

EVALUACIÓN DE LA CORRESPONDENCIA ENTRE LOS MAPAS GENÉTICO (MACHO) Y FÍSICO EN EL GANADO OVINO DE LA RAZA CHURRA

E. García-Gómez¹, B. Gutiérrez-Gil¹, J.P. Sánchez², Y. Bayón¹, L.F de la Fuente¹, F. San Primitivo¹, J.J. Arranz¹

¹Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. ²IRTA Lleida. Av. Alcalde Rovira i Roure, 191 25198 Lleida. E-mail: egarg@unileon.es

INTRODUCCIÓN

El desarrollo de herramientas genómicas en el ganado ovino en los últimos años ha hecho posible la identificación de regiones del genoma asociadas a caracteres de interés en producción animal (Becker et al., 2010; García-Gómez et al., 2011; Zhao et al., 2011; Johnston et al., 2011). La eficacia de estas herramientas está directamente relacionada con la calidad del alineamiento de la secuencia genómica de la que provienen. Debido a la naturaleza, fundamentalmente repetitiva, de los genomas animales y a las cortas longitudes de los fragmentos secuenciados mediante las tecnologías de secuenciación de segunda generación, no siempre la calidad de los mapas físicos derivados de los primeros alineamientos de los genomas presentan una elevada fiabilidad. Además el conocimiento de cómo esta arquitectura física determina los fenotipos de interés hace necesaria la elaboración de un mapa genético necesario para análisis de asociación y de ligamiento. En este último caso, el gran número de meiosis informativas que deben ser analizadas para obtener un mapa de ligamiento de alta densidad, hace que se deban analizar pedigrees de un gran número de animales difícilmente disponibles en un único experimento. En la mayoría de los casos lo que se hace es utilizar los datos medios obtenidos en los análisis del genoma humano (Yu et al., 2001) para convertir las distancias físicas en genéticas considerado que como media a lo largo del genoma 1 cM es igual a 1 Mb. Esta medida es aproximada ya que se ha demostrado que existen regiones cromosómicas más susceptibles a sufrir recombinación (junglas de recombinación) y otras con muy poca tendencia a recombinar (desiertos de recombinación). El objetivo de este trabajo es evaluar la versión 2 del alineamiento del genoma de la oveja mediante la estimación de un mapa de ligamiento basado en las meiosis del sexo masculino en una población de ganado ovino de raza Churra.

MATERIAL Y MÉTODOS

Un población comercial de raza ovina Churra se ha genotipado para el chip de SNPs *Illumina Ovine SNP50BeadChip*. Esta población está formada por 1.696 animales, distribuidos en 16 familias de medio-hermanas, con una media de 105 animales por familia. El mapa físico, con las posiciones de los marcadores en pares de bases (bp), utilizado en este análisis se corresponde con la versión 2 del genoma de la oveja, disponible online (<http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv2.0/>). En dicha versión del genoma existen marcadores que pueden localizarse en diferentes regiones genómicas con una probabilidad muy similar. El control de calidad de los genotipos se realizó, utilizando el software PLINK v1.06 (Purcell et al., 2007), a dos niveles: primero, un control por animal y, posteriormente, un control de calidad por marcador. En la primera fase, se eliminaron todos los animales con un porcentaje de genotipos inferior al 90 %. En la segunda, todos los SNPs con una tasa de genotipado inferior o igual a 95 %, frecuencia del alelo menos frecuente (MAF) inferior o igual a 0,01 y probabilidad de equilibrio Hardy-Weinberg menor de 0,00001, se eliminaron del análisis. Finalmente, todos los SNPs con localización en el genoma desconocida o que se localizan en cromosomas sexuales se excluyeron del análisis aunque hubiesen pasado del control de calidad.

Para la construcción del mapa de ligamiento, se utilizó la opción *fixed* del programa CRI-MAP (Green et al., 1990), v2.503 (proporcionada por JF Maddox). En este análisis, se asume como verdadero el orden de los marcadores, obtenido de la versión 2 del genoma ovino, y se calcula la distancia genética entre ellos en función de la frecuencia de

recombinación en las familias estudiadas. Estos cálculos se llevaron a cabo tantas veces como fue necesario para minimizar los problemas de localización de algunos marcadores. En el caso de que un marcador mostrase un nivel de recombinación con los adyacentes mucho más elevado del esperado por la distancia física entre ellos existen cuatro opciones: (1) si dicho marcador no tiene otra posible localización en el genoma, el mapa se construye sin ese marcador, en caso de que la frecuencia de recombinación entre los marcadores restantes sea adecuada a su distancia, se elimina el marcador problemático y se asume como verdadero el nuevo orden (sin ese marcador); (2) si no tiene otra posible localización, y el mapa genético no mejora al eliminarlo, probamos eliminando los marcadores colindantes, si aun así no mejora, asumimos que puede haber un error, pero los siguientes cálculos se hacen asumiendo como correcto el mapa inicial (con el marcador problemático); (3) si el SNP tiene otra posible localización en el genoma, reconstruimos el mapa con la nueva ubicación y, si mejora, asumimos esta última como verdadera; (4) en caso de que el mapa no mejore, procedemos a las opciones 1 y 2, descritas previamente, eliminando el marcador del análisis. Las posiciones “definitivas” de los mapas físicos (en Megabases, Mb) y de ligamiento (en centiMorgan, cM) se compararon y se ha estimado la correspondencia entre ambas variables tanto por cromosoma como a nivel genómico.

RESULTADOS Y DISCUSIÓN

En total, 15 animales se han eliminado del análisis por no cumplir el criterio de calidad (más del 90% de genotipos válidos), quedando un total de 1.681 animales cuyos genotipos se han utilizado en la construcción del mapa de ligamiento en oveja Churra. De los 54.241 SNPs de chip de Illumina, se eliminaron 6.431 SNPs en el control de calidad: 4.316 SNPs no pasaron el filtro por exhibir más de un 5% de genotipos no válidos; 1.540 SNPs mostraban un MAF por debajo del umbral (0,01); y 575 marcadores no estaban en equilibrio Hardy-Weinberg ($p < 0,00001$) en la población analizada. De los 47.810 SNPs restantes, se han utilizado para la construcción de la primera versión del mapa de ligamiento, los SNPs autosómicos con localización conocida en el genoma, es decir, 46.365.

En total se realizaron 6 iteraciones del mapa de ligamiento hasta obtener la versión “de trabajo”. En el proceso de construcción del mapa 25 SNPs han sido eliminados del mapa de ligamiento y 114 marcadores han sido integrados en el mapa genético ocupado la segunda o tercera posición más verosímil en el mapa físico.

Al comprobar la relación entre los mapas físico y genético en la población de raza Churra estudiada, podemos observar que, como media, 1 Mb equivale a 1,85 cM. Este ratio (cM/Mb) varía a lo largo del genoma, desde 1,50 en OAR2 hasta 2,37 en el OAR20. En la Figura 1 se presentan de manera gráfica estos resultados. Al representar el genoma completo (Figura 1A), se observa claramente que el ratio observado es superior a 1 cM ~ 1 Mb. Las Figuras 1B y 1C muestran un zoom sobre los cromosomas con los ratios mayor y menor, respectivamente a lo largo del genoma. Así como en el cromosoma OAR2 (cM/Mb = 1,50), se puede observar una correspondencia entre el mapa físico (eje Y) y el mapa de ligamiento (eje X) constante sin regiones donde la recombinación sea muy elevada, en el cromosoma OAR20 (cM/Mb = 2,37) se observan diferencias importantes entre las distintas regiones del mismo. Hay que tener en cuenta que las técnicas de secuenciación de segunda generación producen millones de lecturas de tamaño corto (35-500 bp) que se alinean a lo largo del genoma. Estas secuencias forman *contigs*, *scaffolds* y *super-scaffolds*, entre los que puede haber huecos y las distancias físicas estar sub- o sobrestimadas. En el caso del genoma ovino, aún en estado de borrador inicial, este fenómeno puede claramente afectar a la calidad del mapa físico. Además hay que tener en cuenta que la calidad del mapa de ligamiento obtenido está limitada por el reducido tamaño del pedigrí analizado. Las diferencias en la correspondencia cM/Mb entre cromosomas ya se ha puesto de manifiesto en otras especies de mamíferos) existiendo regiones donde la recombinación es mucho mayor que la media del genoma (selvas de recombinación) y otras donde se casi no se observan sucesos recombinatorios (desiertos de recombinación) (Yu et al., 2001; Liu et al 2009).

Los resultados obtenidos en nuestro análisis muestran que la relación cM/Mb es de 1,85. El refinamiento del mapa físico en futuras actualizaciones del genoma, puede ayudar a que este ratio se acerque más al ratio encontrado en las especies con un genoma mucho más elaborado 1 cM ~ 1 Mb. Por el momento, es necesario tener en cuenta, a la hora de interpretar resultados basados en esta versión del genoma, que se encuentra en un estado inicial de desarrollo y esto puede llevar a una incorrecta asociación de regiones genómicas con caracteres productivos. Así, la evaluación de la tasa de recombinación en dichas regiones se convierte en un método adecuado para limitar este tipo problemas

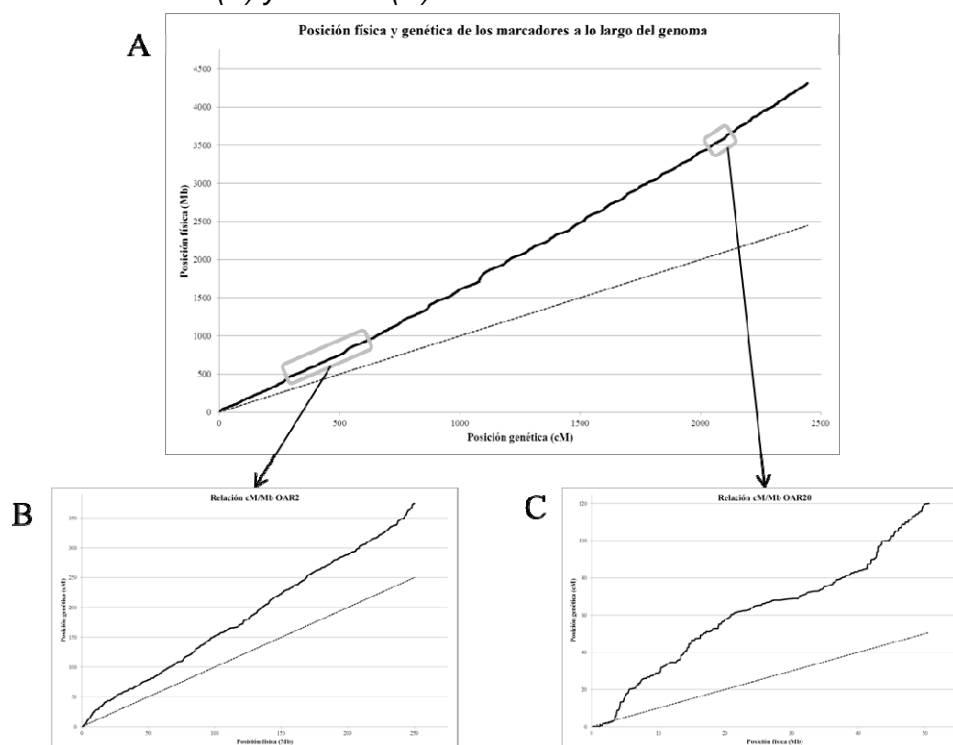
REFERENCIAS BIBLIOGRÁFICAS

▪ Becker D, Tetens J, Brunner A, *et al.* 2010. PLoS ONE. 5: e8689. ▪ García-Gómez E, Reverter A, Whan V, *et al.* 2011. PLoS ONE. 6: e21158. ▪ Green P, Falls K & Crooks S. 1990. Documentation for CRI-MAP, v2.4. ▪ Liu Y, Qin X, Song XZ, *et al.* 2009. BMC Genomics. 10:180 ▪ Purcell S, Neale B, Todd-Brown K, *et al.* 2007. Am. J. Hum. Genet. 81. ▪ Yu A, Zhao C, Fan Y, *et al.* 2001. Nature. 409:951-3. ▪ Zhao X, Dittmer KE, Blair HT, *et al.* 2011. PLoS ONE, 6: e21739.

Agradecimientos:

Este trabajo se ha realizado con financiación del proyecto AGL2009-07000 (Ministerio de Ciencia e Innovación) y “3SR” *Sustainable Solutions for Small Ruminants* (Comisión Europea). EGG es becaria FPU (Ministerio de Educación).

Figura 1. Representación gráfica de la relación entre las posiciones físicas (en Megabases, Mb) y genéticas (en centiMorgan, cM) de los marcadores obtenidas en este estudio y la relación 1 cM ~ 1Mb (línea discontinua). (A) Relación entre posiciones de los mapas físicos y genéticos a lo largo de todo el genoma, se representan las posiciones en cM en el eje X y en Mb en el eje Y. (B) y (C) Posiciones físicas y genéticas para los marcadores situados en los cromosomas OAR2 (B) y OAR20 (C).



ASSESSMENT OF THE CORRELATION BETWEEN GENETIC (MALE) AND PHYSICAL MAPS IN SPANISH CHURRA SHEEP

ABSTRACT: To assess the correspondence between the physical map, derived from Ovine Assembly v2.0, and the genetic map in Spanish Churra sheep, we built a linkage map using the genotypes from the *Illumina Ovine SNP50BeadChip*. After a quality control per animal and per marker, 46,341 autosomal SNPs and 1,681 animals belonging to 16 half-sib families were used to estimate the equivalence between marker positions in cM and Mb. The results reported here show that the average recombination frequency in the 16 sires is 1.85 cM per Mb, higher than the average value in most mammals (1 cM ~ 1 Mb). Moreover, we have detected differences between chromosomes, with a minimum of 1.50 cM/Mb in OAR2 y a maximum of 2.37 cM/Mb in OAR20. These values should be used as a reference in studies where recombination information between markers is important, for example, in QTL detection or effective population size estimation based on molecular information. However, we have to take into account that in this work we have only analysed male meiosis and the number of animals is limited. It is also important that the Ovine Genome Assembly is still in a draft stage and future refinements will improve the sequence.