

Análisis de Estructura Poblacional de la Cabra Florida mediante Información Genealógica y Molecular

Florida Goat Population Structure Analysis by means of Genealogic and Molecular Information

Iñigo Chomón Elosua

Tutores:

Isabel Cervantes Navarro y Juan Pablo Gutiérrez García

Universidad Complutense de Madrid

Resumen

Se emplean herramientas informáticas para determinar la estructura de la población de un grupo de cabras de la raza Florida, a partir de datos de pedigrí y marcadores moleculares (microsatélite y SNP). Tras representar los datos, se observa cómo la población de cabras queda dividida en 7 grupos según el pedigrí, mientras que los marcadores SNP encuentran dos grupos. El análisis de los marcadores microsatélite no produjo ningún resultado concluyente.

Palabras clave: pedigrí, marcadores moleculares, estructura población.

Abstract

We used informatics tools to determine the structure of a Florida breed goat population, based on the pedigree and molecular marker (microsatellite and SNP) data. After plotting the results, we can see that the pedigree divides the population in 7 groups, while the SNP marker data finds two groups. The analysis of the microsatellite markers did not yield any conclusive results.

Keywords: pedigree, molecular markers, population structure.

Trabajo presentado en las XII Jornadas Complutenses, XI Congreso Nacional de Investigación en Ciencias de la Salud para Alumnos Pregraduados y XVI Congreso de Ciencias Veterinarias y Biomédicas.

Agradecimientos a Juan Manuel Serradilla, Antonio Molina, María Jesús Carabaño (Coordinadora del proyecto RTA2011-00108-C02-00, Genética de la tolerancia al estrés por calor en rumiantes de aptitud lechera) y a la Asociación ACRIFLOR, por la cesión de los datos empleados en este trabajo.

Introducción

Con el desarrollo de la genómica (ciencia que estudia los genes, sus funciones y las herramientas que los emplean), se han desarrollado métodos basados en el análisis de material genético individual que permiten determinar las relaciones de parentesco entre individuos de una o varias poblaciones, y/o realizar valoraciones genéticas (Mewissen, Hayes y Goddard, 2001). Esto tradicionalmente se ha realizado por medio del pedigrí, que registra el parentesco entre animales.

Los métodos de análisis molecular del ADN, permiten determinar las relaciones de parentesco de los animales sin tener que precisar de un registro de su ascendencia, lo que los vuelve especialmente útiles en especies donde el registro del pedigrí es difícil o no es muy empleado. Dos de estos métodos son los marcadores microsatélite (secuencias repetidas de nucleótidos que varían en longitud en función

del alelo) y los marcadores “single nucleotide polymorphism” o SNP (posiciones dentro de la cadena de ADN en la que un único nucleótido presenta variación entre individuos, encontrando dos alelos en las poblaciones de una especie).

En este trabajo comparamos la capacidad de los datos de pedigrí y marcadores microsatélite y SNP para determinar la estructura de la población de cabras de la raza Florida.

Materiales y métodos

Se utilizaron datos de marcadores microsatélite y SNP de 625 cabras de la raza Florida (538 hembras y 87 machos), provenientes de 21 ganaderías distintas del sur de España, nacidas entre 01/02/2000 y 01/02/2010. El archivo de pedigrí tiene 19888 animales (19373 hembras y 515

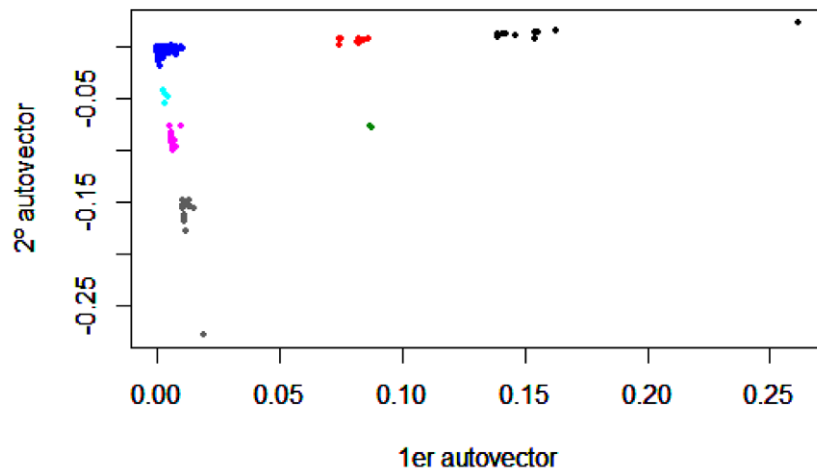


Figura 1. Representación gráfica de los autovectores 1 y 2 obtenidos utilizando la información contenida en la matriz A.;

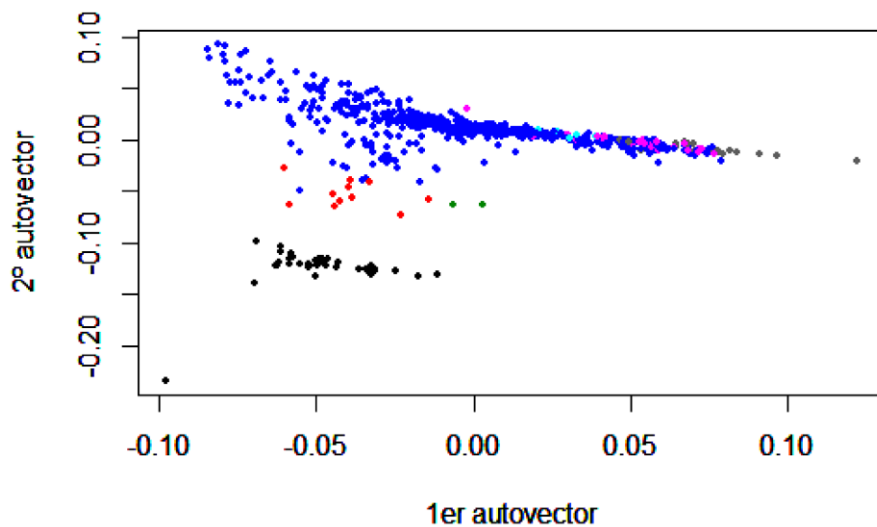


Figura 2. Representación gráfica de los autovectores 1 y 2 obtenidos utilizando la información contenida en la matriz G.

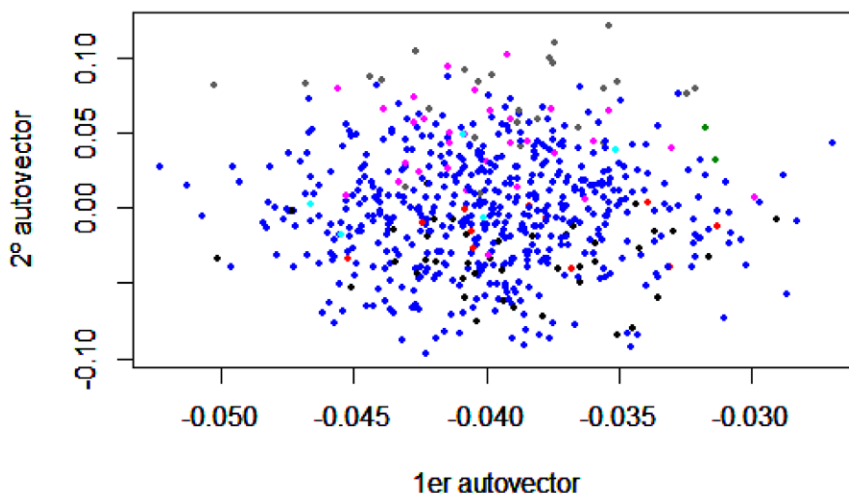


Figura 3. Representación gráfica de los autovectores 1 y 2 obtenidos utilizando la información contenida en la matriz *M*.

Tabla 1

Correlaciones entre los valores de la diagonal, triangular inferior y matrices completas *A*, *G* y *M*.

G-A diagonal	0.341	G-A triangular inferior	0.731	G-A matriz completa	0.729
G-M diagonal	0.189	G-M triangular inferior	0.337	G-M matriz completa	0.318
A-M diagonal	0.175	A-M triangular inferior	0.231	A-M matriz completa	0.222

machos), con los 625 antes mencionados incluidos en él, de 39 ganaderías distintas, nacidos entre el 25/04/1993 y el 01/03/2014.

Para obtener la matriz de coascendencias por pedigrí o matriz *A* se empleó el programa Endog (v4.8, Gutiérrez & Goyache, 2005), mientras que la matriz de coascendencias por microsatélites o matriz *M* se calculó empleando el programa Molkin (v2.0, Gutiérrez, Royo, Álvarez y Goyache, 2005) y la matriz *G* se obtuvo a partir de los datos de los SNP de los animales genotipados, siguiendo una variante del procedimiento diseñado por Van Raden en 2008 (Bouwman, Hayes y Calus, 2016), empleando el programa R (v 3.3.2, R Core Team, 2016).

Una vez se obtienen las matrices, se realiza una descomposición en componentes principales y se representan los dos primeros autovectores obtenidos en una gráfica. También se obtuvieron los valores de las correlaciones entre las matrices *A*, *M* y *G* mediante el programa R.

Resultados

En la Figura 1 se muestran los dos principales autovectores obtenidos de la matriz *A*, diferenciándose 7 subpoblaciones, que representan grupos de animales que están más emparentados entre sí. La población mayoritaria es la azul con 506 animales, mientras que la segunda más grande la población en color negro con 43 individuos.

La representación de los autovectores de la matriz *G* empleando marcadores SNP (Figura 2) nos separa dos subpoblaciones separadas por el segundo autovector (valor del eje $y = -0.08$). La subpoblación principal engloba 6 de los grupos diferenciados en la Figura 1 (estos 6 grupos están más próximos genéticamente) y cuenta con 582 animales, mientras que la segunda población cuenta con 43.

Por último, la gráfica de los autovectores obtenidos con la matriz *M* (marcadores microsatélite) ni mostró grupos claramente separados del resto de animales ni se vio una distribución que indique estructura de la población.

En la tabla 1 se muestran las correlaciones para los valores presentes entre las tres matrices. Observamos que las correlaciones más altas se obtuvieron para los valores entre *G* y *A*, siendo más bajas entre los datos contenidos en *G* y *M*, y entre *A* y *M*.

Discusión

A pesar de que las matrices *A*, *M* y *G* contienen los datos que reflejan las relaciones de parentesco entre distintos animales, la representación de los autovectores difiere mucho. Esto se debe al origen de los datos, la matriz *A* se compone a partir de registros de pedigrí, mientras que las matrices *G* y *M* representan las covarianzas entre los marcadores

genéticos (SNP y microsatélite respectivamente), de forma que separan a los animales de forma más realista en base a la información genética. Por esto llama la atención que la matriz **M** no consiga determinar si hay estructura de población. Esto puede deberse a que la información de microsatélites de los animales es insuficiente. Asimismo, las correlaciones nos muestran que las matrices **G** y **M** guardan poca similitud en sus datos, cuando se esperaba que fuesen más parecidas (ambas contienen información de marcadores genéticos).

Conclusiones

Se encuentra estructura de la población, siendo la información genómica contenida en la matriz **G** la más informativa de las fuentes empleadas.

Referencias

- Bouwman, A. C., Hayes, B. J., & Calus, M. P. L. (Agosto de 2016). *SNP-effects underlying GBLUP depend on the definition of the genomic relationship matrix*. Ponencia presentada en la 67^a Asamblea Anual de la Federación Europea de Ciencia Animal, Belfast, Reino Unido.
- Gutiérrez, J. P., & Goyache, F. (2005). A note on ENDOG: A computer program for analysing pedigree information. *Journal of Animal Breeding and Genetics*, 122(3), 172-176.
- Gutiérrez, J. P., Royo, L. J., Álvarez, I., & Goyache, F. (2005). MolKin v2.0: A computer program for genetic analysis of populations using molecular coancestry information. *Journal of Heredity*, 96(6), 718-721. <https://doi.org/10.1093/jhered/esi118>
- R Core Team (2016) [Software environment for statistical computing and graphics]. Vienna, Austria: R Foundation for Statistical Computing.