

UNA PERSPECTIVA GENÓMICA SOBRE EL ORIGEN DE LA VARIACIÓN GENÉTICA DE LAS CASEÍNAS CAPRINAS

Dailu Guan¹, Emilio Mármol-Sánchez¹, Xavier Such², Vincenzo Landi³, Marcel Amills^{1,2}

¹Centre de Recerca Agrigenòmica (CRAG), Consorci CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra 08193; ²Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra 08193; Departamento de Genética, Universidad de Córdoba, Córdoba. marcel.amills@uab.cat

INTRODUCCIÓN

En el ganado caprino, el polimorfismo de los genes de las caseínas (*CSN1S1*, *CSN1S2*, *CSN2* y *CSN3*) tiene un efecto muy significativo sobre la producción y la composición de la leche, así como sobre el rendimiento quesero de la misma (Martin *et al.*, 1999, Moioli *et al.*, 2007, Amills *et al.*, 2012, Amills, 2014). El origen de la cabra doméstica (*Capra hircus*) se remonta a la domesticación del bezoar (*Capra aegagrus*) en el Creciente Fértil, hace 10.000 años (Zeder and Hesse, 2000). A pesar de que la variación de los genes de las caseínas caprinas se ha descrito ampliamente (Martin *et al.*, 2002, Moioli *et al.*, 2007, Amills *et al.*, 2012, Amills, 2014), todavía no se ha determinado si las variantes de las caseínas emergieron antes o después de la domesticación. En el presente trabajo, hemos abordado esta cuestión mediante el análisis de la variación de los cuatro genes de las caseínas en el bezoar y en cuatro poblaciones de cabras domésticas.

MATERIALES Y MÉTODOS

Se obtuvieron 110 secuencias genómicas pertenecientes a 22 bezoares y 88 cabras domésticas originarias de Europa, Marruecos, China e Irán (Becker *et al.*, 2014, Benjelloun *et al.*, 2015, Reber *et al.*, 2015, Menzi *et al.*, 2016, Wang *et al.*, 2016, Li *et al.*, 2017, Alberto *et al.*, 2018). Los datos en formato SRA se tradujeron al formato fastq mediante la herramienta fastq-dump 2.8.2 disponible en el paquete SRA-toolkits (<https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/>). Los ficheros fastq se filtraron con el programa Trimomatic (Bolger *et al.*, 2014). Las secuencias se alinearon con el genoma caprino ARS1.0 (Bickhart *et al.*, 2017) mediante el algoritmo BWA MEM (Li, 2013). La función *haplotype Caller* del paquete Genome Analysis Toolkit (GATK, version 3.8) se empleó para generar ficheros vcf (McKenna *et al.*, 2010). Finalmente, los polimorfismos se filtraron de acuerdo a una variedad de criterios que asegurasen la calidad de los mismos.

Parte de los polimorfismos nucleotídicos sencillos (SNPs) identificados con GATK se usaron para investigar la diversidad de las distintas poblaciones analizadas. Más concretamente, los SNPs empleados con dicha finalidad se filtraron mediante los comandos “--hwe 0.001 --maf 0.05 --geno 0.3 --indep-pairwise 50 5 0.2” del programa PLINK (Purcell *et al.*, 2007). Se construyó un árbol neighbor-joining con el programa MEGA7 (Kumar *et al.*, 2016) mediante la utilización de una matriz IBS (Purcell *et al.*, 2007). El análisis de componentes principales (PCA) se llevó a cabo con el comando “-pca” del programa PLINK v1.9 (Purcell *et al.*, 2007).

Se tuvo en cuenta las coordenadas genómicas (genoma de referencia ARS1.0) de los loci de las caseínas (*CSN1S1*, *CSN2*, *CSN1S2* and *CSN3*) para extraer los SNPs localizados en dichos genes mediante la herramienta VCFtools (<https://vcftools.github.io/index.html>). Los efectos funcionales de las variantes de las caseínas se predijeron con el programa SnpEff 4.3 (Cingolani *et al.*, 2012).

RESULTADOS Y DISCUSIÓN

Las secuencias procedentes de 110 genomas fueron mapeadas en el genoma caprino ARS1.0 (Bickhart *et al.*, 2017). La media de cobertura de los genomas fue de

9.91× y más del 99% de las secuencias fueron mapeadas con éxito. El análisis de estos datos con el programa GATK permitió identificar 54 millones de SNPs y 9 millones de indels. Se emplearon 11 millones SNPs para analizar las relaciones genéticas entre las cinco poblaciones analizadas mediante un análisis de componentes principales (PCA) y un árbol neighbor-joining (Figura 1a y 1b).

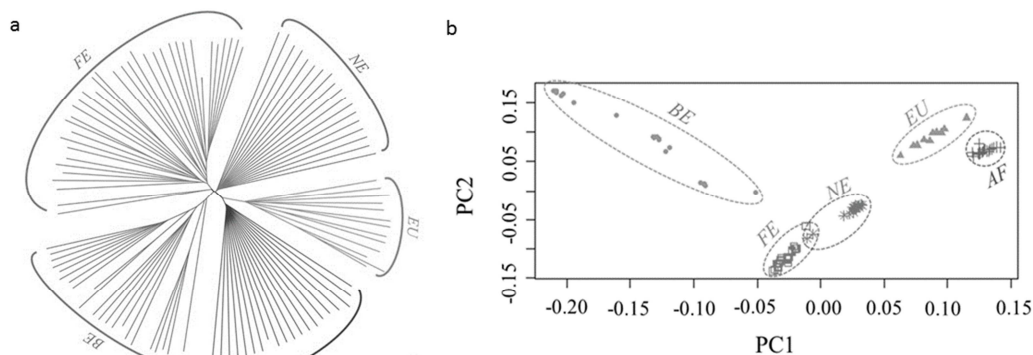


Figura 1 Relaciones genéticas entre bezoares (BE) y cabras domésticas procedentes de Europa (EU), Marruecos (AF), China (FE) e Irán (NE).

El análisis de la **Figura 1** evidencia que las cabras agrupan de acuerdo a su origen geográfico. Las poblaciones de Irán y China presentan un cierto grado de afinidad genética, mientras que las cabras europeas y marroquíes agrupan de forma claramente independiente. Las cabras fueron domesticada en un área geográfica que va de los montes Zagros a Anatolia (Zeder and Hesse, 2000, Naderi *et al.*, 2008) y posteriormente se dispersaron en Europa, África y Asia (Pereira and Amorim, 2010). La diferenciación genética de las cinco poblaciones bajo estudio concuerda con la dispersión de distintos núcleos poblacionales desde el centro de domesticación hacia los tres territorios mencionados anteriormente, esto es, Europa (corredor mediterráneo y danubiano), Asia (estepa rusa y valle del Indo) y África (Pereira y Amorim, 2010). La deriva génica combinada con la existencia de diferencias de manejo productivo y de objetivos de selección probablemente también haya contribuido a establecer las diferencias genéticas que se observan en las cinco poblaciones caprinas analizadas.

Se han identificado centenares de SNPs en los genes de las caseínas (**Figura 2**). La mayor parte de los SNPs identificados son intrónicos, mientras que la segunda categoría más abundante está representada por SNPs situados en las regiones 5' y 3' de los genes. La anotación de los SNPs con el programa SnpEff (Cingolani *et al.*, 2012) reveló que la gran mayoría no tienen un elevado impacto funcional. La comparación de la variación de los genes de las caseínas en bezoares vs cabra doméstica (**Figura 2**), puso de manifiesto que comparten una elevada proporción de los SNPs (38-56%). Por otra parte, la comparación de las cinco poblaciones también evidenció que más de un 50% de las variantes son compartidas entre dos o más poblaciones (**Figura 2**), y un 18-46% de los SNPs son compartidos por todas las poblaciones. Estos resultados indicarían que una proporción importante de la variación genética de las caseínas caprinas ya existía antes de que se produjera la domesticación de la cabra. Esta interpretación concuerda con otros estudios en los que se ha demostrado que diversas variantes genéticas de interés agrícola, como por ejemplo las que determinan el tamaño del tomate y el crecimiento de las ramas axilares del maíz, ya segregaban en los ancestros de dichas plantas (Larson *et al.*, 2014). Otra explicación consistiría en la existencia de flujo génico entre las formas doméstica y salvaje o entre las cinco poblaciones analizadas, sin embargo la inspección de la **Figura 1** no indica la existencia de individuos híbridos. Por otra parte, la variación población-específica podría corresponder a SNPs que han aparecido con posterioridad a la domesticación y dispersión de la cabra doméstica o bien deberse a factores técnicos relacionados con el tamaño muestral y la existencia de errores de secuenciación

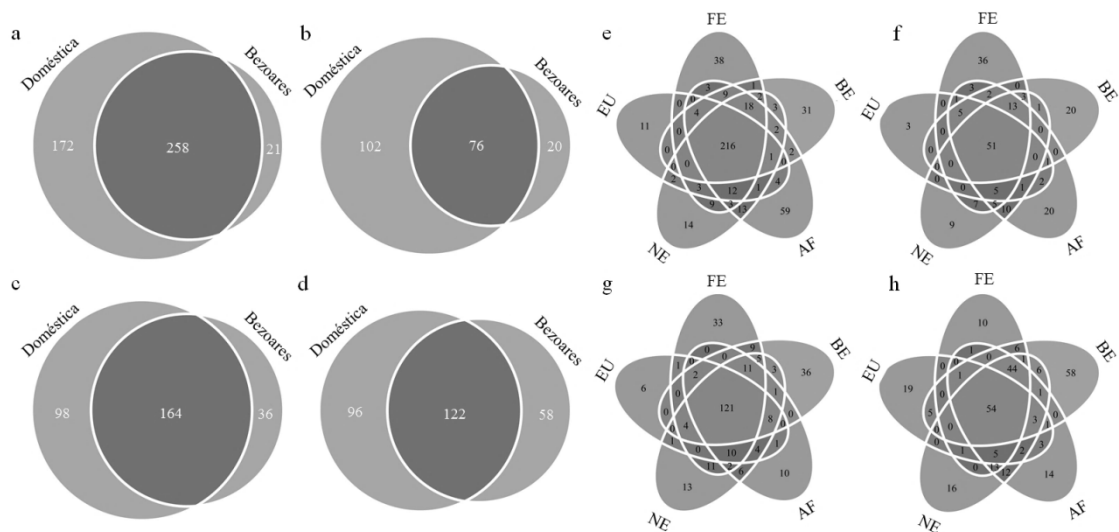


Figura 2. Variación compartida entre bezoar y cabra doméstica (a: *CSN1S1*, b: *CSN2*, c: *CSN1S2*, d: *CSN3*) y entre poblaciones de cabra doméstica de Europa (EU), Irán (NE), China (FE), Marruecos (AF) y bezoar (BE) (e: *CSN1S1*, f: *CSN2*, g: *CSN1S2*, h: *CSN3*)

REFERENCIAS

- Alberto, F. J. *et al.* 2018. *Nat Commun* 9(1):813.
- Amills, M. 2014. *Advances in Biology* 2014:13.
- Amills, M. *et al.* 2012. ed. *InTech*, Rijeka.
- Becker, D. *et al.* 2014. *Animal Genetics* 46(1):50-54.
- Benjelloun, B., *et al.* 2015. *Front Genet* 6:107.
- Bickhart, D. M. *et al.* 2017. *Nat Genet* 49(4):643-650.
- Bolger, A. M., *et al.* 2014. *Bioinformatics* 30(15):2114-2120.
- Cingolani, P. *et al.* 2012. *Fly (Austin)* 6(2):80-92.
- Kumar, S. *et al.* 2016. *Mol Biol Evol* 33(7):1870-1874.
- Larson, G., *et al.* 2014. *P Proc Natl Acad Sci U S A* 111(17):6139-6146.
- Li, H. 2013. *arXiv:1303.3997v2 [q-bio.GN]*.
- Li, X. *et al.* 2017. *Sci Rep* 7(1):15142.
- Martin, P. *et al.* 1999. *Int Dairy J* 9(3-6):163-171.
- McKenna, A. *et al.* 2010. *Genome Res* 20(9):1297-1303.
- Menzi, F. *et al.* 2016. *Sci Rep* 6.
- Moioli, B., *et al.* 2007. *Small Rumin Res* 68(1-2):179-192.
- Naderi, S., *et al.* 2008. *Proc Natl Acad Sci U S A* 105(46):17659-17664.
- Pereira, F. *et al.* 2010. In: *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons, Ltd: Chichester.
- Purcell, S. *et al.* 2007. *Am J Hum Genet* 81(3):559-575.
- Reber, I. *et al.* 2015. *Anim Genet* 46(3):316-320.
- Wang, X. *et al.* 2016. *Sci Rep* 6:38932.
- Zeder, M. A. *et al.* 2000. *Science* 287(5461):2254-2257.

Agradecimientos: Este trabajo ha sido financiado con el proyecto AGL2016-76108-R concedido por el Ministerio de Economía y Competitividad (MINECO). Dailu Guan recibió una beca del *China Scholarship Council*. Emilio Mármol-Sánchez ha sido financiado con un contrato FPU (FPU15/01733) por el Ministerio de Educación Cultura y Deporte (MECD).

A GENOMIC VIEW ABOUT THE ORIGINS OF CASEIN POLYMORPHISM IN GOATS

In the current study, we aimed to investigate if the variation of the caprine casein genes emerged before or after goat domestication. To this end, we have used 110 whole genome sequences from bezoars (the ancestors of domestic goats) and goats from Morocco, China, Iran and Europe to retrieve the SNP variation mapping to the casein genes. We found that bezoars and domestic goats share 38-56% of casein SNPs, and the comparison of the five populations also evidenced a high level of casein SNP sharing. Taken together, these results suggest that a substantial amount of casein variants was already present in the ancestor of domestic goats.

Keywords: domestic goat, casein genes, single nucleotide polymorphism.