

PROCEDIMIENTOS PARA LA DETECCIÓN DE SEÑALES DE SELECCIÓN.

González-Rodríguez, A.¹, Munilla, S.^{1,2}, Mouresan, E. F.¹, Moreno, C.^{1,3}, Altarriba, J.^{1,3}, Varona, L.^{1,3}

¹Unidad de Genética Cuantitativa y Mejora Animal. Facultad de Veterinaria. Universidad de Zaragoza. 50013. Zaragoza. ²Facultad de Agronomía. Universidad de Buenos Aires (Argentina). ³Instituto Agroalimentario de Aragón (IA2). 50013. Zaragoza

La Teoría de Evolución Neutral

Los procedimientos de detección de señales de adaptación están basados en la teoría de evolución neutral (Kimura, 1983) en equilibrio mutación-deriva. Esta teoría postula que:

1. muchas de las mutaciones que surgen en una población son deletéreas y, en consecuencia, son rápidamente eliminadas;
2. una ínfima proporción de las mutaciones que surgen son ventajosas y rápidamente fijadas en la población;
3. que la amplia mayoría de las mutaciones que se observan son selectivamente neutrales (al menos en términos efectivos) y, en consecuencia, son éstas las que contribuyen al polimorfismo y divergencia (diferencias fijadas).

El estudio de las mutaciones del tercer tipo nos ha permitido realizar los estudios de diversidad bajo la hipótesis de que mayores divergencias entre individuos y/o poblaciones están asociadas a una mayor separación evolutiva o, dicho de otro modo, a la presencia más alejada de un ancestro común. Sin embargo, los procesos de selección y adaptación de las poblaciones ganaderas no se han debido exclusivamente al azar, sino que han involucrado ventajas selectivas de algunos genes o regiones del genoma. Por lo tanto, la teoría de evolución neutral nos permite generar una hipótesis nula, en el sentido estadístico, que puede ser rechazada cuando la información molecular observada no cumpla con los patrones esperados por la hipótesis nula.

Los procedimientos desarrollados para la detección de señales de selección abordan el estudio de la información disponible desde distintos puntos de vista, que conllevan definiciones alternativas de la hipótesis nula de ausencia de selección. Una primera clasificación de estos métodos depende del tiempo evolutivo que se pretende estudiar. Así, pueden ser clasificados en un grupo que analiza la divergencia entre especies, o rangos evolutivos amplios, y otro que estudia la variación genética dentro de la misma especie o procesos recientes de selección. Dentro de los procedimientos que estudian un marco evolutivo más reciente, objeto de este trabajo, pueden ser clasificados según Oleksyk *et al.* (2010) en las siguientes categorías: a) Diferenciación entre poblaciones; b) Análisis de la reducción de la variación local; c) Modificaciones del espectro de frecuencias; y d) Extensión del desequilibrio de ligamiento.

Diferenciación entre poblaciones.

El primer grupo de procedimientos para la detección de señales de selección y adaptación hace uso de las metodologías habituales para los análisis de diversidad genética, pero restringidos a regiones específicas del genoma. El punto de partida de estos métodos implica que las poblaciones difieren como consecuencia de procesos de deriva. Sin embargo, si existe alguna región que, además, se ha podido ver afectada por fuerzas selectivas aparecerá reflejada en una mayor divergencia entre las poblaciones en esa región concreta. El análisis entre pares o conjuntos de poblaciones a lo largo de todo el genoma proporciona una aproximación de la distribución del parámetro estadístico bajo la hipótesis nula, que, además, considera los procesos evolutivos que hayan afectado de modo general a cada una de las poblaciones. Entre los procedimientos que analizan los procesos de diferenciación entre poblaciones se pueden destacar:

- *Estadístico F_{ST}* (Wright, 1943): Este procedimiento asume que una población original se ha dividido en subpoblaciones, y calcula la reducción de la heterocigosidad observada en las subpoblaciones, con respecto a la heterocigosidad esperada bajo la frecuencia alélica global. Dicho de otro modo, calcula el cociente entre la variabilidad genética entre subpoblaciones con respecto a la variabilidad genética total. Esta metodología fue propuesta por Wright (1943) junto con otros parámetros: F_{IS} y F_{IT} , que reflejan, respectivamente, el grado de consanguinidad en cada una de las subpoblaciones

y en su conjunto, respectivamente. Una de las limitaciones de la utilización de F_{ST} es que asume que todas las subpoblaciones tienen el mismo tamaño efectivo y derivan independientemente de una misma población ancestral (Qanbari y Simianer, 2014). Esto implica que el estadístico solo puede detectar procesos de selección que operan en sentidos opuestos en las poblaciones. Este procedimiento ha sido muy utilizado, tanto para un solo *locus*, como para haplotipos y/o regiones del genoma (Qanbari y Simianer, 2014). Sin embargo, dado que el análisis de un *locus* simple está sometido a una gran incertidumbre, se propone habitualmente el análisis conjunto de un grupo de marcadores (Qanbari y Simianer, 2014).

- *Selestim* (Vitalis *et al.*, 2014): Este método analiza la divergencia entre las frecuencias alélicas de un conjunto de poblaciones desde una perspectiva Bayesiana. El procedimiento asume un modelo jerárquico, donde la frecuencia alélica para un *locus* en cada una de las poblaciones está determinada a partir de un modelo probabilístico de difusión, donde se integran la tasa de migración entre poblaciones, la frecuencia alélica ancestral, la intensidad y la dirección de la selección a la que está siendo sometido ese *locus*. En jerarquías sucesivas, las intensidades de selección específicas por *locus* y población se asume que se distribuyen a partir de un conjunto de hiperparámetros (efecto medio de la selección para un *locus* en todas las poblaciones y para todos los *loci* en todas las poblaciones). De este modo, el procedimiento permite distinguir entre *loci* sometidos a selección con respecto a aquellos neutrales o cuasi-neutrales, proporcionando medidas de la intensidad y la dirección de la selección de cada *locus*. Además, permite hacer inferencia acerca de la tasa de migración entre las poblaciones y un teórico *pool* común.

Análisis de la reducción de la variabilidad.

Un segundo grupo de procedimientos está basado en la reducción de la variación genética en zonas específicas del genoma y evita, de este modo, la necesidad de disponer de información de más de una población. La hipótesis de partida de estos procedimientos implica que las regiones del genoma que alberguen genes sujetos a procesos de selección positiva deberían mostrar una reducción de la variabilidad. Es evidente que esta reducción puede ser confundida con procesos demográficos, como cuellos de botella o la presencia de un número reducido de individuos fundadores. Sin embargo, estos procesos darían lugar a un patrón generalizado a lo largo del genoma, mientras que la presencia de reducción de la variabilidad en zonas específicas se identificaría con señales de selección, o bien con una tasa de mutación más reducida. Por el contrario, un incremento de la variabilidad se puede atribuir a selección balanceadora, o a tasas más altas de mutación. Entre los métodos que analizan la variación en la diversidad local podemos destacar los siguientes:

- *Medida de Heterocigosidad* (Oleksyk *et al.*, 2008; Rubin *et al.*, 2010): El procedimiento más sencillo mide el grado de heterocigosidad a lo largo de ventanas del genoma. Aquellas ventanas asociadas con un incremento o disminución de la misma se identifican con procesos de selección balanceadora o positiva, respectivamente.
- *Detección de pérdida de variabilidad* (Ramey *et al.*, 2013): Este método intenta localizar regiones del genoma donde se haya producido un proceso de fijación. Con este objetivo, identifica segmentos del genoma donde un grupo de marcadores contiguos presenten una frecuencia alélica del alelo menos representado (MAF) inferior a 0.01.
- *Segmentos de homocigosidad (ROH)* (McQuillan *et al.*, 2008): Este procedimiento está relacionado con el anterior, ya que identifica regiones del genoma donde los individuos son homocigotos para todos los polimorfismos contenidos en ella. El procedimiento se desarrolló inicialmente para cuantificar el grado de consanguinidad, pero también puede interpretarse como una medida de déficit de heterocigosidad. Así, las regiones genómicas con elevados *ROH* pueden ser señales de un proceso selectivo (Curik *et al.*, 2014). Además, el tamaño de los segmentos tienen relación con la antigüedad del proceso selectivo, y, por lo tanto, el método también puede proporcionar información al respecto (Howrigan *et al.*, 2011).
- *Parentesco Molecular*: Este procedimiento ofrece una medida de la diversidad genética en una población a partir de la similitud genética entre individuos debida al parentesco. Si el cálculo de este parentesco se realiza a través de regiones específicas del genoma, desviaciones sobre el patrón general pueden ser atribuidas, como en los procedimientos anteriores, bien a señales de

selección positiva o balanceadora, o bien a tasas de mutación reducidas o incrementadas. Además, las medidas de parentesco molecular entre y dentro de poblaciones están directamente relacionados con las medidas de divergencia y de distancia entre poblaciones (Caballero y Toro, 2002).

- *EigenValue o Norma Espectral* (Munilla *et al.*, 2015): Este método puede considerarse una generalización del anterior. Su fundamento reside en que un proceso selectivo conlleva el aumento del parentesco molecular de la región genómica afectada, ya que una mutación favorable presente en un haplotipo produce un aumento de la frecuencia en la población; y, por ende, un exceso de parentesco molecular. Para poder comparar diferentes matrices de parentesco molecular, es necesaria la utilización de un estadístico que resuma en un único valor la información contenida. Por este motivo, las normas matriciales constituyen una alternativa ideal para este propósito. Entre ellas, el radio o normal espectral, que equivale al máximo valor propio, tiene como principal propiedad de que su valor es menor o igual que cualquier otra norma de la misma matriz (Munilla *et al.*, 2015).

Modificaciones del espectro de frecuencias.

Estos procedimientos definen la hipótesis nula a partir de la distribución esperada de las frecuencias alélicas de los polimorfismos genéticos en una región del genoma bajo el modelo de evolución neutral, el denominado espectro de frecuencias. La hipótesis de partida es que la aparición de una nueva mutación, o alelo derivado, ocurre con muy baja frecuencia, y normalmente es eliminada por fijación del alelo ancestral con rapidez. Por el contrario, si esta nueva mutación presenta una ventaja selectiva o se ve afectada por el efecto de arrastre con alguna otra mutación ventajosa, su frecuencia se incrementará.

Entre los procedimientos que se basan en cambios del espectro de frecuencia se pueden destacar:

- *Estadístico D de Tajima* (Tajima, 1989): Este procedimiento compara dos medidas de diversidad genética (θ). Por un lado cuantifica el número total de sitios segregantes presentes en una determinada región del genoma, $\theta = K/a$, donde K es el número total de sitios, $a = 1 + 1/2 + \dots + 1/n - 1$, y n es el número de secuencias en la muestra (Kimura 1983). Este parámetro corresponde al espectro de frecuencia esperado en este estadístico. Por otra parte, calcula el número promedio de nucleótidos diferentes por sitio en la comparación de las secuencias tomadas de a pares (Nei y Li, 1979) que se identifica como su esperanza entre dos secuencias tomadas al azar (π). Este segundo parámetro identifica al espectro observado. La segunda medida (π) está afectada por la presencia de alelos en mayor frecuencia, mientras que la primera (θ) está determinada, en mayor medida, por el tamaño de la muestra y por la deriva genética. La relación que existe entre ambos estimadores permite determinar si la secuencia de análisis es plausible que siga el modelo neutral o se desvíe del mismo. Si ambos estimadores ofrecen un resultado similar, no se puede descartar la hipótesis de evolución neutral. Por el contrario, si existe selección positiva, se reflejará en el incremento de π y, por el contrario, si existe selección balanceadora, θ se verá incrementado (Tajima, 1989).
- *Estadístico D de Fu y Li* (Fu y Li, 1993): Este estadístico fue desarrollado con la misma lógica del test anterior, pero parte de la idea que las mutaciones en una genealogía de secuencias de una muestra tomada al azar pueden clasificarse en mutaciones externas, que son aquellas que aparecen en las ramas externas del árbol genealógico (mutaciones recientes), e internas, que son aquellas que se producen en las ramas internas (mutaciones antiguas). El número de mutaciones externas es más sensible a procesos de selección que el número de mutaciones internas. Por lo tanto, si existe selección negativa o purificadora se observará un exceso de las mutaciones en las ramas externas, mientras que si existe selección positiva se observará una disminución de las mismas. Por lo tanto, si se compara el número de mutaciones en ramas internas y externas con los valores esperados de acuerdo al modelo neutral (número de diferencias nucleotídicas menos uno) se obtiene una prueba estadística sobre la presencia de selección.
- *Estadístico H de Fay y Wu* (Fay y Wu, 2000): Este test se puede considerar una extensión al test de Tajima (1989). En concreto, trata de detectar un exceso de alelos derivados en segmentos del genoma como consecuencia del efecto de arrastre asociado a un proceso de selección. Este

procedimiento permite discriminar entre selección y otros procesos como expansión o contracción de la población que también pueden proporcionar valores negativos del Estadístico D de Tajima. El principal inconveniente es que necesita conocer el alelo ancestral para cada uno de los *loci* segregantes. Habitualmente, este alelo ancestral se infiere a partir de especies relativamente alejadas en el proceso evolutivo.

- *Cociente de Verosimilitud Compuesto (CLR)* (Kim y Stephan, 2002): Este procedimiento calcula la verosimilitud de las frecuencias alélicas observadas en cada segmento del genoma, bajo el supuesto de una distribución teórica generada mediante coalescencia. Posteriormente, Nielsen *et al.* (2005) han propuesto una modificación que obtiene la distribución bajo la hipótesis nula a partir de la información disponible en la propia población y a lo largo de todo el genoma. Sin embargo, dado que la asunción de que el espectro de frecuencias sea homogéneo a lo largo de todo el genoma implica asumir tasas constantes de mutación y recombinación, Chen *et al.* (2010) han propuesto nuevo método, denominado XP-CLR, que utiliza las frecuencias alélicas de una población de referencia de la misma especie para calcular la verosimilitud de la muestra analizada. Para este cálculo, se asume un modelo de deriva genética basado en movimiento browniano bidireccional. De este modo, se postula como hipótesis nula, que las frecuencias alélicas siguen un modelo Gaussiano, centrado en las frecuencias alélicas de la población de referencia y cuya varianza se genera a partir de los procesos de deriva. Las desviaciones con respecto a este modelo en regiones del genoma son indicadoras de señales de selección. De manera similar, Stella *et al.* (2010) construye la distribución de las frecuencias alélicas a partir de un conjunto de poblaciones y detecta las desviaciones particulares de cada una o de un subconjunto de ellas a lo largo del genoma.

Extensión del desequilibrio de ligamiento.

El último grupo de procedimientos está basado en el estudio del efecto de arrastre generado en los *loci* en desequilibrio de ligamiento con un *locus* sometido a un proceso de selección. Este fenómeno crea un patrón en el genoma denominado barrido selectivo. El barrido selectivo provoca una reducción de la variabilidad alrededor del locus sometido a la selección, y conlleva al incremento de la frecuencia de un determinado haplotipo asociado al alelo seleccionado favorablemente. La longitud y frecuencia de estos haplotipos dependen de la intensidad del proceso selectivo y de la frecuencia de recombinación, de manera que la presencia de haplotipos largos con alta frecuencia sugieren la presencia de una huella de selección reciente, ya que se interpreta que el alelo seleccionado ha incrementado su frecuencia de manera rápida en las últimas generaciones. Por el contrario, la presencia de haplotipos de menor tamaño sugiere una mayor antigüedad en la población, ya que han existido más oportunidades de romper esa asociación a lo largo de las generaciones mediante la recombinación. A partir de estos conceptos básicos se han desarrollado varios test estadísticos que pretenden identificar esta presencia de haplotipos de gran tamaño en alta frecuencia a lo largo del genoma. Entre ellos podemos destacar:

- *Extensión de la homocigosidad haplotípica (EHH)* (Sabeti *et al.*, 2002): Este método evalúa la extensión o longitud de la homocigosidad de los haplotipos (EHH) asociada a un “core” definido por uno o varios *loci*. La EHH se define como la probabilidad de que dos cromosomas elegidos al azar en la población sean idénticos por descendencia dentro de una distancia determinada, y se interpreta como una medida del desequilibrio de ligamiento. La presencia de un alelo en alta frecuencia asociada a una EHH alta se identifica como una señal de selección reciente. Por el contrario, una EHH alta asociada a una frecuencia baja identifica una mutación reciente y una EHH baja está asociada a una mutación antigua. Pese a todo, este procedimiento tiene la limitación de que la EHH es también dependiente de la tasa de recombinación. Por este motivo, Sabeti *et al.* (2002) sugirieron definir la hipótesis nula a partir de los cálculos de EHH para otros “cores” situados en la misma región del genoma. Este procedimiento se denomina *rEHH* (*relative extended haplotype homocigosity*).
- *Integral bajo la curva de homocigosidad (iHS)* (Voight *et al.*, 2006): Este procedimiento se basa en el anterior y propone como medida la integral del área definida por la caída de la EHH desde el “core” o *IHH* (*integrated haplotype homocigosity*). El estadístico de comparación es el logaritmo natural entre la *iHH* definida para el alelo ancestral y para el alelo derivado. Un valor

extremadamente negativo del estadístico refleja un barrido selectivo para el haplotipo del alelo derivado. Por el contrario un valor positivo indica que el alelo seleccionado es el ancestral.

Una de las limitaciones de este procedimiento es que cuando un alelo está fijado o cercano a la fijación en la población analizada, la integral asociada a este alelo no podría ser calculada, o tendrá valores cercanos a cero, proporcionando valores de iHS infinitos. Como consecuencia, este procedimiento es útil para detectar procesos selectivos en desarrollo, pero es incapaz de identificar regiones que han llegado a la fijación. Por otra parte, es también dependiente de la tasa de recombinación de la región del genoma asociada. Para tratar de evitar este problema, Ferrer-Admetlla *et al.* (2014) desarrollaron un procedimiento alternativo (nSL) que evita la utilización del mapa genético, ya que se basa en las recombinaciones observadas en la propia muestra. Al evitar el uso del mapa genético, estos autores demostraron que su método es más robusto ante la variación en las tasas de recombinación.

- *Test de Homocigosidad Haplótipica*: Otros tipos de procedimientos se basan en la medida del grado de homocigosidad entre los haplotipos presentes en un determinado segmento del genoma (Hanchard *et al.*, 2006; Garud *et al.*, 2015). En concreto, estos últimos autores definen los siguientes estadísticos: H1, que corresponde a la suma de los cuadrados de las frecuencias haplotípicas, y H12, que agrupa los dos haplotipos más frecuentes en una única categoría y calcula el sumatorio de las frecuencias al cuadrado. Según estos autores, el estadístico H1 es útil para detectar señales de procesos intensos de selección, mientras que H12 debería ser más potente para detectar procesos de menor intensidad.
- Este tipo de procedimientos se basan en el análisis de una única población. Sin embargo, si se dispone de información de varias poblaciones, es posible utilizar una población de referencia para contrastar los valores obtenidos en los análisis. El objetivo de esta aproximación es aumentar la potencia de los test para identificar procesos selectivos que estén cerca o hayan alcanzado la fijación, ya que se aprovecha la presencia de variabilidad en la población de referencia, que se supone que no ha sufrido ese proceso selectivo y sirve, por lo tanto, como patrón de comparación. Entre ellos se pueden destacar los siguientes: *Extensión de homocigosidad haplotípica entre poblaciones (XP-EHH)* (Sabeti *et al.*, 2007): Este procedimiento calcula el logaritmo natural del cociente entre la iHH (*Integrated Haplotype Homocigosity*) estimada en cada una de las dos poblaciones para cada SNP. Una vez estandarizados, los valores muy positivos sugieren una selección en la población utilizada como numerador, mientras que los valores negativos apuntan a un proceso de selección en la población usada como denominador. Otro procedimiento estrechamente relacionado es denominado Rsb , que fue desarrollado por Tang *et al.* (2007). La principal diferencia entre $XP-EHH$ y Rsb radica en el proceso de estandarización, ya que el primer procedimiento utiliza la media como estadístico central, mientras que el segundo utiliza la mediana, con el objetivo de obtener una mayor robustez a la presencia de valores extremos.
- *Cociente de homocigosidad haplotípica entre poblaciones* (Kimura *et al.*, 2007). Este procedimiento calcula, en cada población, la homocigosidad haplotípica (HH) como la probabilidad que dos haplotipos tomados al azar en una población sean iguales a partir del sumatorio de las frecuencias haplotípicas al cuadrado, y la homocigosidad haplotípica más frecuente (MHH) como el cuadrado de la frecuencia haplotípica máxima. Posteriormente, calcula los cocientes entre estos dos estadísticos entre poblaciones para obtener rHH y $rMHH$, que utiliza como estadísticos de referencia. Según estos autores, estos test permiten identificar procesos de selección en alelos fijados o que estén cerca de ella y son robustos ante procesos demográficos y la densidad de marcadores.
- *Variación del desequilibrio de ligamiento entre poblaciones (VarLD)*: Este método fue propuesto por Teo *et al.* (2009), y evalúa la magnitud de las diferencias en los patrones de desequilibrio de ligamiento entre poblaciones a lo largo de todo el genoma. Dicho de otra manera, identifica regiones genómicas donde el grado de variación del desequilibrio de ligamiento entre dos poblaciones es mayor que el resto del genoma. Para ello, una vez calculado el desequilibrio de ligamiento entre todos los marcadores, obtiene, para cada región y población, matrices de correlaciones entre los marcadores. Posteriormente realiza una descomposición matricial propia para cada una de estas matrices, y utiliza la diferencia entre sus valores propios para cuantificar de manera global la magnitud de las diferencias regionales del desequilibrio de ligamiento.

Como se ha podido apreciar en el apartado anterior existe una gran variedad de procedimientos estadísticos para la detección de señales de selección. Todos ellos tienen objetivos similares, aunque difieren considerablemente entre sí metodológicamente. Además, cada uno de ellos posee sus fortalezas y debilidades con relación a la sensibilidad a procesos demográficos, a la antigüedad del proceso selectivo y a las frecuencias alélicas de los marcadores incluidos en la región del genoma (Simianer *et al.*, 2014). Por este motivo, es razonable pensar que las señales que identifican cada uno de los métodos pueden ser debidas a procesos selectivos de características diferenciadas. En este sentido, Sabeti *et al.* (2007) sugieren que los métodos basados en las modificaciones del espectro de frecuencias y/o medidas de variabilidad pueden detectar procesos selectivos muy antiguos, mientras que los métodos basados en la extensión del desequilibrio de ligamiento identifican procesos mucho más recientes. Los métodos basados en la diferenciación entre poblaciones ocupan una posición intermedia.

Por estos motivos, los resultados de una única prueba no proporcionan la evidencia suficiente para obtener unos resultados claros y reproducibles (Simianer *et al.*, 2014). Por lo tanto, estos mismos autores recomiendan la detección de señales mediante varias metodologías y, a ser posible, usando una combinación de las señales en un único test estadístico conjunto. En este sentido, se han desarrollado varias estrategias que permiten combinar varios de estos procedimientos y entre ellas destacan:

- *Factores de Bayes* (Grossman *et al.*, 2010): Estos autores proponen el cálculo del factor de Bayes asociado a cada método y marcador como el cociente entre la probabilidad de que un marcador determinado se comporte de manera neutral o bajo selección. Posteriormente, el producto de los factores de Bayes de todos los procedimientos proporciona un cociente de probabilidad conjunta. El principal inconveniente de esta aproximación radica en el cálculo de la probabilidad asociada a la hipótesis nula, que no es inmediata, y en que los métodos no son independientes en sí, motivo por el cual el cociente de probabilidades puede ser sesgado.
- *Meta-análisis de valores p* (Utsunomiya *et al.*, 2013): Este procedimiento asume que todas las metodologías proporcionan un valor p, y utiliza la versión ponderada del método de Stouffer *et al.* (1949) para poder combinar los valores p para cada marcador y test. Las limitaciones de este procedimiento radican en que la determinación de un valor p para todas las metodologías no es siempre posible y que, como en el caso anterior, los procedimientos no son independientes.
- *Análisis de componentes principales* (Simianer *et al.*, 2014): Este método propone la combinación de varios test aplicando la factorización canónica de la matriz de correlaciones de los valores dados por cada test, a diferencia de las otras propuestas, esta utiliza los valores estandarizados en lugar de estimar aproximaciones de la probabilidad que un marcador no se comporte de manera neutral. Sin embargo, cada componente principal estará asociado con varios test de modo heterogéneo y, por lo tanto, la determinación de un único valor que resuma todas las pruebas no es inmediata. Así, los propios autores recomiendan usar más de un eje canónico para la inferencia de las señales de selección.

Referencias

- Caballero, A. y Toro, M.A., 2002. *Conservation Genetics*, 3(3):289–299.
- Chen, H., et al. 2010. *Genome Research*, 20(3):393–402.
- Curik, I et al. 2014. *Livestock Science*, 166:26–34.
- Fay, J.C. y Wu, C.I., 2000. *Genetics*, 155:1405–1413.
- Ferrer-Admetlla, A., et al. 2014. *Molecular Biology and Evolution*, 31(5):1275–1291.
- Fu, Y.X. y Li, W.H., 1993. *Genetics*, 133:693–709.
- Garud, N.R. et al. 2015. *PLOS Genetics*, 11(2):e1005004.
- Grossman, S.R et al., 2010. *Science*, 327(5967):883–886.
- Hanchard, N. et al., 2006. *American Journal of Human Genetics*, 78(1):153–159.
- Howrigan, D.P. et al. 2011. *BMC Genomics*, 12(1):460.
- Kim, Y. y Stephan, W., 2002. *Genetics*, 160:765–777.
- Kimura, M., 1983. *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Reino Unido.
- Kimura, R. et al. 2007. *PLOS ONE*, 2(3):e286.

- McQuillan, R. et al. 2008. *American Journal of Human Genetics*, 83:359-372.
- Munilla, S. et al. 2015. En: *XVI JORNADAS SOBRE PRODUCCIÓN ANIMAL*. Zaragoza - España.
- Nei, M. y Li, W.H., 1979. *Proceedings of the National Academy of Sciences of the United States of America*, 76(10):5269–5273.
- Nielsen, R. et al. 2005. *Genome Research*, 15(11):1566–1575.
- Oleksyk, T.K. et al. 2010. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365:185–205.
- Oleksyk, T.K. et al. 2008. *PLOS ONE*, 3(3):e1712.
- Qanbari, S. y Simianer, H., 2014. *Livestock Science*, 116:133–143.
- Ramey, H.R. et al. 2013. *BMC Genomics*, 14(1):382.
- Rubin, C.J. et al. 2010. *Nature*, 464(7288):587–591
- Sabeti, P.C. et al. 2002. *Nature*, 419(6909):832–837.
- Sabeti, P.C. et al. 2007. 2007. *Nature*, 449(7164):913–918.
- Simianer, H. et al. 2014. In *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*. Vancouver - Canada.
- Stella, A. et al. 2010. *Genetics*, 185(4):1451–61.
- Stouffer, S.A. et al. 1949. *Studies in Social Psychology in World War II: The American Soldier. Vol. 1, Adjustment During Army Life*. Princeton: Princeton University Press. Estados Unidos de America
- Tajima, F., 1989. *Genetics*, 123(3):585–595.
- Tang, K. et al. 2007. *PLOS Biology*, 5(7):1587–1602.
- Teo, Y.Y. et al. 2009. *Genome Research*, 19(10):1849–1860.
- Utsunomiya, Y.T. et al. 2013. *PLOS ONE*, 8(5):e64280.
- Vitalis, R. et al. 2014. *Genetics*, 196(3):799–817.
- Voight, B.F. et al. 2006. *PLOS Biology*, 4(3):0446–0458
- Wright, S., 1943. *Genetics*, 28(2):114–138.