html>
- <http://www.tigr.org/>
- <http://omia.angis.org.au/>
- <http://www.livestockgenomics.csiro.au/ibiss/>
- <http://www.thearkdb.org/>
- <http://www.hgsc.bcm.tmc.edu/projects/bovine/>



se pueden aplicar de manera eficaz. No obstante, la selección basada en genes y marcadores puede no ser siempre la mejor solución. Dichas opciones deben evaluarse y optimizarse caso por caso, tomando en consideración los efectos a corto y largo plazo sobre la población, sobre la tasa de endogamia, así como los costos y beneficios en términos socioeconómicos y medioambientales – concretamente sobre el nivel de vida de la población.

Como ocurre con otras tecnologías avanzadas, es muy deseable que los beneficios de los avances científicos en el campo de la caracterización molecular se compartan por todo el planeta, contribuyendo así a una mejor comprensión, utilización y conservación de los recursos zoogenéticos del mundo en beneficio de las generaciones humanas actuales y venideras.

## PARTE 4

## Recuadro 79

## Glosario: marcadores moleculares

Para el material de esta sección se utilizan las siguientes definiciones:

**ADN:** la información genética de un genoma está codificada en el ácido desoxirribonucleico (ADN), que se conserva en el núcleo celular. El ADN tiene dos cadenas estructuradas en una doble hélice, formada por un glúcido (desoxirribosa), fosfato y cuatro bases químicas – los nucleótidos: adenina (A), guanina (G), citosina (C) y timina (T). Una A en una cadena se empareja siempre con una T en la otra mediante dos enlaces de hidrógeno, en tanto que una C siempre se empareja con una G mediante tres enlaces de hidrógeno. Las dos cadenas son, pues, complementarias entre sí.

**ADN complementario (cADN):** secuencias de ADN generadas por la transcripción inversa de las secuencias de mRNA. Este tipo de ADN incluye exones y regiones no traducidas en los extremos 5' y 3' de los genes, pero no incluye el ADN del intrón.

**ARN:** el ácido ribonucleico es un ácido nucleico de una cadena formado por tres de las cuatro bases presentes en el ADN (A, C y G). T, sin embargo, es sustituida por uracilo (U).

**Cebador:** una secuencia corta (de una cadena) de oligonucleótidos utilizados en la reacción en cadena de la polimerasa (PCR).

**Desequilibrio de ligamiento (LD):** es un término utilizado en el estudio de genética poblacional para la asociación no aleatoria de alelos en dos o más loci, no necesariamente en el mismo cromosoma. No es lo mismo que el ligamiento, que describe la asociación de dos o más loci en un cromosoma con un grado limitado de recombinación. LD describe una situación en la que algunas combinaciones de alelos o marcadores genéticos se dan en una población más o menos frecuentemente de lo que sería de esperar de una formación aleatoria de haplotipos de alelos sobre la base de sus frecuencias. El desequilibrio de ligamiento es causado interacciones adaptativas entre genes o por procesos no adaptativos como son la estructura de la población, endogamia, y efectos

estocásticos. En la genética de poblaciones, se afirma que el desequilibrio de ligamiento caracteriza la distribución del haplotipo en dos o más loci.

**Gen candidato:** cualquier gen que pudiera plausiblemente causar diferencias en las características observables de un animal (p. ej., en resistencia a las enfermedades, producción de proteína en leche, o crecimiento). El gen puede ser un candidato porque esté situado en una determinada región cromosómica de la que se sospecha que está implicada en el control del carácter, o porque su producto proteico pueda sugerir que puede estar implicado en el control del carácter (p. ej., genes de proteína láctea en la producción de proteína láctea).

**Haplotipo:** contracción de la expresión «genotipo haploide»; se refiere a la constitución genética de un cromosoma individual. En el caso de organismos diploides, el haplotipo contendrá un miembro de la pareja de alelos para cada locus. Puede referirse a un conjunto de marcadores (p. ej., polimorfismos de un solo nucleótido – SNP) que estadísticamente están asociados a un único cromosoma. Sabido esto, se cree que la identificación de unos cuantos alelos de un bloque de haplotipo puede identificar de manera inequívoca el resto de loci polimórficos de la región. Dicha información es de gran utilidad para investigar la genética de los caracteres complejos.

**Ligamiento:** la asociación de genes y/o marcadores cercanos entre sí en un cromosoma. Los genes y marcadores ligados tienden a heredarse juntos.

**Marcador genético:** un polimorfismo del ADN que se puede detectar fácilmente mediante análisis fenotípico o molecular. El marcador puede hallarse dentro de un gen o en un ADN sin función conocida. Dado que los segmentos de ADN que se encuentran próximos entre sí en un cromosoma tienden a heredarse juntos, los marcadores se suelen utilizar como maneras indirectas de seguir la pista del patrón

• continúa

**Recuadro 79 cont.****Glosario: marcadores moleculares**

de herencia de un gen que aún no se ha identificado, pero cuya localización aproximada sí es conocida.

**Tecnología de microarray:** una nueva forma de estudiar cuántos genes interaccionan entre sí y cómo las redes reguladoras de la célula controlan simultáneamente vastas baterías de genes. El método utiliza un robot que aplica con precisión pequeñas gotitas de ADN funcional sobre portaobjetos de

vidrio. Luego los investigadores unen marcadores fluorescentes al mRNA o al cADN de la célula que están estudiando. Las sondas marcadas se enlazan a cadenas de cADN en el portaobjetos, y estos se colocan en un microscopio de rastreo que puede medir el fulgor de cada punto fluorescente; el brillo revela la cantidad presente de un mRNA concreto, indicador de su grado de actividad.

**Referencias**

- Aebersold, R. y Mann, M. 2003. Mass spectrometry-based proteomics. *Nature*, 422 (6928): 198–207. Review.
- Ajmone-Marsan, P., Negrini, R., Milanesi, E., Bozzi, R., Nijman, I.J., Buntjer, J.B., Valentini, A. y Lenstra, J.A. 2002. Genetic distances within and across cattle breeds as indicated by biallelic AFLP markers. *Animal Genetics*, 33: 280–286.
- Akey, J.M., Zhang, G., Zhang, K., Jin, L. y Shriver, M.D. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, 12(12): 1805–14.
- Aravin, A. y Tuschl, T. 2005. Identification and characterization of small RNAs involved in RNA silencing. *Febs Letters*, 579(26): 5830–40.
- Bachem, C.W.B., Van der Hoeven, R.S., De Bruijn, S.M., Vreugdenhil, D., Zabeau, M. y Visser, R.G.F. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analyses of gene expression during potato tuber development. *The Plant Journal*, 9: 745–753.
- Bamshad, M. y Wooding, S.P. 2003. Signatures of natural selection in the human genome. *Nature Reviews Genetics*, 4(2): 99–111. Review.
- Baumung, R., Simianer, H. y Hoffmann, I. 2004. *Genetic diversity studies in farm animals – a survey*, *Journal of Animal Breeding and Genetics*, 121: 361–373.
- Beaumont, M.A. y Balding, D.J. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, 13(4): 969–80.
- Beja-Pereira, A., Alexandrino, P., Bessa, I., Carretero, Y., Dunner, S., Ferrand, N., Jordana, J., Laloe, D., Moazami-Goudarzi, K., Sanchez, A. y Cañon, J. 2003. Genetic characterization of southwestern European bovine breeds: a historical and biogeographical reassessment with a set of 16 microsatellites. *Journal of Heredity*, 94: 243–50.
- Berthier, D., Quere, R., Thevenon, S., Belemsaga, D., Piquemal, D., Marti, J. y Maillard, J.C. 2003. Serial analysis of gene expression (SAGE) in bovine trypanotolerance: preliminary results. *Genetics Selection Evolution*, 35 (Supl. 1): S35–47.
- Bertone, P, Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., Gerstein, M. y Snyder, M. 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science*, 306: 2242–2246.

## PARTE 4

- Black, W.C., Baer, C.F., Antolin, M.F. y DuTeau, N.M. 2001. Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, 46: 441–469.
- Bruford, M.W., Bradley, D.G. y Luikart, G. 2003. DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, 4: 900–910.
- Buntjer, J.B., Otsen, M., Nijman, I.J., Kuiper, M.T. y Lenstra, J.A. 2002. Phylogeny of bovine species based on AFLP fingerprinting. *Heredity*, 88: 46–51.
- Campbell, D. y Bernatchez, L. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, 21(5): 945–56.
- Cañon, J., García, D., García-Atance, M.A., Obexer-Ruff, G., Lenstra, J.A., Ajmone-Marsan, P., Dunner, S. y The ECONOGENE Consortium. 2006. Geographical partitioning of goat diversity in Europe and the Middle East. *Animal Genetics*, 37: 327–334.
- Chen, S.Y., Su, Y.H., Wu, S.F., Sha, T. y Zhang, Y.P. 2005. Mitochondrial diversity and phylogeographic structure of Chinese domestic goats. *Molecular Phylogenetics and Evolution*, 37: 804–814.
- Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H. y Nielsen, R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, 15: 1496–1502.
- Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibe, B., Bouix, J., Caiment, F., Elsen, J.M., Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C. y Georges, M. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics*, 38: 813–818.
- De Marchi, M., Dalvit, C., Targhetta, C. y Cassandro, M. 2006. Assessing genetic diversity in indigenous Veneto chicken breeds using AFLP markers. *Animal Genetics*, 37: 101–105.
- Donson, J., Fang, Y., Espiritu-Santo, G., Xing, W., Salazar, A., Miyamoto, S., Armendarez, V. y Volkmuth, W. 2002. Comprehensive gene expression analysis by transcript profiling. *Plant Molecular Biology*, 48: 75–97.
- Excoffier, L., Smouse, P.E. y Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131: 479–491.
- Farnir, F., Grisart, B., Coppieters, W., Riquet, J., Berzi, P., Cambisano, N., Karim, L., Mni, M., Moisis, S., Simon, P., Wagenaar, D., Vilkki, J. y Georges, M. 2002. Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics*, 161: 275–287.
- Fields, S. y Song, O. 1989. A novel genetic system to detect protein–protein interactions. *Nature*, 340: 245–246.
- Freeman, A.R., Bradley, D.G., Nagda, S., Gibson, J.P. y Hanotte, O. 2006. Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. *Animal Genetics*, 37: 1–9.
- Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L. y Feldman, M.W. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 139: 463–471.
- Goldstein, D.B. y Schlötterer, C. 1999. *Microsatellites: evolution and applications*. Nueva York, EE.UU. Oxford University Press.

- Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. y Snell, R. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research*, 12: 222–231.
- Guo, J., Du, L.X., Ma, Y.H., Guan, W.J., Li, H.B., Zhao, Q.J., Li, X. y Rao, S.Q. 2005. A novel maternal lineage revealed in sheep (*Ovis aries*). *Animal Genetics*, 36: 331–336.
- Haley, C. y de Koning, D.J. 2006. Genetical genomics in livestock: potentials and pitfalls. *Animal Genetics*, 37(Suppl 1): 10–12.
- Hanotte, O., Bradley, D.G., Ochieng, J.W., Verjee, Y. y Hill, E.W. 2002. African pastoralism: genetic imprints of origins and migrations. *Science*, 296: 336–339.
- Hayes, B.J., Visscher, P.M., McPartlan, H.C. y Goddard, M.E. 2003. A novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, 13: 635–643.
- Hill, E.W., O’Gorman, G.M., Agaba, M., Gibson, J.P., Hanotte, O., Kemp, S.J., Naessens, J., Coussens, P.M. y MacHugh, D.E. 2005. Understanding bovine trypanosomiasis and trypanotolerance: the promise of functional genomics. *Veterinary Immunology and Immunopathology*, 105: 247–258.
- Hill, W.G. 1981. Estimation of effective population size from data on linkage disequilibrium. *Genetics Research*, 38: 209–216.
- Hillel, J., Groenen, M.A., Tixier-Boichard, M., Korol, A.B., David, L., Kirzhner, V.M., Burke, T., Barre-Dirie, A., Crooijmans, R.P., Elo, K., Feldman, M.W., Freidlin, P.J., Maki-Tanila, A., Oortwijn, M., Thomson, P., Vignal, A., Wimmers, K. y Weigend, S. 2003. Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. *Genetics Selection Evolution*, 35: 533–557.
- Hood, L., Heath, J.R., Phelps, M.E. y Lin, B. 2004. Systems biology and new technologies enable predictive and preventative medicine. *Science*, 306: 640–643.
- Ibeagha-Awemu, E.M., Jann, O.C., Weimann, C. y Erhardt, G. 2004. Genetic diversity, introgression and relationships among West/Central African cattle breeds. *Genetics Selection Evolution*, 36: 673–690.
- Jarne, P. y Lagoda, P.J.L. 1996. Microsatellites, from molecules to populations and back. *Tree*, 11: 424–429.
- Joshi, M.B., Rout, P.K., Mandal, A.K., Tyler-Smith, C., Singh, L. y Thangaraj, K. 2004. Phylogeography and origin of Indian domestic goats. *Molecular Biology and Evolution*, 21: 454–462.
- Joost, S. 2006. *The geographical dimension of genetic diversity. A GIScience contribution for the conservation of animal genetic resources*. École Polytechnique Fédérale de Lausanne, Suiza. (Tesis de doctorado.)
- Kayser, M., Brauer, S. y Stoneking, M. 2003. A genome scan to detect candidate regions influenced by local natural selection in human populations. *Molecular Biology and Evolution*, 20: 893–900.
- Lai, S.J., Liu, Y.P., Liu, Y.X., Li, X.W. y Yao, Y.G. 2006. Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loop sequence variation. *Molecular Phylogenetics and Evolution*, 38: 146–54.
- Lan, L., Chen, M., Flowers, J.B., Yandell, B.S., Stapleton, D.S., Mata, C.M., Ton-Keen Mui, E., Flowers, M.T., Schueler, K.L., Manly, K.F., Williams, R.W., Kendziorski, C. y Attie, A.D. 2006. Combined expression trait correlations and expression quantitative trait locus mapping. *PLoS Genetics*, 2: 51–61.
- Liang, P. y Pardee, A.B. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 257: 967–997.

## PARTE 4

- Liu, Y.P., Wu, G.S., Yao, Y.G., Miao, Y.W., Luikart, G., Baig, M., Beja-Pereira, A., Ding, Z.L., Palanichamy, M.G. y Zhang, Y.P. 2006. Multiple maternal origins of chickens: out of the Asian jungles. *Molecular Phylogenetics and Evolution*, 38: 12–19.
- Lueking, A., Possling, A., Huber, O., Beveridge, A., Horn, M., Eickhoff, H., Schuchardt, J., Lehrach, H. y Cahill, D.J. 2003. A nonredundant human protein chip for antibody screening and serum profiling. *Molecular and Cellular Proteomics*, 2: 1342–1349.
- Luikart, G., England, P.R., Tallmon, D., Jordan, S. y Taberlet, P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4: 981–994.
- Luikart, G., Gielly, L., Excoffier, L., Vigne, J.D., Bouvet, J. y Taberlet, P. 2001. Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Science USA*, 98: 5927–5932.
- Mburu, D.N., Ochieng, J.W., Kuria, S.G., Jianlin, H. y Kaufmann, B. 2003. Genetic diversity and relationships of indigenous Kenyan camel (*Camelus dromedarius*) populations: implications for their classification. *Animal Genetics*, 34(1): 26–32.
- McPherron, A.C. y Lee, S.J. 1997. Double muscling in cattle due to mutations in the myostatin gene. *Proceedings of the National Academy of Science USA*, 94: 12457–12461.
- Negrini, R., Milanesi, E., Bozzi, R., Pellecchia, M. y Ajmone-Marsan, P. 2006. Tuscany autochthonous cattle breeds: an original genetic resource investigated by AFLP markers. *Journal of Animal Breeding and Genetics*, 123: 10–16.
- Nei, M. 1972. Genetic distance between populations. *The American Naturalist*, 106: 283–292.
- Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89: 583–590.
- Nei, M. y Roychoudhury, A.K. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics*, 76: 379–390.
- Nei, M., Tajima, F. y Tateno, Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution*, 19: 153–170.
- Nielsen, R. y Signorovitch, J. 2003. Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology*, 63: 245–55.
- Nijman, I.J., Otsen, M., Verkaar, E.L., de Ruijter, C. y Hanekamp, E. 2003. Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. *Heredity*, 90: 10–16.
- Pariset, L., Cappuccio, I., Joost, S., D'Andrea, M.S., Marletta, D., Ajmone Marsan, P., Valentini A. y ECONOGENE Consortium 2006. Characterization of single nucleotide polymorphisms in sheep and their variation as an evidence of selection. *Animal Genetics*, 37: 290–292.
- Pritchard, J.K., Stephens, M. y Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–959.
- Rabie, T.S., Crooijmans, R.P., Bovenhuis, H., Vereijken, A.L., Veenendaal, T., van der Poel, J.J., Van Arendonk, J.A., Pakdel, A. y Groenen, M.A. 2005. Genetic mapping of quantitative trait loci affecting susceptibility in chicken to develop pulmonary hypertension syndrome. *Animal Genetics*, 36: 468–476.

- Saitou, N. y Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4: 406–425.
- Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S. y Altshuler, D.; International SNP Map Working Group. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409: 928–933.
- SanCristobal, M., Chevalet, C., Haley, C.S., Joosten, R., Rattink, A.P., Harlizius, B., Groenen, M.A., Amigues, Y., Boscher, M.Y., Russell, G., Law, A., Davoli, R., Russo, V., Desautes, C., Alderson, L., Fimland, E., Bagga, M., Delgado, J.V., Vega-Pla, J.L., Martinez, A.M., Ramos, M., Glodek, P., Meyer, J.N., Gandini, G.C., Matassino, D., Plastow, G.S., Siggens, K.W., Laval, G., Archibald, A.L., Milan, D., Hammond, K. y Cardellino, R. 2006a. Genetic diversity within and between European pig breeds using microsatellite markers. *Animal Genetics*, 37: 189–198.
- SanCristobal, M., Chevalet, C., Peleman, J., Heuven, H., Brugmans, B., van Schriek, M., Joosten, R., Rattink, A.P., Harlizius, B., Groenen, M.A., Amigues, Y., Boscher, M.Y., Russell, G., Law, A., Davoli, R., Russo, V., Desautes, C., Alderson, L., Fimland, E., Bagga, M., Delgado, J.V., Vega-Pla, J.L., Martinez, A.M., Ramos, M., Glodek, P., Meyer, J.N., Gandini, G., Matassino, D., Siggens, K., Laval, G., Archibald, A., Milan, D., Hammond, K., Cardellino, R., Haley, C. y Plastow, G. 2006b. Genetic diversity in European pigs utilizing amplified fragment length polymorphism markers. *Animal Genetics*, 37: 232–238.
- Sauer, S., Lange, B.M.H., Gobom, J., Nyarsik, L., Seitz, H. y Lehrach, H. 2005. Miniaturization in functional genomics and proteomics. *Nature Reviews Genetics*, 6: 465–476.
- Sodhi, M., Mukesh, M., Mishra, B.P., Mitkari, K.R., Prakash, B. y Ahlawat, S.P. 2005. Evaluation of genetic differentiation in *Bos indicus* cattle breeds from Marathwada region of India using microsatellite polymorphism. *Animal Biotechnology*, 16: 127–137.
- Storz, G., Altuvia, S. y Wassarman, K.M. 2005. An abundance of RNA regulators. *Annual Review of Biochemistry*, 74: 199–217.
- Sunnucks, P. 2001. Efficient genetic markers for population biology. *Tree*, 15: 199–203.
- Syvänen, A.C. 2001. Accessing genetic variation genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, 2: 930–941.
- Takezaki, N. y Nei, M. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, 144: 389–399.
- Tapio, M., Tapio, I., Grisliis, Z., Holm, L.E., Jeppsson, S., Kantanen, J., Miceikiene, I., Olsaker, I., Viinalass, H. y Eythorsdottir, E. 2005. Native breeds demonstrate high contributions to the molecular variation in northern European sheep. *Molecular Ecology*, 14: 3951–3963.
- Tilquin, P., Barrow, P.A., Marly, J., Pitel, F., Plisson-Petit, F., Velge, P., Vignal, A., Baret, P.V., Bumstead, N. y Beaumont, C. 2005. A genome scan for quantitative trait loci affecting the *Salmonella* carrier-state in the chicken. *Genetics Selection Evolution*, 37: 539–61.
- Troy, C.S., MacHugh, D., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C. y Bradley D.G. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410: 1088–1091.

## PARTE 4

- Velculescu, V.E., Vogelstein, B. y Kinzler, K.W. 2000. Analyzing uncharted transcriptomes with SAGE. *Trends in Genetics*, 16: 423–425.
- Velculescu, V.E., Zhang, L., Vogelstein, B. y Kinzler, K.W. 1995. Serial analysis of gene expression. *Science*, 270: 484–487.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J. y Kuiper, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23: 4407–1444.
- Weir, B.S. y Basten, C.J. 1990. Sampling strategies for distances between DNA sequences. *Biometrics*, 46: 551–582.
- Weir, B.S. y Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, 38: 1358–1370.
- Wienholds, E. y Plasterk, R.H. 2005. MicroRNA function in animal development. *FEBS Letters*, 579: 5911–5922.
- Wong, G.K., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z., Meng, Q., Zhou, J., Li, D., Zhang, J., Ni, P., Li, S., Ran, L., Li, H., Zhang, J., Li, R., Li, S., Zheng, H., Lin, W., Li, G., Wang, X., Zhao, W., Li, J., Ye, C., Dai, M., Ruan, J., Zhou, Y., Li, Y., He, X., Zhang, Y., Wang, J., Huang, X., Tong, W., Chen, J., Ye, J., Chen, C., Wei, N., Li, G., Dong, L., Lan, F., Sun, Y., Zhang, Z., Yang, Z., Yu, Y., Huang, Y., He, D., Xi, Y., Wei, D., Qi, Q., Li, W., Shi, J., Wang, M., Xie, F., Wang, J., Zhang, X., Wang, P., Zhao, Y., Li, N., Yang, N., Dong, W., Hu, S., Zeng, C., Zheng, W., Hao, B., Hillier, L.W., Yang, S.P., Warren, W.C., Wilson, R.K., Brandstrom, M., Ellegren, H., Crooijmans, R.P., van der Poel, J.J., Bovenhuis, H., Groenen, M.A., Ovcharenko, I., Gordon, L., Stubbs, L., Lucas, S., Glavina, T., Aerts, A., Kaiser, P., Rothwell, L., Young, J.R., Rogers, S., Walker, B.A., van Hateren, A., Kaufman, J., Bumstead, N., Lamont, S.J., Zhou, H., Hocking, P.M., Morrice, D., de Koning, D.J., Law, A., Bartley, N., Burt, D.W., Hunt, H., Cheng, H.H., Gunnarsson, U., Wahlberg, P., Andersson, L., Kindlund, E., Tammi, M.T., Andersson, B., Webber, C., Ponting, C.P., Overton, I.M., Boardman, P.E., Tang, H., Hubbard, S.J., Wilson, S.A., Yu, J., Wang, J., Yang, H.; International Chicken Polymorphism Map Consortium. 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature*, 432: 717–722.
- Zhu, H., Bilgin, M. y Snyder, M. 2003. Proteomics. *Annual Review of Biochemistry*, 72: 783–812.