

ESTUDIO DE LA ESTRUCTURA POBLACIONAL DE CUATRO RAZAS OVINAS ESPAÑOLAS MEDIANTE UN CHIP DE SNPs

García-Gámez, E.¹, Gutiérrez-Gil, B.¹, Kijas, J.², Calvo, J.H.³, *International Sheep Genomics Consortium (ISGC)*, y Arranz J.J.¹

¹Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. ²CSIRO Livestock Industries (Australia). ³Unidad de Tecnología en Producción Animal, CITA, Zaragoza. E-mail: egarg@unileon.es

INTRODUCCIÓN

Como consecuencia de la era genómica han surgido una gran cantidad de herramientas que, en el ámbito de la genética animal, nos permiten analizar el genoma de las diferentes especies. Estas herramientas surgen como consecuencia de los diferentes proyectos de secuenciación de los genomas animales y ofrecen unas grandes posibilidades de utilización en la mejora genética. Entre ellas podemos destacar, la selección de los reproductores con mayor eficiencia y menor intervalo generacional (Hayes et al., 2009) y la detección de forma eficiente de genes que controlan caracteres de importancia económica. En este último caso la técnica más comúnmente empleada es, siguiendo los esquemas empleados en la especie humana, el análisis de asociación del genoma completo (GWAS, por sus siglas en inglés). En la especie humana se han realizado numerosos estudios de asociación, principalmente con enfermedades (<http://genome.gov/gwastudies/>) y muchas de ellas no han podido ser replicadas en estudios posteriores existiendo dudas sobre la veracidad de dichos hallazgos. Una de las causas fundamentales para la detección de falsos positivos en los estudios de asociación es el la presencia de subestructuras poblacionales en aquellas poblaciones que se analizan como homogéneas (estratificación poblacional) (Price et al., 2010). En los animales domésticos ya se han publicado numerosos estudios de asociación con caracteres de importancia económica (Becker et al., 2010; Snelling et al., 2010) y también se ha detectado el efecto de la subdivisión poblacional en poblaciones que en principio se creían muy homogéneas (y que pueden sesgar los resultados de asociación Decker et al., 2010). En el presente trabajo pretendemos estudiar la estructura genética de las razas españolas analizadas en el experimento *Sheep HapMap* como paso previo a posteriores estudios de asociación utilizando esta herramienta.

MATERIAL Y MÉTODOS

Los animales utilizados en este estudio, a través de la participación de nuestro grupo de investigación en el proyecto *Sheep HapMap*, fueron: la raza Churra (CHU) (tronco Churro), razas Castellana (CAS) y Rasa Aragonesa (RAA) (tronco Entrefino), Ojalada (OJA) (tronco Ibérico) y Awassi (AWA) (raza extranjera, procedente de Israel). Un total de 190 animales pertenecientes a las 5 razas, fueron seleccionados para el presente trabajo. Las distribución por raza fue la siguiente: CHU (n = 100), CAS (n = 23), RAA (n = 20), OJA (n = 24) y AWA (n = 23). De los 54.321 marcadores se han eliminado aquellos que no cumplían los requisitos de calidad: MAF (frecuencia del alelo raro) igual o inferior a 0,01, falta de ajuste a las condiciones del equilibrio Hardy Weinberg ($P < 4,5 \cdot 10^{-6}$) o presencia de inconsistencias en el modo de herencia. Entre los genotipos disponibles una vez aplicados estos filtros (49.034 SNPs), se seleccionaron dos subconjuntos. En el caso de análisis clúster con STRUCTURE se han elegido aquellos marcadores bien posicionados y con una distancia media entre ellos de 20 Mb y una representación de cada cromosoma equivalente a la existente en el *Ovine SNP50BeadChip* (Illumina, Inc), eligiéndose un total de 154 SNPs. Para el análisis con ADMIXTURE se han elegido aquellos marcadores que no muestran LD ($r^2 > 0,1$) en ventanas de 50 marcadores y se han utilizado un total de 11.869 SNPs. Las opciones disponibles en el software PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/>) se utilizaron para la selección de SNPs (--keep), el cálculo de las frecuencias alélicas (--freq), el test de equilibrio Hardy-Weinberg (--hardy), el cálculo del porcentaje de genotipos ausentes (--missing) y para la selección de marcadores para ADMIXTURE (--indep-pairwise). La medida de la distancia entre las razas estudiadas se llevó a cabo calculando el valor de F_{ST} por parejas de razas. Para dicho cálculo se utilizó el programa GENEPOP versión 4 (Raymond & Rousset, 1995).

El estudio de la estructura poblacional y la detección del número más probable de subpoblaciones, se realizaron mediante el software STRUCTURE versión 2.3 (Pritchard et

al., 2000). El análisis se realizó utilizando un modelo de no ligamiento, en el que se asume que ha habido mezcla entre las poblaciones de las que provienen los animales objeto de estudio, y suponiendo frecuencias alélicas independientes entre poblaciones. Para calcular el número de *clusters* (K), se utilizaron valores entre 1 y 10 con 5 repeticiones independientes del análisis. Cada una de las repeticiones, tuvo un periodo de *burn-in* de 20.000 iteraciones y, posteriormente, 100.000 réplicas para la obtención de resultados. El alineamiento de los resultados obtenidos en las repeticiones del análisis, se realizó con el programa CLUMPP 1.1.2 (Jakobsson & Rosenberg, 2007); posteriormente, el software DISTRUCT se utilizó para la visualización de los resultados. Los mismos cálculos se realizaron con el programa ADMIXTURE (Alexander et al., 2009) que realiza el mismo sistema de cálculo, pero es capaz de utilizar un mayor conjunto de datos.

RESULTADOS Y DISCUSIÓN

Los resultados del análisis de las frecuencias genotípicas de los SNPs seleccionados, nos aportan información acerca de las diferencias entre razas. El valor de MAF medio en el conjunto de animales seleccionados fue variable entre 0,1042 y 0,4974. Estudiando cada una de las razas por separado, podemos observar que existen SNPs en las razas Awassi (17 SNPs) y Rasa Aragonesa (1 SNP) para los que no encontramos las dos variantes alélicas.

El estudio de la distancia entre poblaciones se basó en el cálculo del valor de F_{ST} medio entre todos los *loci* para las razas utilizadas en este trabajo. Las razas más cercanas fueron Ojalada y Rasa Aragonesa, con un valor de F_{ST} de 0,0181; las más alejadas fueron Awassi y Rasa Aragonesa ($F_{ST} = 0,2037$). Como media, el valor del F_{ST} entre Awassi y cada una de las razas ibéricas estudiadas, fue un orden de magnitud mayor que el existente entre las razas ibéricas.

En la Figura 1 se muestran los resultados obtenidos en STRUCTURE v.2.3, representados a través del programa DISTRUCT. El cálculo de la verosimilitud en forma logarítmica, mostró un aumento entre los valores de $K = 1$ hasta $K = 6$ (Figura 1A). A partir de este valor de K , la representación del Ln de la probabilidad, alcanza una forma asintótica, no mejorando los valores de verosimilitud al aumentar K hasta 10, e incluso descendiendo ligeramente a partir de $K = 9$. Los resultados para el caso del programa ADMIXTURE son plenamente coincidentes con los anteriormente descritos. Este resultado muestra como un subconjunto de marcadores informativos distribuidos a lo largo del genoma puede ser plenamente informativo para los fines de clasificación racial en el ganado ovino.

En cuanto al momento de separación entre las razas, la Awassi, más alejada filogenéticamente del resto, es la primera que se separa cuando K toma el valor 2 (Figura 1B). Cuando aumentamos el valor de K , observamos que la raza Churra es la primera de las españolas en mostrar una diferenciación en subpoblaciones con respecto al resto, mostrando una clara estructura intrarracial. Este hecho nos indica que la estructura racial y familiar ha de ser tenida en cuenta en los análisis de asociación para buscar genes de importancia económica en esta raza ovina, en la que se basan la mayor parte de los estudios desarrollados por nuestro grupo de investigación. Además, la variabilidad dentro de esta raza es superior a la existente entre las otras razas ibéricas; este hecho puede ser debido al mayor tamaño muestral de la Churra ($n = 100$) respecto al resto. En el caso del resto de poblaciones, los resultados obtenidos muestran una elevada semejanza entre las tres de las cuatro razas españolas estudiadas, Castellana, Ojalada y Rasa Aragonesa, con una clara homogeneidad entre ellas, a pesar de su pertenencia a distintos troncos filogenéticos. Este estudio representa una primera aproximación, a través del uso de SNPs, para el estudio de la estructura poblacional en el ganado ovino de nuestro país.

REFERENCIAS BIBLIOGRÁFICAS

- Alexander, D. H., Novembre, J. & Lange, K. 2009. Genome Research 19:1655–1664.
- Becker, D., Tetens, J., Brunner, A., Bürstel, D., Ganter, M., Kijas, J., International Sheep Genomics Consortium & Drögemüller C. 2010. PLoS One. 5:e8689.
- Decker, J. E., Vasco, D. A., McKay, S. D., Rolf, M. M., Taxis, T. M., Chapple, R. H., Gregg, S. J., Kim, J. W., Schnabel, R. D. & Taylor, J. F. 2011. Plant Animal Genome XVII. P563.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. 2009. J. Dairy Sci. 92:433-443.
- Price, A. L., Zaitlen, N. A. Reich, D. & Patterson, N. 2010. Nat. Rev. Genet. 11:459-463.

Pritchard, J. K., Stephens, M. & Donnelly, P. 2000. *Genetics* 155:945-959. • Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J. & Sham, P. C. 2007. *Am. J. Hum. Genet.* 81:559-575. • Raymond, M. & Rousset, F., 1995. *J. Heredity* 86:248-249. • Rosenberg, N. A. 2004. *Mol. Ecol. Notes* 4:137-138. • Snelling, W. M., Allan, M. F., Keele, J. W., Kuehn, L. A., McDanel, T., Smith, T. P., Sonstegard, T. S., Thallman, R. M. & Bennett, G. L. 2010. *J. Anim. Sci.* 88:837-848.

Agradecimientos:

Este trabajo forma parte del proyecto AGL2009-07000 financiado por el Ministerio de Ciencia e Innovación. EGG es becaria FPU (Ministerio de Educación). BGG es investigadora contratada dentro del programa Juan de la Cierva (Ministerio de Ciencia e Innovación).

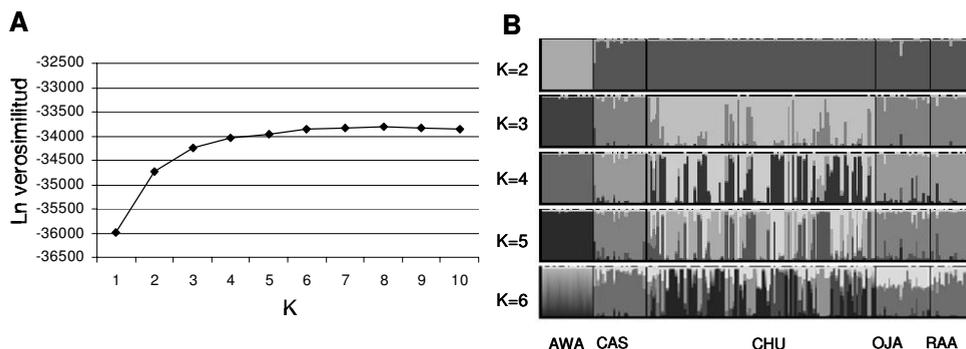


Figura 1. A) Representación gráfica de los valores del Ln de la verosimilitud, estimados con el software STRUCTURE v.2.3 para valores de K entre 1 y 10. B) Representación de la estructura poblacional obtenida para las cinco razas ovinas analizadas.

GENETIC STRUCTURE OF FOUR SPANISH SHEEP BREEDS USING A SNP-CHIP

ABSTRACT: Modern genomic tools allow genome-wide association studies (GWAS) to find genes underlying economically important traits in animals. In human, these studies have found many associations that cannot be always replicated. One of the reasons for this is the presence of population stratification in groups considered to be homogeneous. The objective of this work was to study the genetic structure of four Spanish sheep breeds (Castellana, Churra, Ojalada and Rasa Aragonesa) included in the *Sheep HapMap*. We selected 154 SNPs from the *Ovine SNP50BeadChip*, using PLINK options. Five independent analyses were made with STRUCTURE software to calculate the most likely number of clusters. The closest breeds were Ojalada and Rasa Aragonesa; Awassi was the most different when compared with any of the Spanish breeds. Population structure analysis revealed K=6 to be the most likely number of clusters. According to these results, Awassi was the most different breed (K=2), whereas among the Spanish populations, Churra was the only one clustering separately from the rest (K=3). Moreover, there were more differences within this breed than among the others (K=6). This study represents a first approach to assess/study population structure in Spanish sheep breeds and understand how these populations will behave in future GWAS.

Keywords: population structure, Spanish sheep, SNP-chip