

Capítulo III

Lineamientos para la caracterización de la diversidad de los recursos genéticos animales

Luis Telo da Gama

Facultad de Medicina Veterinaria de la Universidad de Lisboa, Portugal.

Carlos Edmundo Lucero Casanova

Investigador PhD de AGROSAVIA. Centro de Investigación Tibaitatá.



Conservación de la diversidad genética

De acuerdo con la teoría evolucionista, las poblaciones con baja variabilidad genética tienen menor potencial adaptativo (por tanto, menor capacidad de acoplarse a los cambios ambientales) que las poblaciones con elevados niveles de variabilidad genética. En consecuencia, preservar la diversidad genética es esencial para asegurar la supervivencia de las poblaciones a largo plazo, y eso implica conservar las distintas poblaciones y preservar la diversidad genética intrarracial a lo largo de las generaciones. Así, la diversidad genética de una población es importante como factor de adaptación a las fluctuaciones a corto plazo y a las alteraciones ambientales a largo plazo.

Cuando se considera la posible pérdida de diversidad genética de una población, el riesgo de extinción es normalmente el criterio más obvio, por lo que la conservación de varias poblaciones distintas es esencial para maximizar el potencial evolutivo de una especie y minimizar el riesgo de extinción a largo plazo. Desde un punto de vista general, los factores que pueden determinar la extinción de una población se pueden clasificar en dos grandes tipos:

1. Factores determinísticos, incluyendo la destrucción de los hábitats, la polución, la sobreexplotación, las alteraciones climáticas, etcétera.
2. Factores estocásticos, incluyendo la deriva genética, la consanguinidad, la proporción de sexos (agravados por la depresión consanguínea), etcétera.

Estos principios generales se aplican a las especies salvajes y a las especies de animales domésticos, así las posibles estrategias para preservar la diversidad genética sean distintas. La unidad biológica de trabajo en las especies domésticas es la raza y la conservación de la variabilidad genética debe tener en cuenta la diversidad entre razas y dentro de cada una de estas. En el caso particular de



Recursos *zoogenéticos*

las especies de animales domésticos, los principales factores que contribuyen a un mayor riesgo de extinción son los siguientes:

Económicos y de mercado.

- Políticas inadecuadas.
- Estrategias de conservación deficientes.
- Inestabilidad sociopolítica.
- Falta de apoyo institucional.
- Control sanitario deficiente.
- Dificultades para la mano de obra.
- Ambiente de producción degradado.
- Funcionalidad de las razas sustituidas (por ejemplo, ovinos de lana, razas gordas de cerdos, etcétera).

El riesgo de extinción de una raza se evalúa normalmente por el número de reproductores que existen, aunque otros criterios pueden y deben ser utilizados en esta evaluación del riesgo. De acuerdo con la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, 2011) y teniendo en cuenta el censo como criterio de evaluación, de las razas de animales domésticos que existen actualmente en el mundo, cerca de 1/3 están amenazadas y 1/3 tienen un estatus desconocido. Como consecuencia, muchos países están desarrollando programas de conservación de los recursos genéticos animales, que tienen como objetivo, en una primera etapa, preservar las razas que hay, bien sea por programas de conservación de germoplasma o por programas de conservación de rebaños en los productores.

Durante muchos años, la preocupación fundamental ha sido preservar la diversidad genética interracial y, por lo tanto, el enfoque ha sido evitar la pérdida de las razas en mayor riesgo de extinción. La pérdida de diversidad genética dentro de cada raza también es muy importante y no puede ser ignorada, aunque no haya sido siempre considerada de manera adecuada.



Indirectamente, la pérdida de variabilidad intrarracial resulta en una mayor debilidad de la raza y, a largo plazo, puede llevar a su extinción. La menor diversidad genética de una población se traduce en:

- Deriva genética (con fijación/pérdida de genes).
- Menor respuesta a la selección.
- Menor capacidad de adaptación.
- Depresión consanguínea.
- Mayor riesgo de extinción.

El número de reproductores de una raza tiene un impacto directo en su viabilidad, ya que afecta la evolución de la consanguinidad y, por lo tanto, de la diversidad genética. Cuando el censo de una raza se reduce, la consanguinidad se acumula y las consecuencias de la depresión consanguínea son inevitables; esto tiene un impacto sobre todo en las características asociadas con la supervivencia de la especie (mortalidad, fertilidad, prolificidad, etcétera). Como resultado, la tasa de reproducción de la raza se va reduciendo gradualmente, la población se presenta con un tamaño cada vez menor y la consanguinidad se vuelve más problemática. Esta cadena de eventos conduce al fenómeno conocido como *vórtice de extinción*, con el cual la secuencia censo reducido/consanguinidad/depresión consanguínea termina extinguiendo una raza.

Aspectos generales

La primera tarea de cualquier programa de gestión de los recursos genéticos animales (RGA) es caracterizar lo que existe y los factores que pueden condicionar el mantenimiento y uso de esos recursos. Solo después de esa caracterización es posible que se desarrollen programas de mejora genética o conservación de esos RGA, que pueden y deben complementarse. Genéricamente, la estrategia puede ser como se muestra en el figura 2.



Recursos *zoogenéticos*



Figura 2. Diagrama para la estructuración de programas de mejora genética o conservación.
Fuente: Elaboración propia

Las actividades propuestas en la figura 2 son complementarias y no necesariamente excluyentes; por ejemplo, un programa de selección debe complementarse con un programa de conservación para mantener la diversidad genética a largo plazo. De la misma forma, la selección y el cruzamiento pueden y deben complementarse, así como la conservación *in situ* y *ex situ*.

El desarrollo de las estrategias de gestión de los RGAⁿ tiene como punto de partida la caracterización de lo que existe, pues sin eso no hay forma de decidir cuáles son las opciones más adecuadas. Se puede considerar que la caracterización debe incidir en diversas perspectivas, que se resumen de la siguiente forma:



- Caracterización fenotípica
 - Morfológica.
 - Productiva.
- Caracterización del entorno productivo
 - Ambiental.
 - Socioeconómica.
- Caracterización genética
 - Demográfica.
 - Molecular.
 - Genómica.

A continuación, se detallan algunos ejemplos de acciones específicas para desarrollar en cada una de estas etapas de la caracterización de los RGA. Otros aspectos se pueden encontrar en la lista de referencias.

Caracterización fenotípica-morfológica

Desde el principio de la domesticación, los animales han sido seleccionados en condiciones ambientales específicas y con determinados objetivos, que han resultado en el desarrollo progresivo de características específicas (por ejemplo, la lana fina de los merinos, el gran tamaño de algunas razas bovinas usadas para trabajo o la adaptación a condiciones tropicales de los cebuínos). A finales del siglo XIX la selección artificial organizada terminó en modificaciones profundas de la capacidad productiva y, consecuentemente, de las características morfológicas de muchas razas. Es importante resaltar que varias razas han desarrollado un prototipo morfológico, que corresponde al “ideal” de la raza y que sirve de criterio, por ejemplo, para los animales que sean inscritos en un libro genealógico.



Recursos *zoogenéticos*

Hay una lista interminable de características morfológicas que pueden ser consideradas y que naturalmente son distintas de una especie a otra. Algunos ejemplos de indicadores físicos para considerar en el proceso de caracterización morfológica de una raza son los siguientes:

- Características mesurables
 - Dimensiones, pesos, ángulos, proporciones, entre otros.
- Características calificables
 - Coloración (de capa, cuernos, cascos, entre otros).
 - Perfil frontal (convexo).
 - Forma del cuerpo (longilíneo, brevilíneo, entre otros).
 - Forma de los cuernos.
 - Orientación de las orejas.
 - Pelo (largo, mediano, corto).
- Características de expresión binomial (presencia/ausencia)
 - Cuernos.
 - Marmellas.
 - Cresta.

Las características morfológicas de un grupo representativo de animales pueden ser analizadas de una forma descriptiva, para conseguir una caracterización morfométrica y cualitativa de la raza. En este caso, es importante conocer la media, la distribución, la variabilidad, las frecuencias, entre otros, como indicadores de la uniformidad o heterogeneidad de la raza.

También se puede hacer un análisis comparativo, con el objetivo de cuantificar la similitud o discrepancia de las razas analizadas. Para este fin, muchas veces se usan análisis multivariantes (componentes principales, análisis discriminante, distancias entre grupos, etcétera), que pueden ayudar a identificar razas más alejadas, además de discriminar o asignar correctamente a la raza de origen.



Caracterización fenotípica-productiva

Con el proceso de selección para distintos objetivos y la adaptación a condiciones muy diversas se han conseguido razas con capacidades productivas muy distintas. Un punto fundamental de la caracterización de los RGAn es el conocimiento profundo de sus capacidades productivas. Está claro que hay un número muy grande de características productivas que pueden ser evaluadas y que tienen distinta importancia según la especie y el sistema de producción. Por lo general, los principales tipos de características productivas para tener en cuenta son:

- Producción.
- Reproducción.
- Calidad de productos.
- Comportamiento.
- Adaptación.
- Funciones de beneficio indirecto.

Caracterización del entorno

Otro punto fundamental es el entorno, dada la importancia de conocer los factores que condicionan la “existencia” de una raza (considerados *sensu latu*) o que pueden comprometer su supervivencia. Para poder determinarlos, se mencionan los siguientes:

- Distribución de la raza
 - Distribución geográfica.
 - Especies/razas.
 - Sistemas de información geográfica.
- Productores
 - Número.
 - Dimensión.



Recursos *zoogenéticos*

- Dispersión.
- Distribución de las edades.
- Raza pura o cruce.
- Factores condicionantes de naturaleza ambiental
 - Clima
 - Temperatura.
 - Humedad.
 - Precipitación.
 - Nieve.
 - Viento.
 - Duración de luz.
 - Radiación solar.
 - Suelo/territorio
 - Elevación.
 - Declive.
 - pH suelo.
 - Superficie del territorio (piedras, desierto, pantano, etcétera).
 - Vegetación predominante.
 - Recursos alimentarios y agua
 - Disponibilidad cuantitativa y cualitativa.
 - Variación anual y estacional.
 - Competición con otras especies.
 - Riesgos sanitarios
 - Enfermedades predominantes.
 - Resistencia/susceptibilidad.
 - Predadores.
- Factores socioeconómicos
 - Caracterización socioeconómica
 - Usos principales de la raza (carne, leche, fibra, productos transformados, razas no productivas, etcétera).



- Productos y mercados a los que se dirige (exterior vs. autoconsumo).
- Nichos de mercado.
- Venta de reproductores.
- Capacidad de manejo
 - Sistemas de producción.
 - Nivel de confinamiento.
 - Protección climática.
 - Control de enfermedades.
 - Disponibilidad de alimento y agua.
 - Control reproductivo.
- Aspectos relacionados con el género
 - Toma de decisiones.
 - Responsabilidades de trabajo.

Caracterización genética por análisis genealógico

La caracterización genética de una raza por análisis de la información del pedigrí ofrece información sobre aspectos importantes y de gran utilidad que aportan en la identificación de cuellos de botella y en el desarrollo de programas de conservación. De manera global, los aspectos aportados por la caracterización son los siguientes:

- Demografía de la raza
 - Distribución de edades, intervalo generacional, eficiencia reproductiva.
- Profundidad de las genealogías
 - Grado de completitud y número de generaciones conocidas.
- Erosión genética
 - Consanguinidad, parentesco y censo efectivo.



Recursos *zoogenéticos*

- Cuellos de botella en la población
 - Contribuciones genéticas de fundadores y ascendentes.
- Estructura racial
 - Origen de los genes fundadores, explotaciones que producen reproductores e intercambio de animales.

Muchos de los análisis aquí descritos se pueden hacer con softwares desarrollados para este fin; por ejemplo, ENDOG, POPREP, PEDIG, EVA, RELAX2, etcétera. Todos los análisis parten de un archivo relativamente simple, con la siguiente estructura básica:

- Animal.
- Padre.
- Madre.
- Sexo.
- Fecha de nacimiento.
- Rebaño (grupo) de origen.
- Rebaño actual.

Desde el comienzo hay que asegurar la validez de los datos (compatibilidad de identificación, fiabilidad de las paternidades, fechas, sexos, entre otros). Este es el punto fundamental, y probablemente una de las mayores dificultades, en la gestión genealógica de una raza.

Los distintos análisis pueden agruparse en algunos grandes temas de estudio, los cuales se mencionan a continuación.

Indicadores demográficos

Los indicadores demográficos dan una primera mirada de la población con la que se está trabajando y permiten identificar algunas de las principales amenazas y factores de riesgo.



- Censo
 - Número de animales registrados (actual y evolución).
 - Tamaño de los efectivos, número de reproductores/rebaño/año.
- Distribución por edades
 - Edad de los machos y las hembras reproductores.
- Intervalo entre generaciones
 - Edad media de los padres cuando nacen los hijos que los sustituyen, calculada para las cuatro vías de selección (padres de machos, padres de hembras, madres de machos, madres de hembras).
- Precocidad y longevidad media
 - Edad media y dispersión del primer y último hijo.
- Número de descendientes/reproductores activos
 - Media y distribución para padres y madres.

Profundidad de las genealogías

El conocimiento del pedigrí condiciona la información que se puede obtener en un análisis genealógico; por ejemplo, si solo los padres son conocidos (o si los abuelos de uno de los lados son desconocidos), la consanguinidad calculada es 0. Por tanto, una mayor profundidad en el conocimiento de las genealogías permite obtener una información más fiable y coherente.

- Porcentaje de ascendientes conocidos en cada generación
 - Porcentaje de individuos con padre/madre, abuelo/abuela materno y paterno conocidos, entre otros.
- Número equivalente de generaciones completas
 - Número equivalente de generaciones que serían conocidas si estuvieran completas las genealogías.
- Integridad del pedigrí
 - Proporción de ascendientes conocidos hasta una determinada generación anterior.



Caracterización fenotípica-morfológica

La pérdida de diversidad genética es inevitable en una población finita y cerrada. Existen distintas formas de evaluar esto en una población dinámica, y cuando se usan datos genealógicos para este objetivo, los criterios están fundamentalmente relacionados con el parentesco y la consanguinidad, calculados a partir del pedigrí.

- Consanguinidad (F)
 - Media y distribución (por rebaño, año).
 - Evolución.
 - Proporción de apareamientos consanguíneos.
- Parentesco
 - Coeficiente de parentesco (a_{ij}).
 - Coascendencia ($= \frac{1}{2} a_{ij}$).
- Tasa de consanguinidad
 - Pérdida de heterocigosis en cada generación, expresada proporcionalmente en la heterocigosis todavía existente (1-F).
 - Evolución de la consanguinidad media por año y por generación.
- Censo efectivo de la población
 - Número de reproductores que, si tuvieran la estructura de una población ideal, darían origen a la tasa de consanguinidad observada.
- Desviación del apareamiento aleatorio
 - Estimada por la discrepancia (α) entre la consanguinidad observada y la que resultaría de la coascendencia media en la generación anterior $P(1-F_t) = (1-f_{t-1})(1-\alpha)$.

Cuellos de botella en la población y probabilidad de origen de los genes

Estos parámetros permiten evaluar la desigual contribución de los padres para la generación siguiente. De esta manera, es posible estimar el balance entre los fundadores en la contribución a lo largo de las generaciones, teniendo en



cuenta la selección realizada y la variación en el tamaño de las familias, con posible pérdida de genes resultante de cuellos de botella en el pedigrí.

El objetivo es identificar la contribución a la población actual (considerada como población de referencia).

- Fundadores.
- Ascendientes.
- Rebaños fundadores.
 - Número total y número efectivo
 - El número efectivo de fundadores/ascendientes corresponde al número de fundadores/ascendientes que, si todos tuvieran la misma contribución, daría origen a la misma diversidad genética observada en la población de referencia.
 - Contribuciones acumuladas.
- Representatividad del crY y mtDNA fundadores.
- Índice de conservación genética (GCI, por sus siglas en inglés); este traduce el equilibrio de la contribución de los distintos fundadores para un individuo.

Estructura racial

- Origen de los genes fundadores
 - Países, rebaños y animales fundadores.
- Número efectivo de rebaños que originan padres, abuelos paternos, bisabuelos.
- Estructura piramidal.
 - Cuantificación de los seleccionadores, multiplicadores y productores comerciales.
- Distancia genética entre rebaños (índice de fijación [F_{st}]) (Nei, 1972)
 - Calculada a partir de la relación de parentesco entre ellos.



Caracterización genética por análisis de marcadores moleculares

Emplear distintos tipos de marcadores genéticos puede ser útil en el estudio de la diversidad, la estructura, las relaciones y demás factores de las razas de animales domésticos. Cada tipo de marcador nos puede dar una información distinta, ya que la clase de transmisión y la neutralidad a la selección son diferentes.

Las aplicaciones de los principales tipos de marcadores se encuentran resumidas en la tabla 1.

Tabla 1. Aplicaciones de los principales tipos de marcadores

Marcadores	Transmisión	Inferencia
mtDNA	Línea materna	Especies ancestrales Lugares de domesticación Orígenes maternos Intrapoblación: censo efectivo, expansión Variación entre poblaciones/regiones
crY	Línea paterna	Especies ancestrales Orígenes paternos Introgresión paterna
Autosómicos	Mendeliana-biparental	Intrapoblación: • Medidas de diversidad • Historia de las razas • Variación adaptativa Interpoblaciones • Relaciones entre poblaciones

Fuente: Elaboración propia, a partir de datos de Lenstra et al. (2012)

Dependiendo del comportamiento que tengan frente a la selección, los marcadores genéticos de uso más comunes son:

- Neutros (microsatélites, marcadores uniparentales)



- Reflejan principalmente los efectos de la deriva genética y relaciones entre poblaciones (migración).
- No neutros (genes codificantes, poliformismo de nucleótido único [SNP])
 - Reflejan el efecto de la selección y (quizás en menor escala) de los otros factores.

De esta manera, los resultados con distintos marcadores no son los mismos, por lo que siempre hay que considerar el tipo de marcador que se está usando en los análisis de diversidad genética o de relación interracial.

Durante años se usaron marcadores microsatélites para la caracterización genética de especies domésticas, por su elevado polimorfismo, facilidad de genotipado y neutralidad, y codominancia. En la actualidad, se están empleando cada vez más los paneles de SNP de densidad variable, que permiten algunas inferencias adicionales, además de lo que se consigue con los microsatélites.

Independiente del tipo de marcador genético, lo fundamental es que sea polimórfico, esto es, que exista más de una forma del gen estudiado. Normalmente un gen es considerado polimórfico si el alelo más frecuente tiene $p < 0,99$.

Los parámetros básicos que pueden calcularse con los marcadores genéticos son:

- Genotipos
 - Frecuencias genotípicas.
 - Heterocigosis.
- Alelos
 - Número total de alelos/locus.
 - Frecuencias alélicas.



Recursos *zoogenéticos*

Después de estimar los parámetros, se estudia la diversidad genética, que puede hacerse entre razas o intraraza. Genéricamente, cuando se usan marcadores clásicos, las fases para este estudio son las siguientes:

1. Diversidad intrarracial
 - Variabilidad génica/alélica.
 - Variabilidad genotípica
 - Heterocigosis.
 - Consanguinidad.
 - Equilibrio.
2. Diversidad interracial
 - Relaciones entre poblaciones
 - Distancias genéticas.
 - Análisis filogenético.
 - Análisis de clúster.
 - Estructura poblacional
 - Origen.

Para cada uno de estos objetivos hay estrategias específicas que permiten conocer en mayor detalle los niveles de diversidad genética existentes. En una primera aproximación, consideraremos el uso de marcadores autosómicos neutros, como son los microsatélites. Más adelante consideraremos otros tipos de marcadores genéticos.

Diversidad genética intrarracial

En esta sección se tienen en cuenta los parámetros más comunes de evaluación de la diversidad genética intrarracial. Admitamos que p_i es la frecuencia del alelo i y n_{ij} es la frecuencia observada de individuos con el genotipo correspondiente a los alelos i y j .

- Heterocigosis



- Heterocigosis esperada en la población (H_e). Corresponde a la diversidad genética y es la probabilidad de que, en un locus, dos alelos elegidos al azar en la población sean diferentes:

H_e en un locus con dos alelos

$$H_e = 1 - (p^2 + q^2)$$

- Heterocigosis observada (H_o):

$$H_o = \frac{\sum_{ij} n_{ij}}{N}$$

- H_e en un locus k con varios (i) alelos:

$$H_{e_k} = 1 - \sum p_i^2$$

- H_e en un conjunto de L loci:

$$H_e = \frac{\sum_k H_{e_k}}{L}$$

- Contenido de información polimórfica de un locus (PIC):

$$PIC = 1 - \sum p_i^2 - \sum_{i,j} p_i^2 p_j^2$$

- Déficit de heterocigosis en una raza o un locus, conocido como consanguinidad molecular:

$$F = 1 - \frac{H_{obs}}{H_{esp}}$$

- Diversidad alélica

- Número total de alelos/locus

- Conteo directo:
- N.º medio de alelos/locus (n)
- Admitiendo k loci, con ai alelos:

$$a = \frac{\sum a_i}{k}$$

- Número efectivo de alelos (A_e)



Recursos *zoogenéticos*

- Número de alelos que, si tuvieran la misma frecuencia, darían origen a la misma H_e :

$$A_e = \frac{1}{\sum p_i^2} = \frac{1}{1 - H_e}$$

- Alelos privados:
 - Alelos presentes exclusivamente en una población.
- Riqueza alélica (R_d)
 - El número de alelos encontrado normalmente depende del tamaño de la muestra (contrariamente a la H_e).
 - Método de rarefacción (simulación) que estima el número de alelos que estaría presente en una población con tamaño definido.
 - Se calcula a partir del número de alelos que, si estuvieran presentes en una población, no estarían representados en una muestra de tamaño n .
- Alelos nulos
 - Resultan de una posible mutación en la secuencia flanqueante del microsatélite, lo cual no permite la hibridación del primer ejemplar en esa región. Aumenta artificialmente la frecuencia de genotipos homocigos. Se traduce en ausencia de equilibrio con la ley Hardy-Weinberg (en adelante H-W).

Test de equilibrio

Normalmente se parte del principio de que la población está en equilibrio de acuerdo con la ley H-W. Se puede testar si la población está realmente en equilibrio, si la distribución de genotipos está de acuerdo con la ley H-W; es posible, por ejemplo, que exista ventaja selectiva de uno de los genotipos.

- Test de χ^2
 - Comparar
 - Distribución observada de genotipos.



- Distribución esperada según la ley H-W.
- Diferencia entre valores observados y esperados
 - Grande, no hay equilibrio.
 - Pequeña, hay equilibrio.
- Cómo llevarlo a cabo
 - H_0 : población en equilibrio.
 - Calcular:
$$\chi^2 = \sum \frac{(O - E)^2}{E}$$
 - Comparar con valor crítico tabulado u obtener P-value.
- Pruebas alternativas
 - La prueba χ^2 no es muy fiable cuando la muestra es pequeña (cuando una de las celdas tiene valor esperado < 5).
- Alternativa recomendada: prueba exacta de Fisher.

Diversidad genética interracial

A menudo, en estudios de diversidad genética es importante cuantificar el grado de similitud o divergencia entre razas distintas, y hay varias formas de hacerlo. En este capítulo, solo se considerarán los métodos de uso más común, aunque existen otros que también pueden ser utilizados.

- Estadísticos F de Wright (o índices de fijación).
 - Miden el déficit de heterocigosis respecto a las proporciones esperadas, si la población se encontrara en equilibrio según la ley H-W.
 - Este déficit puede ser debido a:
 - Subdivisión de la población.
 - Forma de apareamiento.
 - Se obtienen y comparan los niveles de heterocigosis en los diferentes niveles:
 - Población global.
 - Subpoblación (raza).



Recursos zoogenéticos

- Individuo.
- Si se admite que tenemos dos subpoblaciones (s1 y s2), cada una con n individuos, dentro de una población global T, se pueden representar los correspondientes parámetros de diversidad genética como en la figura 3.

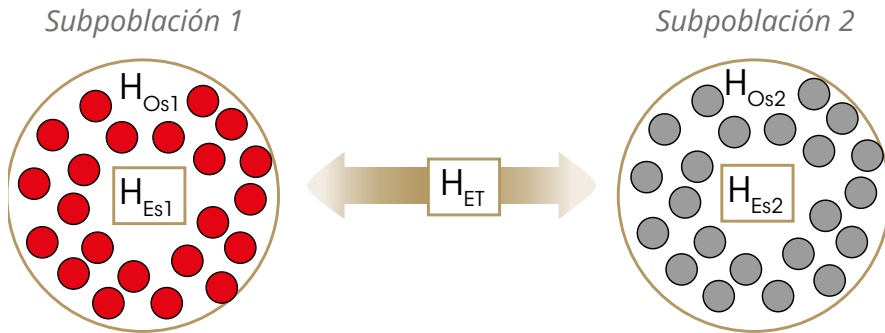


Figura 3. Representación gráfica de una metapoblación con dos subpoblaciones (razas).
Nota: Con las correspondientes heterocigosis esperadas (HE) y observadas (HO), en las subpoblaciones 1 y 2 (s1 y s2) y en el total (T). Cada círculo coloreado representa un individuo.
 Fuente: Elaboración propia

Además, se pueden calcular los parámetros, admitiendo un locus con h alelos, en el cual p_{s1i} es la frecuencia del alelo i en la subpoblación 1 y p_{Ti} es la frecuencia del mismo alelo en la población total.

- H_{Os1} = heterocigosis observada en la subpoblación 1.
- H_{Es1} = heterocigosis esperada en la subpoblación 1

$$H_{Es1} = 1 - \sum_{i=1}^h p_{s1i}^2$$

- H_{ET} = heterocigosis esperada en la población total

$$H_{ET} = 1 - \sum_{i=1}^h p_{Ti}^2$$

- Lo mismo para la subpoblación 2.



En este caso, se pueden considerar varios tipos de desviaciones:

- Individuos en relación con su subpoblación

$$H_{Os1} - H_{Es1} \qquad H_{Os2} - H_{Es2}$$

- Subpoblaciones en relación con la población total

$$H_{Es1} - H_{ET} \qquad H_{Es2} - H_{ET}$$

- Individuos en relación con la población total

$$H_{Os1} - H_{ET} \qquad H_{Os2} - H_{ET}$$

Para obtener estas desviaciones, hay que calcular las heterocigosis medias:

- Heterocigosis observada media en las dos subpoblaciones

$$\bar{H}_O = \frac{H_{Os1} + H_{Os2}}{2}$$

- Heterocigosis esperada media en las dos subpoblaciones

$$\bar{H}_E = \frac{H_{Es1} + H_{Es2}}{2}$$

- Heterocigosis total

$$H_{ET} = 1 - \sum_{i=1}^h p_{Ti}^2$$

Se puede descomponer la variabilidad (desviación de la heterocigosis) de la siguiente manera:

- F_{IS} = déficit de heterocigosis de los individuos en relación con su subpoblación

— Indica la consanguinidad media de las dos subpoblaciones; también puede ser efecto Whalund.

$$F_{IS} = \frac{\bar{H}_E - \bar{H}_O}{\bar{H}_E}$$

- F_{ST} = déficit de heterocigosis de las subpoblaciones en relación con la población total



Recursos *zoogenéticos*

— Indica el grado de diferenciación genética entre subpoblaciones

$$F_{ST} = \frac{H_{ET} - \bar{H}_E}{H_{ET}}$$

- F_{IT} = déficit de heterocigosis de los individuos en relación con la población total

— Refleja los efectos de la diferenciación genética y de la forma de apareamiento

$$F_{ST} = \frac{H_{ET} - \bar{H}_O}{H_{ET}}$$

Los distintos parámetros se relacionan como:

$$F_{IT} = F_{IS} + F_{ST}(F_{IS})(F_{ST})$$

- F_{ST} : es la forma común de comparar el grado de diferenciación entre poblaciones
 - Entre dos razas corresponde a la distancia F_{ST} entre ellas.
 - Entre n razas corresponde a la proporción de la diversidad genética explicada por las diferencias entre razas.
 - F_{ST} varía entre 0 y 1.
- $F_{ST} = 1 \Rightarrow$ cada raza está fijada para un alelo diferente.
- $F_{ST} = 0 \Rightarrow$ las razas tienen las mismas frecuencias alélicas
 - Algunas variaciones de F_{ST}
 - R_{ST} : tiene en cuenta la tasa de mutación de los loci microsatélites.
 - G_{ST} : considera la posibilidad de varios alelos.
 - G'_{ST} : corresponde al grado de diferenciación observado, ajustado para la diferenciación máxima posible.
 - θ (Weir and Cockerham): corrige para el posible sesgo de muestreo de las razas.
 - F_{ST} : es la forma común de comparar el grado de diferenciación entre poblaciones; normalmente, se considera $F_{ST} > 0,15$ como un elevado grado de diferenciación.



- F_{IS} : déficit de heterocigosis; por lo general, se interpreta como el nivel de consanguinidad de la población, pero también puede indicar subestructura de la población. Cuando $F_{IS} < 0$ indica que se está evitando la consanguinidad o que hay migración, con introducción de animales de fuera.
- Efecto Wahlund: traduce el déficit de heterocigosis que resulta de la subdivisión de una población (raza) en grupos más pequeños (por ejemplo, rebaños aislados).
 - La heterocigosis encontrada en una población subdividida es siempre $\leq H_e$, porque las líneas aisladas divergen por deriva genética.
 - Cuando hay subdivisión, la heterocigosis media observada en la raza global es:

$$\bar{H}_o = 2\bar{p}\bar{q} - 2\sigma_q^2$$

Donde σ_q^2 es la varianza de las frecuencias génicas entre las k subpoblaciones, i.e.

$$\sigma_q^2 = \frac{\sum (q_i - \bar{q})^2}{k}$$

- Análisis molecular de la varianza (AMOVA).

Algunas de las características del AMOVA son las siguientes:

- Es una extensión de las estadísticas F de Wright, que considera la distribución de la diversidad en un nivel jerárquico, con agrupamiento de subpoblaciones en distintos niveles, por ejemplo, dentro de una región, un continente, etcétera.
- En este caso, la variabilidad es considerada en niveles
 - Entre grupos.
 - Entre razas dentro de grupos.
 - Entre animales dentro de razas.



Recursos *zoogenéticos*

- Los componentes de varianza son estimados para los tres niveles, aplicando los principios del ANOVA jerárquica (*nested*) a datos moleculares.
 - Las pruebas de hipótesis son realizadas por permutación, lo que dispensa el supuesto de distribución normal.
 - En función de los componentes de varianza para cada factor en el modelo, se obtienen estimadores con interpretación similar a las estadísticas F, que corresponden a la proporción de la varianza explicada por cada factor, así:
 - ϕ_{GT} , entre grupos en la población total.
 - ϕ_{SG} , entre subpoblaciones dentro de los grupos.
 - ϕ_{ST} , entre subpoblaciones dentro de la población total.
- Distancias genéticas
 - En poblaciones genéticamente aisladas, la deriva genética y la mutación llevan a su progresivo distanciamiento, hasta la fijación de alelos en una u otra población.
 - La distancia genética permite una cuantificación de la divergencia genética entre poblaciones/individuos; la distancia entre dos poblaciones depende del tiempo transcurrido desde que se separaron.
 - Una distancia reducida entre poblaciones puede indicar una separación reciente o la existencia de flujo genético entre ellas.
 - La distancia genética normalmente es calculada con base en la proporción de alelos en común entre las poblaciones, pero la escala no es absoluta y las distancias son siempre relativas (unos grupos en relación con los otros).
 - Hay muchas propuestas alternativas de distancias genéticas y no hay una única que pueda calificarse como la mejor, ya que cada una depende de un modelo evolutivo distinto.
 - Diferentes distancias asumen distintos presupuestos:
 - Nei: deriva y mutación.



– Reynolds: solo deriva genética.

— Para calcular las principales distancias genéticas, se hacen algunos cálculos preliminares.

Admitamos:

Razas X e Y

Alelo u

r loci analizados

X_u = frecuencia alelo u en la raza X

Cálculos preliminares:

$$J_X = \frac{\sum_l \sum_u X_u^2}{r} \quad J_Y = \frac{\sum_l \sum_u Y_u^2}{r} \quad J_{XY} = \frac{\sum_l \sum_u X_u Y_u}{r}$$

• Árboles de distancias

— Después de obtener la matriz de distancias genéticas entre un grupo de razas, esta se puede convertir en un árbol representativo de dichas distancias. Estos árboles permiten visualizar de forma cuantitativa las relaciones entre razas y reconstituir su historia evolutiva.

— Los árboles pueden ser con o sin raíz. En el primer caso, se asume que hay un origen común a todas las razas y la distancia a la raíz traduce una trayectoria evolutiva; en el árbol sin raíz, hay solamente una representación de la distancia entre razas.

Los dos métodos más comunes de representación de los árboles de distancias son:

• *Unweighted pair-group method using arithmetic averages (UPGMA)*

— Método de grupos de pares no ponderados utilizando promedios aritméticos.

— Asume una tasa de evolución constante para los grupos; por consiguiente, las ramas son iguales para todos los grupos.

— La evolución constante implica existencia de una raíz.

• *Neighbour-joining*

— “Juntar vecinos”.



Recursos *zoogenéticos*

- Descomposición en estrella de árbol de distancias, minimizando el tamaño de las ramas (mínimo de alteración evolutiva).
- Los árboles pueden ser con o sin raíz y las ramas no tienen el mismo tamaño.
- El método *neighbor-joining* (NJ) es el más común en estudios con razas domésticas.

La robustez de un árbol se puede inferir usando la metodología *bootstrap*, con la cual se hace un remuestreo retirando un marcador a la vez, y se obtiene un árbol. El valor del *bootstrap* indica el porcentaje de ocasiones que un par de razas aparece en el mismo nodo.

Es recomendable introducir un *outgroup* en los análisis, especialmente si las razas analizadas están muy próximas, que sirve como punto de referencia externo y da una mayor estabilidad a los análisis.

Estructura de la población

Cuando se estudia la diversidad genética de una población, es importante conocer la posibilidad de ocurrencia de subestructura en esa población o la existencia de mezcla de poblaciones; para esto, existen diferentes alternativas, que se presentan a continuación.

Distancia entre individuos

Una forma posible de evaluar la existencia de subestructura o mezcla de poblaciones es obtener un árbol de distancias genéticas entre individuos, construido a partir de la proporción de alelos en común. Después de representar el árbol de distancias individuales, se analiza la tendencia a que algunos individuos de la misma raza se agrupen y constituyan subgrupos separados, o que se mezclen con individuos de otras razas. El hecho de que los animales de una determinada raza se sobrepongan



en el árbol con los de otra raza puede indicar algún grado de mezcla entre esas dos razas o una clasificación incorrecta de los animales. Estos resultados pueden ser interpretados junto con otro tipo de información que ayude a esclarecer el historial demográfico y selectivo de las razas analizadas.

Perspectiva bayesiana

En los últimos años, ha ganado mucha popularidad el análisis de la estructura poblacional usando la aproximación bayesiana desarrollada por Pritchard et al. (2003), y la cual ha sido incorporada al software Structure. En esta metodología, se admite que existe un conjunto de subpoblaciones (razas) en el presente que pueden (o no) tener un origen común en un grupo de poblaciones ancestrales, y en el que durante el proceso evolutivo puede, ocasionalmente, presentarse una mezcla entre esas poblaciones.

Structure emplea una metodología bayesiana para establecer clusters, utilizando la estimación por Monte Carlo Markov Chains (MCMC). El proceso se inicia admitiendo que, subyacente a las razas actuales, existe un número variable (K) de poblaciones ancestrales, que calcula la verosimilitud de que las frecuencias génicas actuales sean coherentes con el K asumido. Los cálculos se hacen de forma iterativa, testando valores crecientes de K . En cada nivel de K , se estima la probabilidad posterior de los datos para el nivel de K considerado; o sea, la verosimilitud del número de clusters considerados dada la diversidad genética observada. Los niveles crecientes de K son testados hasta que se llegue a un valor de la función de verosimilitud que ya no mejora, lo que indica haber alcanzado el nivel óptimo de K . En cada evaluación con un determinado valor de K , se calcula el coeficiente de pertenencia de cada animal de las poblaciones ancestrales, y después se puede obtener una media para el grupo de animales de determinada raza actual.



Recursos *zoogenéticos*

En la figura 4, se muestra una representación gráfica del planteamiento usado por el software Structure y se admite, en este caso, la contribución de tres poblaciones ancestrales para cuatro razas actuales. En este ejemplo, es claro que las razas A, B y C reciben una contribución casi exclusiva de las poblaciones ancestrales X, Y y Z, respectivamente. La raza D resulta del cruzamiento de animales con origen en las poblaciones ancestrales Y y Z, y aunque la contribución media de Y y Z para la raza D es bastante equilibrada, no todos los animales de la raza D tienen la misma contribución de las dos poblaciones ancestrales.

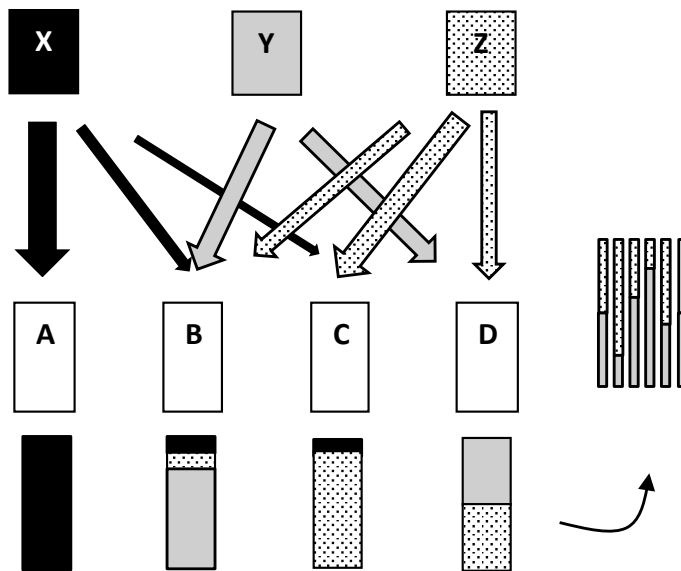


Figura 4. Contribución media de las poblaciones ancestrales X, Y y Z para las razas actuales A, B, C y D.

Nota: Se utilizó el software Structure. Para el caso de la raza D, se indica también la contribución para cada uno de los animales.

Fuente: Elaboración propia

Existen varias metodologías para estimar el K óptimo, aunque no hay un consenso sobre cuál es la mejor opción. De cualquier forma, la interpretación de los resultados obtenidos con niveles crecientes de K puede por sí misma



ser informativa. Por ejemplo, en un análisis de influencias genéticas en razas bovinas criollas, con $K=2$ se detectan las influencias en las razas actuales de las poblaciones ancestrales correspondientes a Taurus e Indicus; después con $K=3$ se separan las influencias ibéricas, británicas, etcétera (Ginja et al., 2019).

En resumen, la metodología basada en el algoritmo bayesiano de *clustering* implementado por el software Structure es extremadamente potente en estimar la distancia genética y la diferenciación entre razas, así como en identificar el grado de mezcla entre ellas, detectar la subestructura racial, permitir la exclusión de animales que se encuentran “contaminados”, asignar animales a razas, entre otras. Otros softwares han sido desarrollados con el mismo objetivo, aunque parecen ser más eficientes cuando se usa información de SNP (Admixture, Faststructure, Análisis Bayesiano de la Estructura Poblacional [BAPS, por sus siglas en inglés]).

Análisis factorial de correspondencias

El análisis factorial de correspondencias (AFC) es una metodología estadística que permite evaluar la distancia entre razas o individuos por su distribución espacial, y corresponde esencialmente a un análisis de componentes principales, aplicado en este caso a los datos de una tabla de contingencia. En un AFC, se extrae información de una tabla de contingencia en la que existe alguna relación de correspondencia entre las líneas y las columnas, donde la variabilidad inicial (inercia) es descompuesta; así, se identifica un número reducido de factores que justifican la desviación entre valores observados y esperados.

A partir de la información genotípica (o genómica) se calcula el grado de similitud genética entre pares de individuos y se obtienen ejes ortogonales de variación por combinaciones lineales de varios marcadores genéticos. Cuando los componentes principales se calculan de forma decreciente, reflejan la variabilidad genética que resulta de las diferentes contribuciones



Recursos *zoogenéticos*

ancestrales en la muestra; de esta manera, se interpreta que los individuos que tienen el mismo valor en un determinado componente principal tendrán la misma contribución en aquel eje.

El AFC ha sido usado para analizar información de marcadores genéticos, con el objetivo de encontrar la relación entre la diversidad genética observada y los factores subyacentes a los componentes principales que justifican esa diversidad. Estos componentes pueden, por ejemplo, traducir la distribución geográfica de las poblaciones en un territorio o el flujo de genes entre subpoblaciones. El AFC no tiene implicaciones particulares en lo que respecta a las presunciones genéticas, y esencialmente pretende encontrar patrones de distribución de la diversidad que pueden traducir la distancia genética entre poblaciones, la existencia de subestructura y la posible mezcla entre grupos. Para eso, el AFC permite visualizar la distribución de las observaciones (tanto de individuos como de los valores centrales de las poblaciones) en un espacio bidimensional o tridimensional, en función de los diferentes componentes principales. Es una metodología muy usada para identificar posibles patrones de distribución de la variabilidad genética. Es importante tener en cuenta la proporción de la variabilidad explicada por cada uno de los componentes principales.

En la figura 5, se muestra un ejemplo que representa la distribución de las observaciones después de un AFC basado en la información de marcadores genéticos en un grupo de animales pertenecientes a tres razas. Los resultados indican que el componente principal 1 (eje horizontal) separa bien la raza A de las razas B y C, pero estas dos están poco diferenciadas, posiblemente por mezcla entre ellas. Habría que investigar qué factores podrían estar subyacentes al componente 1 y a la menor diferenciación B/C (geografía, historial de las razas, etcétera).



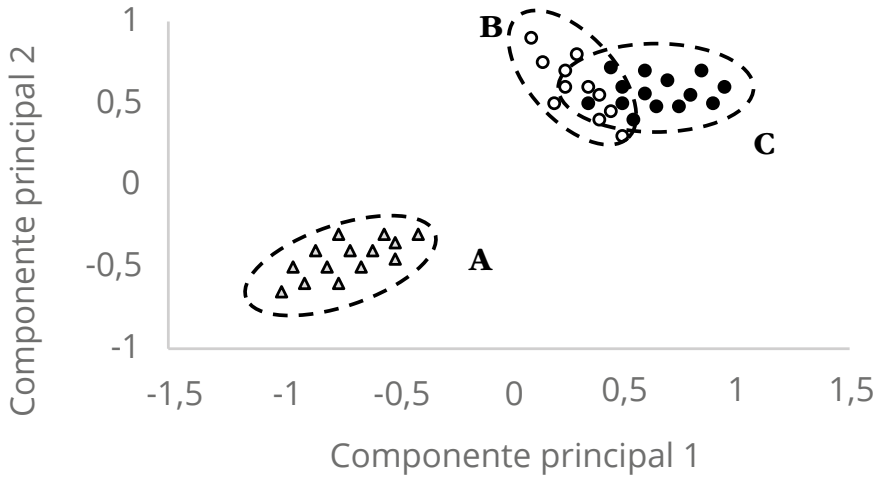


Figura 5. Resultados de un APC, incluyendo animales de razas A, B y C genotificados con marcadores moleculares.
Fuente: Elaboración propia

Caracterización genética por análisis genómico

En los últimos años, los enormes avances conseguidos en la secuenciación del genoma de diversas especies han permitido el desarrollo de paneles de SNP de densidad variable, que pueden traer grandes beneficios a los programas de caracterización de la diversidad genética. En otros capítulos de esta publicación, se presentan datos detallados acerca de la utilización de la información genómica con este objetivo, por lo que en el presente capítulo se abordan brevemente algunos de los aspectos principales de dicha metodología.



Recursos *zoogenéticos*

Relación de parentesco

La primera gran ventaja de los marcadores SNP es que permiten cuantificar de una forma objetiva el grado de similitud (parentesco) que existe entre dos individuos. Mientras que en la relación de parentesco genealógica lo que se calcula es el valor esperado de la similitud entre animales (por ejemplo, 1/4 entre medio-hermanos), en el caso de la relación de parentesco genómica, la similitud es medida por la proporción de alelos que los individuos comparten. Por tanto, esta metodología ha suplantado el parentesco convencional, tanto en programas de selección como de caracterización/conservación.

Estructura poblacional

De la misma manera que con los marcadores convencionales, la información obtenida con los marcadores SNP puede usarse para estudiar la estructura de una población, por ejemplo, haciendo un análisis AFC o el abordaje bayesiano implementado por Structure. Los análisis con paneles de SNP permiten un nivel de detección de diversidad bastante más refinado que con los marcadores clásicos, por lo que los SNP suponen una alternativa muy interesante para estudiar la estructura poblacional.

Segmentos de homocigosis

Los segmentos más o menos largos de homocigosis que se encuentran en el genoma se interpretan como resultado de apareamientos consanguíneos que han dado origen al individuo. Los segmentos largos de ROH representan la consanguinidad reciente, mientras que los segmentos cortos corresponden a una consanguinidad más antigua, y globalmente el coeficiente de consanguinidad de un individuo puede ser calculado como la proporción de su genoma, que está en segmentos ROH.



La consanguinidad estimada a partir de los registros genealógicos está correlacionada con la consanguinidad estimada por ROH, pero la correlación no es 1 como consecuencia, entre otras razones de errores en la genealogía; por ejemplo, porque en el primer caso estamos considerando el valor esperado de homocigosis, mientras que en el segundo tenemos el valor observado. Las correlaciones son más altas cuando se usan segmentos más largos de ROH para estimar la consanguinidad.

Desequilibrio de ligamiento

Los loci que están próximos en el genoma tienden a segregar conjuntamente, y de esta manera generan el llamado desequilibrio de ligamiento (*linkage disequilibrium* [LD]). Este LD se reduce con el tiempo, ya que el *crossing-over* separa los segmentos que están ligados. Después de un periodo, el LD tiende a ser menor, y lo mismo sucede si se utiliza un número de reproductores más alto. Por tanto, la forma como están ligados los loci más próximos o más alejados en el genoma es consecuencia de cuellos de botella en el pasado; por esta razón, el LD se puede usar para estimar de forma retrospectiva el censo efectivo en generaciones pasadas.

Distancia entre razas

Los valores de LD también se pueden comparar entre razas y, a partir de estos resultados, estimar las distancias genéticas entre esas razas (admitiendo que tienen un origen común). Además, la información genómica también sirve para estimar los estadísticos F , como ya se expuso. En el caso de paneles de SNP, es posible no solo obtener la distancia F_{ST} entre razas para todo el genoma, sino también locus por locus, lo que puede ser importante para identificar huellas de selección en determinadas regiones del genoma.



Genética del paisaje

Es importante considerar que las diferencias entre razas en determinados puntos del genoma se pueden asociar a la presencia de ciertas razas que están en puntos geográficos específicos y que pueden indicar que las condiciones ambientales en las que se encuentran los animales han permitido la selección a favor de determinados alelos que son beneficiosos para las condiciones consideradas (temperatura, humedad, etcétera). En estas circunstancias, es importante investigar los loci involucrados y los mecanismos fisiológicos de adaptación.

Referencias

- Ginja, C., Gama, L. T., Cortés, O., Martín Burriel, I., Vega-Pla, J. L., Penedo, C., Sponenberg, P., Cañón, J. Sanz, A. Alves do Egito, A., Álvarez, L. A., Giovambattista, G., Agha, S., Rogberg-Muñoz, A. Cassiano Lara, M. A. C. Consortium, B., Delgado, J. V., & Martínez, A. (2019). The genetic ancestry of American Creole cattle inferred from uniparental and autosomal genetic markers. *Scientific Reports*, 9(1), 11486.
- Lenstra, J. A., Groeneveld, L. F., Eding, H., Kantanen, J., Williams, J. L., Taberlet, P., Nicolazzi, E. L., Sölkner, J., Simianer, H., Ciani, E., García, J. F., Bruford, M. W., Ajmone-Marsan, P., & Weigend, S. (2012). Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Animal Genetics*, 43(5), 483-502.
- Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, 106(949), 283-229.
- Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). (2011). *Molecular genetic characterization of animal genetic resources*.
- Pritchard, J. K., Stephens, M., & Falush, D. (2003). Inference of population structure using multilocus genotype data. *Genetics*, 164(4), 1567-1587.

