

Estudios de asociación genómica en poblaciones animales independientes

*Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires,
Área Ciencias Agropecuarias*

Yeni Liliana Bernal Rubio

Zootecnista - Universidad Nacional de Colombia - 2006

Magister en Biometría y Mejoramiento - Universidad de Buenos Aires - 2013

Lugar de trabajo: Facultad de Agronomía, Universidad de Buenos Aires.



FAUBA

Escuela para Graduados Ing. Agr. Alberto Soriano
Facultad de Agronomía – Universidad de Buenos Aires



COMITÉ CONSEJERO

Director de tesis

Rodolfo Juan Carlos Cantet

Ing. Agr. (Universidad de Buenos Aires, Argentina)
MSc. (Montana State University, Estados Unidos de América)
MSc. (University of Illinois, Estados Unidos de América)
PhD (University of Illinois, Estados Unidos de América)

Consejero de Estudios

Juan Pedro Steibel

Ing. Agr. (Universidad Nacional de la Pampa, Argentina)
MSc. (Universidad de Buenos Aires, Argentina)
PhD (Michigan State University, Estados Unidos de América)

JURADO DE TESIS

Director de tesis

Rodolfo Juan Carlos Cantet

Ing. Agr. (Universidad de Buenos Aires, Argentina)
MSc. (Montana State University, Estados Unidos de América)
MSc. (University of Illinois, Estados Unidos de América)
PhD (University of Illinois, Estados Unidos de América)

JURADO

Mónica Graciela Balzarini

Ingeniera Agrónoma (Universidad Nacional de Córdoba, Argentina)
MSc. Biometría (Universidad de Buenos Aires, Argentina)
Carrera docente Especialidad en Estadística (Universidad Nacional de Córdoba, Argentina)
PhD (Louisiana State University, Estados Unidos de América)

JURADO

Guillermo Giovambattista

Licenciado en Biología (Universidad Nacional de La Plata, Argentina)
Doctor en Ciencias Naturales (Universidad Nacional de La Plata, Argentina)

JURADO

Daniel Omar Maizón

Médico Veterinario (Universidad de Buenos Aires, Argentina)
MSc Biometría (Universidad de Buenos Aires, Argentina)
PhD (Cornell University, Estados Unidos de América)

Fecha de defensa de la tesis: 22 de Julio de 2015

"Hay una fuerza motriz más poderosa que el vapor, la electricidad y la energía atómica: la voluntad (Albert Einstein)".

A mi familia y a David, por su amor incondicional.

AGRADECIMIENTOS

Considerando el apoyo y colaboración que recibí durante mi formación doctoral, quiero agradecer a todas las personas, grupos de amigos y de investigación que hicieron posible el desarrollo de este trabajo:

En primer lugar, a los miembros del jurado, Dra. Mónica Balzarini, Dr. Guillermo Giovambattista y Dr. Daniel Maizón, tanto por aceptar ser revisores de esta tesis, como por su disposición, tiempo y valiosos aportes a la versión final de este trabajo.

A la Escuela para graduados *Alberto Soriano* (FAUBA) así como al departamento de producción animal y en especial a la cátedra de mejoramiento genético animal, por facilitarme el acceso a los recursos durante mi investigación.

A las instituciones que con su financiación hicieron posible el desarrollo de esta tesis: Consejo Nacional de Investigaciones Científicas (CONICET, Argentina), Departamento Administrativo de Ciencia, Tecnología e Investigación (COLCIENCIAS, Colombia) y al Departamento de Agricultura de Estados Unidos (USDA, USA).

A los docentes de la cátedra de mejoramiento genético animal (FAUBA), por sus enseñanzas, sabiduría y por la compañía recibida en estos años. A Anita Birchmeier, por sus consejos y cuidar de sus hijos FAUBA y darme la oportunidad de conocer una persona tan bondadosa como Noemí.

Al grupo de estadística genética – genómica animal de Michigan State University (USA), especialmente a la Dra Cathy Ernst y al Dr. Ronald Bates, por su ayuda, paciencia y por permitirme ser parte de actividades de investigación y de integración.

A los colaboradores e investigadores del Departamento de Agricultura de Estados Unidos (USDA, USA) y del centro de investigación *Meat Animal Research Center* (MARC, Nebraska, USA). Esta tesis es sin duda, el resultado de su apoyo y colaboración.

A mi consejero de estudios, el Dr. Juan Pedro Steibel, por la oportunidad de trabajar en este proyecto, por la confianza, motivación, por las correcciones y sugerencias y por recordarme siempre la importancia de tener un espíritu crítico y de análisis.

A mis amigos y hermanos de investigación y “esparcimiento” tanto en Argentina (Andrés Rogberg, Carolina García, José Luis Gualdrón, Juan David Corrales, María José Suarez, Natalia Forneris y Sebastián Munilla), como en Michigan State University (Yvonne Badke, Deborah Velez, Kaitlyn Perry, Kaitlin Wurtz, Yongfang Lu, Pablo Reeb, Chunyu Chen, Sebastian Casiro, Scott Funkhouser). Y por supuesto, a todos mis amigos del querido parchecito Latino en Lansing.

A mi director Dr. Rodolfo Cantet, por sus enseñanzas a nivel profesional como personal y por permitirme ser parte de esta familia tan maravillosa como es la cátedra de mejoramiento genético de la FAUBA. Gracias por la confianza, ayuda, buena disposición y por enseñarme el valor de trabajar de manera independiente.

Finalmente a mis padres y hermanos por su comprensión, amor, ayuda y apoyo, los cuales he recibido siempre de modo incondicional. Y desde luego a David, por su amor, compañía, apoyo, y porque al igual que mi familia, ha hecho que la distancia no sea un impedimento para la cercanía espiritual. Este logro es para ellos.

Declaro que el material incluido en esta tesis es, a mi mejor saber y entender, original producto de mi propio trabajo (salvo en la medida en que se identifique explícitamente las contribuciones de otros), y que este material no lo he presentado, en forma parcial o total, como una tesis en ésta u otra institución.

Yeni Liliana Bernal Rubio.

PUBLICACIONES DERIVADAS DE LA TESIS

1. Bernal Rubio, Y.L., Gualdrón Duarte, J.L., Bates, R.O., Ernst, C.W., Nonneman, D., Rohrer, G.A., King, A., Shackelford, S.D., Wheeler, T.L., Cantet, R.J.C. y Steibel, J.P. 2014. Meta-analysis of genome wide association studies for pork quality traits. En Proceedings of the 10th World congress of Genetics Applied to Livestock Production – WCGALP. Vancouver, Canada. August 17-22.

ÍNDICE GENERAL

	Página
DECLARACIÓN.....	VII
PUBLICACIONES DERIVADAS.....	IX
ÍNDICE GENERAL.....	XI
ÍNDICE DE TABLAS.....	XV
ÍNDICE DE FIGURAS.....	XVII
ABREVIATURAS.....	XIX
RESUMEN.....	XXI
ABSTRACT.....	XXIII
CAPÍTULOS	
1. INTRODUCCIÓN GENERAL	1
2. ANÁLISIS CONJUNTO DE ASOCIACIÓN Y ESTRUCTURA GENÉTICA DE LA POBLACIÓN.....	9
2.1. Introducción.....	11
2.2. Materiales.....	13
2.2.1. Población experimental de cerdos Michigan State University (MSUPRP; East Lansing, Michigan, USA).....	13
2.2.2. Población del centro de investigación Meat Animal Research Center (MARC; Clay Center, Nebraska, USA).....	13
2.2.3. Población comercial.....	14
2.3. Métodos.....	15
2.3.1. Imputación de genotipos en la población Comercial.....	15
2.3.2. Desarrollo de estudios de asociación genómica basados en modelos GBLUP.....	15
2.3.3. Estimación de efectos de SNP y su respectiva varianza.....	16
2.3.4. <i>p</i> -valores para significancia de los efectos de SNP.....	16
2.3.5. Análisis de la estructura genética de la población.....	17
2.3.6. Análisis de asociación conjunto.....	17
2.3.6.1. Estimación de componentes de varianza asumiendo una matriz de relaciones genómicas G diagonal en bloques y homogeneidad de varianzas.....	18
2.3.6.2. Estimación de componentes de varianza asumiendo varianza genética heterogénea pero varianza residual homogénea.....	19
2.3.6.3. Estimación de componentes de varianza asumiendo varianzas genéticas y residuales heterogéneas.....	19
2.3.6.4. Estimación de componentes de varianza a partir de la combinación de genotipos y fenotipos provenientes de cada población.....	20
2.4. Resultados.....	20
2.4.1. Exactitud en la imputación de genotipos.....	20
2.4.2. Variación fenotípica entre poblaciones.....	21
2.4.3. Estimación de componentes de varianza y heredabilidad.....	21
2.4.4. Análisis de la estructura genética de la población.....	23

2.4.5. Análisis de asociación poblacional para CIE a*	25
2.4.6. Análisis de asociación conjunto.....	26
2.5. Discusión.....	30
3. META-ANÁLISIS DE ASOCIACIÓN A PARTIR DE MODELOS DE PREDICCIÓN GENÓMICA.....	35
3.1. Introducción.....	37
3.2. Materiales.....	38
3.3. Métodos.....	39
3.3.1. Estimación del desequilibrio de ligamiento y persistencia de la fase entre poblaciones.....	39
3.3.2. Desarrollo de estudios de asociación genómica basados en modelos GBLUP.....	40
3.3.3. Análisis de la estructura genética dentro de población.....	40
3.3.4. Equivalencia entre test estadísticos basados en modelos animales centrados y test basados en modelos asumiendo efectos de SNP fijos.....	41
3.3.5. Meta-análisis de estudios de asociación genómica (MA-GWA)...	45
3.3.5.1. Cálculo de estadísticos z (z-scores).....	46
3.3.5.2. Esquemas de ponderación.....	46
3.3.5.2.1. Ponderación empleando la inversa de la varianza de los efectos estimados de SNP....	46
3.3.5.2.2. Ponderación por tamaño muestral.....	46
3.3.5.3. Combinación de estadísticos z.....	47
3.3.6. Comparación con análisis conjunto de asociación.....	47
3.4. Resultados.....	47
3.4.1. Desequilibrio de ligamiento y persistencia de la fase a través de través de poblaciones.....	47
3.4.2. Variación fenotípica, componentes de varianza y análisis de asociación a nivel poblacional.....	49
3.4.3. Meta-análisis de estudios de asociación genómica (MA-GWA)	50
3.4.4. Comparación meta-análisis de asociación y análisis de asociación conjunto.....	52
3.5. Discusión.....	53
4. ESTUDIOS DE ASOCIACIÓN Y META-ANÁLISIS PARA CARACTERES DE CALIDAD DE CARNE EN CERDOS.....	57
4.1. Introducción.....	59
4.2. Materiales.....	60
4.2.1. Poblaciones animales.....	60
4.2.2. Información genotípica y fenotípica.....	61
4.3. Metodología.....	62
4.3.1. Análisis de asociación genómica para caracteres de calidad en carne de cerdo.....	62
4.3.2. Análisis de la estructura genética de la población.....	63
4.3.3. Test estadísticos por población.....	63
4.3.4. Meta-análisis de estudios de asociación genómica (MA-GWA)..	64

4.3.5. Búsqueda de genes candidatos para caracteres de calidad de carne considerando bloques de desequilibrio de ligamiento.....	64
4.4. Resultados.....	65
4.4.1. Variación fenotípica entre poblaciones.....	65
4.4.2. Variación genética entre poblaciones.....	66
4.4.3. Análisis de asociación poblacional y meta-análisis de GWA.....	68
4.4.4. Búsqueda de genes candidatos para caracteres de calidad de carne.....	76
4.5. Discusión.....	80
5. DISCUSIÓN GENERAL.....	87
CONCLUSIONES.....	95
BIBLIOGRAFÍA.....	99

ÍNDICE DE TABLAS

Tabla	Página
2.2.1. Descripción de información genotípica para las poblaciones Comercial, MARC y MSUPRP.....	14
2.4.2.1. Estadística descriptiva para CIE a* a través de las poblaciones evaluadas.....	21
2.4.3.1. Componentes de varianza y heredabilidades para CIE a* a través de las poblaciones evaluadas y bajo análisis conjunto de asociación...	22
3.4.1.1. Estimaciones promedio de LD en términos de correlación r^2 para distancias de 0.5Mb, 1Mb, 5Mb y 10Mb entre marcadores en las tres poblaciones de cerdos evaluadas.....	48
3.4.1.2. Comparación entre pares de poblaciones de la correlación de la fase para varias distancias entre pares de marcadores.....	49
3.4.3.1. Asociaciones significativas con CIE a* bajo meta-análisis.....	51
4.4.1.1. Número de registros, media y desviación estándar para cada carácter dentro de población.....	66
4.4.2.1. Efectos fijos considerados para cada carácter dentro de población.....	67
4.4.2.2. Estimaciones de componentes de varianza y heredabilidades para cada carácter y conjunto de datos.....	68
4.4.3.1. Asociaciones significativas para caracteres de calidad de carne usando GWA a nivel poblacional y meta-análisis basado en ponderación por inversa de la varianza	69
4.4.4.1. Asociaciones significativas identificadas por meta-análisis de asociación para caracteres de calidad agrupadas por cromosoma.....	76
4.4.4.2. Genes candidatos identificados para caracteres de calidad de carne.....	79

ÍNDICE DE FIGURAS

Figura	Página
2.4.1.1. Gráficos de exactitud de la imputación de genotipos.....	20
2.4.4.1. Proporción de la varianza explicada por los componentes principales	23
2.4.4.2. Comparación de p-valores entre modelos incluyendo y omitiendo componentes principales.....	24
2.4.5.1. Gráficos Manhattan para CIE a* para todas las poblaciones.....	26
2.4.6.1. Comparación de p-valores entre GWA poblacional y JA asumiendo homocedasticidad en componentes de varianza.....	27
2.4.6.2. Comparación de p-valores entre GWA poblacional y JA asumiendo heterocedasticidad en varianza genética.....	28
2.4.6.3. Comparación de p-valores entre GWA poblacional y JA asumiendo heterocedasticidad en componentes de varianza.....	29
2.4.6.4. Gráfico de Manhattan para CIE a* en análisis conjunto (JA).....	30
3.4.1.1. Estimación de LD y persistencia de la fase entre poblaciones.....	48
3.4.3.1. Gráficos Manhattan para CIE a* en MA-GWA.....	50
3.4.3.2. Comparación de p-valores obtenidos bajo MA.....	52
3.4.4.1. Comparación de p-valores obtenidos bajo el meta-análisis (MA) y el análisis conjunto de asociación (JA).....	52
4.4.3.1. Gráfico Manhattan para pH 24 horas post-sacrificio considerando GWA en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWA.....	70
4.4.3.2. Gráficos de Manhattan para resistencia al corte considerando GWA en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWA.....	71
4.4.3.3. Gráficos de Manhattan para porcentaje de pérdida de líquidos a partir de GWA en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWA.....	71
4.4.3.4. Gráficos de Manhattan para porcentaje de pérdidas por cocción a partir de GWA en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWA.....	72

4.4.3.5. Gráficos de Manhattan para porcentaje de grasa intramuscular a partir de GWA en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWA.....	72
4.4.3.6. Gráficos de Manhattan para CIE L* (luminosidad) a partir de GWA en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWA.....	73
4.4.3.7. Gráficos de Manhattan para CIE a* (color rojo) a partir de GWA en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWA..	73
4.4.3.8. Gráficos de Manhattan para CIE b* (color amarillo) a partir de GWA en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWA.....	74
4.4.4.1. Gráficos de LD y p-valores para asociaciones con: a. CIE a* (Cromosoma 1: 307.9Mb-309.9Mb); b. SF (Cromosoma 2: 5.1Mb-6.2Mb).....	77
4.4.4.2. Gráficos de LD y p-valores para asociaciones con: a. SF (Cromosoma 2:105.4Mb-107.4Mb); b. IMF y CKL (Cromosoma 6: 0Mb-1Mb).....	77
4.4.4.3. Gráficos de LD y p-valores para asociaciones con: a. CIE a* (Cromosoma 6: 48.5Mb-63.1Mb); b. pHu, PRL y CKL (Cromosoma 15:133.2Mb-135.7Mb).....	78

ABREVIATURAS

CIE a*	Medida de color relacionada con color rojo de la carne
CIE b*	Medida de color relacionada con color amarillo de la carne
CIE L*	Medida de color relacionada con luminosidad de la carne
CKL	Porcentaje de pérdidas por cocción
EMMAX	Algoritmo de asociación eficiente de modelos mixtos (Efficient Mixed Model Association eXpedited)
ERT	Caracteres de relevancia económica
FDR	Tasa de "descubrimientos falsos"
GBLUP	Predictor genómico lineal insesgado de mínima varianza
GEBV	Valor de cría genómico estimado
GWAS	Estudio de asociación del genoma completo.
h ²	Heredabilidad
IMF	Porcentaje de grasa intramuscular
JA	Análisis conjunto de asociación
K	miles de SNP
Kb	Kilobases
LD	Desequilibrio gamético (desequilibrio de ligamiento)
MA-GWA	Meta-análisis de estudios de asociación
MAF	Frecuencia del alelo menor
MARC	Población de cerdos "Meat Animal Research Center"
Mb	Megabase
MDS	Escalamiento multidimensional
ML	Máxima verosimilitud
mRNA	Ácido Ribonucleico (ARN) mensajero
MSUPRP	Población de cerdos "Michigan State University"
PCA	Análisis por componentes principales
pH	Medida de acidez o alcalinidad
pHu	pH medido 24 horas post-sacrificio
PRL	Porcentaje de pérdida de líquidos
PSE	Pálida, Suave y Exudativa.
QTL	Locus de un carácter cuantitativo
Q-Q	Gráficos cuantil-cuantil
REML	Método de máxima verosimilitud restringida
SAM	Selección asistida por marcadores
SF	Resistencia al corte
SG	Selección genómica
SNP	Polimorfismo de nucleótido simple
SSF	Resistencia al corte en una porción
WBS	Resistencia al corte por el método de la cuchilla Warner-Bratzler
WGR	Modelos de regresión en todo el genoma
Φ	Distribución normal estándar acumulada
σ_a^2	Varianza genética
σ_e^2	Varianza Residual

Título: Estudios de asociación genómica en poblaciones animales independientes.

RESUMEN

En mejoramiento genético animal, uno de los principales enfoques para explicar la arquitectura genética de caracteres de importancia económica son los estudios de asociación del genoma completo (GWAS), que no suelen ser particularmente potentes dado el reducido tamaño muestral. Un enfoque para aumentar la potencia de detección de QTL es combinar datos de poblaciones diferentes en un análisis conjunto de asociación (JA). Sin embargo, este enfoque tiene limitaciones tales como la definición de efectos a ser modelados entre poblaciones y la dificultad en el acceso a los genotipos de poblaciones comerciales. Alternativamente, se pueden combinar resultados obtenidos de GWAS independientes mediante el meta-análisis (MA-GWAS). El objetivo central de esta tesis es describir e implementar métodos para GWAS a nivel poblacional y también, combinando datos de varias poblaciones (JA), así como combinando resultados de GWAS independientes (MA-GWAS). La aplicación en datos reales mostró que MA aumentó la potencia para detectar QTL significativos en contraste con los GWAS poblacionales y el JA, considerando la estructura genética, la heterogeneidad de los componentes de varianza entre poblaciones y evitando problemas del JA tales como el acceso a los datos y definición de los efectos fijos. Los resultados del MA fueron usados en la búsqueda de genes candidatos, identificando nuevas posiciones para algunos de los caracteres de calidad de carne porcina evaluados. En conclusión, el trabajo describe métodos novedosos para integrar resultados de evaluaciones genómicas independientes, a fin de detectar asociaciones significativas entre poblaciones e identificar nuevos genes asociados con caracteres de importancia económica.

Palabras clave: estudios de asociación genoma completo, análisis conjunto de asociación, meta-análisis, GBLUP, calidad de carne en cerdos

Title: Genome wide association studies for independent animal populations.

ABSTRACT

In animal breeding, one of the most used approaches to elucidate genetic architecture of pork quality traits is to perform genome-wide association (GWAS) studies. However, power of GWAS can be low due to the reduced sample sizes in livestock populations. Although an alternative to increase power of detection of significant QTL is to pool data from different databases in a joint analysis (JA), the approach is limited due to the definition of the effects to be modeled between populations and the difficulties for accessing genotypes from commercial populations. An alternative approach is to combine the results from independent GWAS in a meta-analysis (MA-GWAS). The main goal of this dissertation is to describe and to implement methods for GWAS: 1) by population, 2) by pooling data from different populations in a JA, and 3) by combining results from independent GWAS in a MA-GWAS. Application to real datasets showed that MA increases power for detecting significant associations in comparison to population GWA and JA, while accounting for population structure and heterogeneity of variance components across populations, and by avoiding issues present in JA such as data sharing and definition of fixed effects. Results from MA were used in the search of candidate genes, which retrieved genes previously reported and also, novel candidate genes for most of the evaluated traits. To summarize, this dissertation provided original methods to integrate results from independent genomic evaluations, which can be used to summarize significant associations across populations and to identify novel genes associated with economically relevant traits.

Key Words: genome-wide association studies, joint analysis, Meta-analysis, GBLUP, pork quality

Capítulo 1

Introducción general

Introducción general

Los caracteres de calidad de carne de cerdo constituyen un objetivo de selección complejo y desafiante en los programas de mejoramiento genético. Esto debido principalmente a que el término “calidad de carne” incluye combinaciones de caracteres relacionados con un producto de apropiado contenido nutricional, atractivo en apariencia y con óptima palatabilidad (Kauffman *et al.*, 1990). A partir de 1950, los cambios en la demanda de alimentos de origen animal indujeron cambios en los objetivos de selección en porcinos, pasando de evaluarse características relacionadas con el procesamiento y la palatabilidad a considerar caracteres de rendimiento de res y calidad de carne, buscando reducir el porcentaje de grasa total en la carne de cerdo comercializada (Topel, 1986). Este cambio de objetivos trajo aparejado la necesidad de determinar las fuentes de variación genética de caracteres que afectan el rendimiento de la carne durante el procesamiento, su valor nutricional y por lo tanto, la aceptación por parte del consumidor y precio del producto generado (Cameron *et al.*, 1990; Rothschild y Ruvinsky 1998; Lonergan *et al.*, 2001).

En cerdos, los caracteres de calidad de carne particularmente relevantes son el pH medido 24 horas post-sacrificio, las mediciones objetivas de color (CIE L*, a* y b*; CIE International, 1976) y los caracteres de procesamiento tales como el porcentaje de pérdidas por cocción y la resistencia al corte. Cameron *et al.*, (1990), De Vries *et al.*, (1994), Lo *et al.*, (1992) y Hermesch *et al.*, (2000) estimaron los parámetros genéticos de (co)varianza para distintos caracteres de calidad de carne por medio del análisis de varianza y modelos lineales mixtos, mostrando la existencia de variación genética sustancial en dichos caracteres. No obstante, dichas metodologías enmarcadas dentro de la genética cuantitativa clásica, tienen un alcance limitado para determinar con precisión qué genes y regiones genómicas son responsables de la variación genética observada en caracteres de calidad de carne (Zhu y Zhao, 2007). Actualmente la genética cuantitativa se centra en explorar la arquitectura genética (Gianola *et al.*, 2010) respecto del número, la distribución y la interacción entre loci, que afectan la variación de los caracteres cuantitativos de relevancia económica (ERT, Golden *et al.*, 2000) incorporando la información molecular dentro de las metodologías de evaluación genética animal. De este modo, la genética molecular permitió determinar la función, expresión y posición dentro del genoma de genes asociados con los caracteres de interés (Hayes y Goddard, 2010). Inicialmente se la asoció con la selección asistida por marcadores (SAM), donde se estimaban los efectos aditivos (QTL) relacionados con unos pocos marcadores moleculares en la predicción de los valores de cría (Meuwissen y Goddard, 1996; Spelman *et al.*, 1999). Dado el limitado número de marcadores empleado en SAM, los QTL involucrados explicaban solo una fracción pequeña de la varianza genética relacionada con el objetivo de selección, con lo cual la mayor parte de la varianza no explicada por los marcadores debía ser determinada con métodos de evaluación tradicionales (Dekkers, 2004; Meuwissen *et al.*, 2013). En ese contexto de “cauto optimismo”, como lo llamara Dekkers (2004), surge entonces la selección genómica (SG) por sugerencia de Meuwissen *et al.*, (2001). A diferencia de SAM, la SG intenta estimar los efectos de un gran número de marcadores moleculares conocidos como polimorfismos de nucleótido simple (SNP, por su definición en inglés como “Single Nucleotide Polymorphism”) simultáneamente en una población genotipada de tamaño

razonablemente grande, bajo el supuesto que toda la varianza genética puede explicarse por aquellos marcadores ubicados en posiciones genómicas cercanas a los QTL causales (Meuwissen *et al.*, 2013). Una vez obtenidos los efectos de SNP, es posible predecir los valores de cría genómicos o GEBV (Genomic Estimated Breeding Values) de los individuos evaluados bajo el supuesto que estos últimos son iguales a la suma de los efectos de los SNP evaluados en el panel de marcadores. La disponibilidad de GEBV a edades tempranas condujo a una rápida adopción de la metodología en el mejoramiento genético animal (Hayes y Goddard, 2010), particularmente en bovinos de leche. Sin embargo, el éxito de la SG depende, entre otros aspectos, del nivel de asociación entre los SNP y QTL, condición estocástica que es conocida como desequilibrio gamético o de ligamiento (LD). Asimismo, la SG depende de la densidad de los SNP disponibles, afectando la exactitud de las predicciones GEBV de los candidatos a la selección.

La disponibilidad de plataformas de evaluación de SNP en altas densidades condujo a un interés creciente en identificar regiones candidatas que se encuentren asociadas con la variabilidad genética de caracteres complejos o poligénicos. Surgen así los estudios de asociación del genoma completo o GWAS (Genome Wide Association Studies), enfoque que permite determinar variantes asociadas significativamente con caracteres ERT (Bush y Moore, 2012). Al igual que la SG, los resultados obtenidos a partir de GWAS dependerán, entre otros factores, de la correlación o existencia de LD entre los marcadores y el alelo causal, así como del tamaño muestral de la población evaluada (Hirschhorn y Daly, 2005). En producción animal se efectuaron numerosos análisis GWAS, tales como los trabajos de Maxa *et al.*, (2012) en producción de leche en bovinos; García-Gómez *et al.* (2012) en ovinos lecheros provenientes de una población comercial; Nonneman *et al.* (2013) en cerdos para caracteres de calidad de carne; Casas *et al.* (2013) en vacunos de carne para identificar regiones genómicas asociadas con la concentración de vitamina D y Liu *et al.* (2013) en pollos parrilleros para el peso de la res. Ahora bien, la potencia estadística de estos análisis para detectar asociaciones significativas entre SNP y QTL de ERT es afectada por distintos factores, los cuales deben ser considerados al momento de diseñar las poblaciones para el estudio de asociación, metodología que constituye el tema central de esta tesis.

En primer lugar, el grado de LD en poblaciones de cerdos fue evaluado por Badke *et al.* (2012) y Wang *et al.* (2013). Se reportaron valores de LD promedio entre 0.36 y 0.46 y entre 0.50 y 0.55 para pares de SNP adyacentes (distancia entre SNP de 60Kb a 70Kb) de razas americanas y europeas, respectivamente. En cerdos, Du *et al.* (2007) señalaron la necesidad de un LD mayor a 0.2 entre SNP adyacentes para que el GWAS sea efectivo. Debido al costo, es común genotipar cierto número de individuos parentales en alta densidad mientras que los animales de generaciones posteriores son genotipados en baja densidad (Anderson *et al.*, 2008; Habier *et al.*, 2009). Una alternativa para analizar esta información consiste en predecir o imputar los genotipos en baja densidad, condicionando en los genotipos en alta densidad (Marchini y Howie, 2010). Los genotipos imputados permiten aumentar el número de SNP a probar, incrementando hasta un 10% la potencia de los GWAS (Spencer *et al.*, 2009). Consecuentemente, existe una relación entre el nivel de LD, la exactitud de la imputación y los resultados obtenidos a partir de estudios de asociación: si bien la exactitud de imputar genotipos dependerá del nivel de LD entre los marcadores (Hickey

et al., 2012; Gualdrón Duarte *et al.*, 2013), el incremento en el número de SNP disponibles luego de imputar permitirá aumentar la posibilidad de identificar variantes causales en GWAS para poblaciones con bajos niveles de LD.

Además de su importancia en el desarrollo de estudios de asociación y en la imputación de genotipos, el LD también puede ser considerado al momento de identificar genes candidatos para los caracteres de interés. En los inicios del empleo de genética molecular en producción animal, los genes candidatos eran secuenciados en un número diferente de animales, evaluando la asociación entre secuencias del gen y la variación fenotípica del carácter (Meuwissen *et al.*, 2001). Sin embargo, la disponibilidad de marcadores en alta densidad comprendiendo altos niveles de LD puede ser de utilidad en la caracterización de regiones genómicas significativas, que serán utilizadas en la búsqueda de genes candidatos. Do *et al.* (2014) consideraron la existencia de bloques de LD en regiones cromosómicas que incluían varios SNP significativos, con el fin de caracterizar dichos intervalos y su asociación con caracteres de consumo en cerdos. Sin embargo, un problema que pueden sufrir los GWAS es la reducción del LD por la presencia de subgrupos en la población que muestran valores de desequilibrio distintos (Pritchard y Rosenberg, 1999; Pritchard *et al.* 2000). Esta observación lleva al concepto de poblaciones estratificadas, las cuales hacen referencia a una muestra de individuos dividida en varias sub-poblaciones mutuamente independientes, cada una de las cuales está caracterizada por diferentes frecuencias alélicas (Pritchard y Rosenberg, 1999). Dicha estratificación puede afectar la varianza de las estimaciones de las frecuencias alélicas y la distribución de los genotipos en la población (Laird y Lange, 2011) y producir un nivel elevado de asociaciones significativas aún en marcadores que no están en LD a un QTL, dando lugar a falsas asociaciones entre los SNP y los ERT (Pritchard y Donnelly, 2001). Actualmente, se dispone de herramientas estadísticas para detectar la estructura genética de una población y por lo tanto, evaluar su impacto en la variabilidad genética. Entre tales alternativas, la incorporación de componentes principales (PCA – “Principal component analysis”) como covariables en el modelo de análisis constituye una de las herramientas que permite corregir por estratificación en los estudios GWAS (Price *et al.*, 2006). Básicamente, el enfoque de PCA realiza una corrección en aquellos marcadores con grandes diferencias entre las frecuencias alélicas de distintas poblaciones. Tanto el empleo del PCA como de metodologías derivadas (tales como escalamiento multidimensional, MDS – Multidimensional Scaling) fueron empleadas extensamente en el desarrollo de GWAS dentro del mejoramiento genético animal (Maxa *et al.*, 2012; Minozzi *et al.*, 2012; Zhang *et al.*, 2013; Bermingham *et al.*, 2014), a partir de la descomposición de la matriz de relaciones genómicas, para corregir por la presencia de estructura genética en la población.

Con el desarrollo de la SG en el mejoramiento genético animal, se han empleado distintas metodologías estadísticas enfocadas en predecir los GEBV para los individuos candidatos a selección. La principal diferencia estructural entre dichos métodos radica en el supuesto distribucional de los efectos de los marcadores SNP. Existen dos enfoques: 1) aquellos centrados en efectos de marcador independientes (Meuwissen *et al.*, 2001; Gianola *et al.*, 2010; Habier *et al.*, 2011), y 2) los modelos de regresión del genoma completo (WGR – Whole-genome regression; de Los Campos *et al.*, 2010). Los

WGR comprenden la metodología GBLUP (Genomic Best Linear Unbiased Prediction), y asumen que conjuntamente todos los SNP producen una variable aleatoria que sigue una distribución normal con matriz de (co)varianza igual a una matriz de “relaciones genómicas” G , la cual es calculada para todos los individuos genotipados en la población. Si bien ambos métodos han sido ampliamente difundidos en la predicción de valores de cría genómicos para diferentes especies animales (Chen *et al.*, 2011; Saatchi *et al.*, 2011; Wang *et al.*, 2014), los resultados obtenidos a partir de cada enfoque dependerán de la arquitectura genética del carácter. Daetwyler *et al.* (2010) observaron que la metodología GBLUP tuvo una exactitud mayor en la predicción de los GEBV cuando el número de QTL fue elevado, teniendo cada QTL un efecto pequeño sobre el carácter. Implícito en el supuesto de una distribución normal de los valores de cría genómicos es que la variabilidad en el ERT se distribuye uniformemente a lo largo del genoma. En tales condiciones, la metodología GBLUP es más conveniente en términos computacionales en comparación con los métodos Bayesianos de muestreo (Shepherd *et al.*, 2013). Independientemente de su empleo como metodología predictiva, los GEBV fueron utilizados recientemente en el cálculo de efectos de SNP para probar asociaciones en GWAS (McClure *et al.*, 2012; Rolf *et al.*, 2012; Bernal Rubio *et al.*, 2014; Gualdrón Duarte *et al.*, 2014; Wang *et al.*, 2014).

En los GWAS realizados con paneles de alta densidad es deseable contar con un número alto de SNP en relación con el número de individuos genotipados. Esto se debe a la redundancia de los marcadores en LD que restringe el número de grados de libertad para las pruebas de hipótesis e identificación de asociaciones significativas (Gao *et al.*, 2010). Entonces, es necesario aplicar correcciones a nivel estadístico (valor alfa) de las pruebas conjuntas de hipótesis para que el número de falsos positivos no crezca exponencialmente (Bush y Moore, 2012). Un modo sencillo de corregir pruebas de hipótesis múltiples es el procedimiento de Bonferroni, donde se ajusta el valor alfa por el número de pruebas realizadas. Dado el supuesto que cada una de las pruebas de asociación es independiente de las demás, la corrección de Bonferroni es conservadora en los GWAS con muchas hipótesis y controla el número de falsos positivos a expensas de aumentar el número de falsos negativos (Gao *et al.*, 2010). Aun cuando existen métodos alternativos (Benjamini y Hochberg, 1995; Gao *et al.*, 2010; Fu *et al.*, 2014), la corrección de Bonferroni continúa teniendo gran difusión en la mejora genética animal (Sahana *et al.*, 2010; García-Gómez *et al.*, 2012; Minozzi *et al.*, 2012; Jung *et al.*, 2014) fundamentalmente por la simplicidad de implementación (Fu *et al.*, 2014).

Una manera efectiva de mejorar la potencia de los GWAS es aumentar el tamaño muestral. Sin embargo, los costos económicos y el bajo tamaño efectivo de las poblaciones animales limitan esta estrategia. Una alternativa es realizar un análisis conjunto de asociación (JA, del inglés *Joint analysis*), combinando la información genotípica y fenotípica de varias poblaciones, y analizar los datos conjuntamente (Hirschhorn y Daly, 2005). La combinación de observaciones mejora la potencia para detectar asociaciones que no han sido observadas en los estudios poblacionales individuales y permite confirmar la presencia de QTL detectados en una población individual (Lander y Kruglyak, 1995). En el ámbito del mejoramiento genético animal existen antecedentes del JA en cerdos (Walling *et al.*, 2000; Kim *et al.*, 2005; Muñoz *et al.*, 2009), ganado lechero (Minozzi *et al.*, 2012) y ganado de carne (Reverter *et al.*,

2004), investigaciones que evidenciaron las ventajas del JA comparado con el análisis poblacional, de modo de lograr mayor potencia estadística en GWAS. Sin embargo, el JA muestra ciertas dificultades relacionadas con la disponibilidad y acceso a los datos, la complejidad genética del carácter y con el hecho que las poblaciones individuales muestran características particulares, involucrando distintas razas de animales, definiciones del carácter y protocolos experimentales para la recolección de fenotipos (Riggio *et al.*, 2014). Además el JA tiene un elevado requerimiento computacional, especialmente cuando el número de genotipos y poblaciones es grande y cuando la estructura de (co)varianzas de la interacción SNP \times población es compleja, como en las poblaciones animales (Möhring y Piepho, 2009). Todo esto lleva a la necesidad de buscar un enfoque de análisis alternativo que permita ganar potencia estadística al detectar asociaciones significativas y consistentes entre poblaciones, y que no implique la combinación de datos y posterior análisis conjunto, disminuyendo los problemas computacionales así como aquellos relacionados con la definición de los fenotipos y los efectos fijos y/o aleatorios a incluir en el modelo. Estos problemas no se observan con el meta-análisis (MA, Cochran, 1954; Fleiss, 1993) de estudios de asociación (MA-GWAS) que aumenta la potencia al identificar variantes fenotípicas de pequeño efecto (Evangelou y Ioannidis, 2013). Si bien el MA-GWAS ha sido implementado fundamentalmente en humanos, más recientemente ha sido utilizado con poblaciones animales (Khatkar *et al.*, 2004; Akanno *et al.*, 2013; Bernal Rubio *et al.*, 2014; Bolormaa *et al.*, 2014). En general, los MA-GWAS utilizan información de dominio público, combinando los resultados reportados mediante estadísticos que son únicos para cada marcador con respecto a todas las poblaciones analizadas (Riggio *et al.*, 2014). Möhring y Piepho (2009) señalaron que el MA puede visualizarse como un análisis en dos etapas. En la primera, cada población es evaluada individualmente y se obtienen estimaciones de los efectos de SNP particulares a cada una de ellas. Posteriormente, dichas estimaciones son analizadas de manera conjunta, y son ponderadas durante el cálculo de pruebas estadísticas para los efectos de SNP únicos, de modo diferencial para cada población, al mismo tiempo que son sencillas de computar. Existen distintos enfoques de ponderación en el MA (Nordmann *et al.*, 2012), pero no hay información sobre qué esquema pondera más apropiadamente las pruebas estadísticas de GWAS calculados a partir de GBLUP.

Sobre la base de todo lo expuesto, el objetivo general de la tesis es realizar GWAS para datos de calidad de carne en distintas poblaciones de cerdos, determinando la consistencia de las asociaciones entre marcadores moleculares y regiones del genoma asociadas con la variabilidad genética de los ERT analizados. A tal efecto, se desarrollan extensamente los modelos de evaluación genética para calcular los GBLUP y su asociación con la estimación de los efectos de los SNP y las pruebas de hipótesis para buscar asociaciones entre SNP y regiones genómicas responsables de la variación en los caracteres de calidad de carne. Al especificar los modelos, se consignan en detalle las distintas especificaciones de las varianzas genéticas y ambientales (heterogeneidad de los componentes de varianza) entre poblaciones. En relación con las pruebas de hipótesis, se demuestra la equivalencia de la prueba de efectos de SNP considerados variables aleatorias y una prueba en que los efectos de SNP son asumidos como fijos con hipótesis nula de igualdad a cero, siendo ajustados al modelo de uno a la vez. El capítulo siguiente a la presente introducción muestra el análisis conjunto (JA-

GWAS), mientras que en el siguiente se desarrolla el meta-análisis (MA-GWAS) y se lo compara con el JA-GWAS. Ambos métodos buscan identificar regiones genómicas relacionadas con la variación en los ERT de calidad de carne porcina, de modo de poder considerar luego genes candidatos para los caracteres analizados. Todos los análisis en la tesis son realizados mediante programas escritos en lenguaje *R*.

Capítulo 2

Análisis conjunto de asociación y estructura genética de la población

Análisis conjunto de asociación y estructura genética de la población

2.1. Introducción

La disponibilidad actual de genotipos de marcadores SNP en alta densidad permitió el desarrollo de la selección genómica (SG) en el mejoramiento animal (Meuwissen *et al.*, 2001; Goddard y Hayes, 2007). Dicha metodología permite la estimación de los “valores de cría genómicos” (GEBV) para caracteres productivos y reproductivos de relevancia económica (González-Recio *et al.*, 2008; Hayes *et al.*, 2009; VanRaden *et al.*, 2009; de los Campos *et al.*, 2009). Uno de los métodos más utilizados para la SG en el mejoramiento genético animal es el enfoque GBLUP, metodología que pertenece a la categoría de modelos de regresión del genoma completo (del inglés “Whole-genome regression methods”; Meuwissen *et al.*, 2001; Misztal *et al.*, 2009; Aguilar *et al.*, 2010; de Los Campos *et al.*, 2013), y que también ha sido utilizado recientemente en análisis de asociación (Wang *et al.*, 2012, 2014; Gualdrón Duarte *et al.*, 2014). Una vez que se han obtenido los GEBV a partir de las evaluaciones genómicas es posible desarrollar “estudios de asociación del genoma completo” (abreviados como GWAS, Wang *et al.*, 2012; Wang *et al.*, 2014; Gualdrón Duarte *et al.*, 2014). El objetivo de los GWAS es identificar regiones del genoma que explican una porción sustancial de la variación genética de caracteres complejos (Hirschhorn y Daly, 2005; Visscher *et al.*, 2007). Dado que los GWAS no asumen *a priori* una ubicación en el genoma de los segmentos asociados, constituyen una metodología insesgada en la búsqueda de genes causales (Hirschhorn y Daly, 2005).

Una limitación de los GWAS, dentro de la disciplina del mejoramiento genético animal, es el débil efecto de las pruebas de asociación entre los marcadores y los fenotipos evaluados, situación que tiene como consecuencia el requerimiento de un gran número de individuos para detectar asociaciones significativas (Minozzi *et al.*, 2012). Sin embargo, el aumento del tamaño muestral es costoso y complejo dentro del contexto de la producción animal, dadas las dificultades para obtener poblaciones con fenotipos comparables y la limitada disponibilidad de información proveniente de los sistemas de producción comercial (Houlston *et al.*, 2010). Además, las poblaciones animales suelen ser genéticamente estratificadas, aspecto que al ignorarse conduce a falsas asociaciones entre los marcadores y los caracteres de interés (Rabinowitz, 1997; Hirschhorn y Daly, 2005). Por lo tanto, es posible combinar los datos provenientes de diferentes poblaciones mediante el desarrollo de un *análisis conjunto de asociación* (JA, del inglés “Joint Analysis”; Allison y Heo, 1998; Skol *et al.*, 2006) cuyo objeto es aumentar: 1) el tamaño muestral, 2) la precisión en el mapeo y consecuentemente, 3) el poder en la detección de variantes con efecto reducido. El JA consiste en analizar simultáneamente los datos de diferentes poblaciones bajo un modelo apropiado, y permite aumentar la potencia de detección de QTL que no son identificados por medio del análisis de asociación dentro de las poblaciones individuales, o validar la existencia de QTL detectados en una sola población (Lander y Kruglyak, 1995).

En el contexto del mejoramiento genético animal, existen varios ejemplos de aplicaciones de dicha metodología. Tal es el caso del trabajo realizado por Walling *et al.* (2000) quienes mostraron el potencial del JA en la detección de QTL significativos para caracteres de crecimiento y producción en poblaciones de cerdos, evidenciando la utilidad del JA, tanto para obtener estimaciones más precisas de los efectos y ubicación de QTL comunes entre poblaciones, como para examinar diferencias en los efectos de tales QTL entre los conjuntos de datos analizados. De modo similar, Kim *et al.* (2005) y Pérez-Enciso *et al.* (2005) observaron que la aplicación del JA en modelos multirraciales y multicaracter permitió detectar un número de QTL más alto y de mayor significancia en caracteres de producción en cerdos, comparado con los resultados obtenidos a partir de análisis de una única población. Asimismo, Muñoz *et al.* (2009) compararon análisis de QTL a nivel poblacional versus JA para detectar QTL relacionados con caracteres de crecimiento y grasa corporal en cerdos, observando la necesidad de contar con un elevado y suficiente número de individuos para detectar QTL significativos y para la resolución del mapeo. En bovinos, Minozzi *et al.* (2012) emplearon JA para identificar genes y marcadores asociados con susceptibilidad a paratuberculosis en ganado lechero. Estos autores confirmaron que el JA permitió identificar nuevos loci relacionados con el carácter de interés, los que no habían mostrado un efecto significativo en los GWAS individuales. Recientemente, Riggio *et al.* (2014) implementaron JA para identificar regiones genómicas relacionadas con la variación en la resistencia a nemátodos en tres poblaciones de ovejas no relacionadas, mostrando una vez más el poder del JA sobre los análisis individuales para detectar regiones relacionadas con el carácter de interés. Si bien los trabajos mencionados sientan precedente en el uso de JA en mejoramiento genético animal, existen ciertos aspectos a resolver antes de implementar un JA: 1) el uso de diferentes marcadores en las distintas poblaciones, 2) la incorporación de diferentes fenotipos y definiciones de modelos estadísticos y, 3) las diferencias cualitativas y cuantitativas en los componentes de varianza de las diferentes poblaciones (Walling *et al.*, 2000; Minozzi *et al.*, 2012). Sin embargo, los trabajos mencionados son antecedentes de la aplicación del JA en producción animal, mostrando su utilidad al momento de cuantificar la variación genética que ocurre en caracteres complejos.

Un aspecto de importancia capital en los estudios de asociación es la presencia de estructura genética de la población, la cual tiende a generar falsas asociaciones entre el carácter y los marcadores o, en otros casos, ocultar asociaciones verdaderas. Entre otras razones, esto puede deberse a un tamaño efectivo reducido en la población, así como a la presencia de animales estrechamente relacionados, hecho que se observa con frecuencia en poblaciones animales (Minozzi *et al.*, 2012). Consecuentemente, dentro de los análisis de asociación existen diferentes enfoques para corregir por la estructura genética subyacente dentro de la población. Uno de ellos es la utilización de análisis por componentes principales (PCA, del inglés “principal component analysis”; Price *et al.*, 2006), metodología que realiza una corrección para marcadores con grandes diferencias en las frecuencias alélicas en las distintas poblaciones. Un ejemplo de esta aplicación es el trabajo de Minozzi *et al.* (2012), quienes emplearon escalamiento multidimensional (relacionado con PCA) para evaluar la existencia de estructura genética de la población, mediante la identificación de animales con diferente composición genética.

Por lo tanto, en el presente capítulo se describe la metodología para el desarrollo de GWAS basados en modelos GBLUP, empleando un JA que tiene en cuenta la estructura genética de la población, la amplia difusión de los modelos WGR mediante GBLUP (Lans *et al.*, 2012) y la disponibilidad de los datos originales a partir de tres poblaciones individuales de cerdos para caracteres de calidad de carne. Adicionalmente, se describe el proceso utilizado para la creación de la población conjunta empleando distintos enfoques para considerar 1) la estructura genética de la población y 2) la heterogeneidad en los componentes de varianza entre poblaciones.

2.2. Materiales.

Para la implementación del análisis conjunto, se analizaron datos correspondientes a mediciones objetivas del color rojo de la carne o CIE a* (CIE International, 1976) en la superficie del musculo *longissimus dorsi* después del sacrificio, empleándose registros de tres poblaciones de cerdos, las cuales se describen a continuación.

2.2.1. Población experimental de cerdos Michigan State University (MSUPRP; East Lansing, Michigan, USA).

La población F_0 surgió del apareamiento entre cuatro machos Duroc (no emparentados) con 15 hembras Pietrain, mediante inseminación artificial (Edwards *et al.*, 2008). De los animales resultantes en la F_1 , se mantuvieron 50 hembras y seis machos (hijos de 3 padres F_0) para ser usados como padres de la generación F_2 , compuesta por 1259 cerdos pertenecientes a 142 camadas y 11 grupos de destete. Los apareamientos se organizaron evitando cruzar hermanos enteros o medios hermanos. En la población F_2 , se realizaron mediciones de caracteres de crecimiento, carcasa y composición. Asimismo, los animales fueron genotipados usando dos paneles de marcadores: 1) 411 animales (cuatro machos Duroc F_0 , 15 hembras F_0 Pietrain, seis machos F_1 , 50 hembras F_1 y 336 cerdos F_2) fueron genotipados con el chip de alta densidad PorcineSNP60 BeadChip (Illumina, Inc.) diseñado por Ramos *et al.* (2009); 2) 612 animales F_2 fueron genotipados con el chip de baja densidad GeneSeek Genomic Profiler-Porcine LD (GeneSeek Neogen Company, Lincoln, NE; Badke *et al.* 2013).

2.2.2. Población del centro de investigación Meat Animal Research Center (MARC; Clay Center, Nebraska, USA).

La población experimental MARC fue creada apareando hembras Yorkshire-Landrace ($n = 220$) con padres Duroc o Landrace (12 machos de cada raza), los cuales fueron asignados a las hembras al azar. Para las siguientes generaciones, los apareamientos se diseñaron de la siguiente manera: se cruzaron individuos cuya raza paterna era Duroc, con animales cuyos padres fueron Landrace. Posteriormente, la población se apareó al azar evitando cruzamientos dentro de línea paterna. En total, la población MARC estuvo compuesta por 1237 animales, los cuales fueron fenotipados en las generaciones 4 (531 hembras), 6 (223 machos y hembras) y 7 (483 machos y

hembras). Se recopilaron registros de composición de carcasa y caracteres de calidad de carne, tal como se describe en Nonneman *et al.* (2013). Asimismo, todos los animales fueron genotipados usando el chip de alta densidad Illumina PorcineSNP60 BeadChip (Ramos *et al.*, 2009).

2.2.3. Población comercial.

Siguiendo el proceso descrito en King *et al.* (2011) y Shackelford *et al.* (2012), se obtuvieron lomos de cerdo sin hueso 24 horas post-sacrificio ($n = 1920$), a partir de cuatro instalaciones de procesamiento en gran escala. Los lomos fueron empacados al vacío, almacenados en cajas y transportados a -2.8°C hacia el centro de investigación MARC. A los 14 días post-mortem, se obtuvo una pieza de 2.54 cm de grosor de la región cercana a la Costilla No.14, e inmediatamente se determinó el color de la superficie expuesta del lomo mediante el uso de un colorímetro (Minolta ColorTec PCM; color-tec.com, Clinton, NJ). Posteriormente, se registraron mediciones de composición de res o carcasa y caracteres de calidad de carne. En total se genotiparon 450 lomos usando el chip de alta densidad Illumina PorcineSNP60 BeadChip (Ramos *et al.*, 2009), mientras que 1440 lomos fueron genotipados usando el chip de baja densidad GeneSeek Genomic Profiler-Porcine LD (GeneSeek Neogen Company, Lincoln, NE; Badke *et al.*, 2013).

En todas las bases se descartaron aquellos datos provenientes de individuos con más de 10% de genotipos faltantes así como marcadores con $\text{MAF} < 0.05$. Estos criterios de edición fueron los mismos que los empleados por Badke *et al.* (2012). Así, la Tabla 2.2.1 contiene el número inicial de marcadores y los individuos genotipados en alta y baja densidad, así como el número final de SNP y de individuos luego de aplicar los criterios de edición.

Tabla 2.2.1. Descripción de información genotípica para las poblaciones Comercial, MARC y MSUPRP.

Item	Población		
	Comercial	MARC	MSUPRP
Nº inicial de SNP antes de edición	61565	61565	62163
Nº inicial de animales con genotipo alta densidad	480	1237	398
Nº final de SNP después de edición	45688	44020	40569
Nº final de animales con genotipo alta densidad	474	1234	324
Nº final de animales con genotipo baja densidad	1418	0	604
Nº total de individuos	1892	1234	928

2.3. Métodos

2.3.1. Imputación de genotipos en la población Comercial.

Considerando que la población MARC fue genotipada en alta densidad, y que los animales de la población MSUPRP genotipados en baja densidad fueron imputados en un estudio previo (Gualdrón Duarte *et al.*, 2013), se realizó la imputación de genotipos para animales de la población Comercial que habían sido genotipados en baja densidad GeneSeek Genomic Profiler-Porcine LD (GeneSeek Neogen Company, Lincoln, NE; Badke *et al.*, 2013), siguiendo la metodología empleada por Badke *et al.* (2012). Específicamente, se utilizó el programa impute.R (RDC Team, 2013), disponible en https://www.msu.edu/~steibelj/JP_files/imputation.html. Dicho programa permite, entre otras funciones, construir un panel de referencia a partir de los haplotipos disponibles para los SNP genotipados en alta densidad y posteriormente, imputar las dosis alélicas para los SNP con genotipos faltantes utilizando el software BEAGLE 3.3.1 (Browning y Browning, 2007). Una vez realizada la inferencia de genotipos faltantes, Impute.R permite evaluar la exactitud de la imputación, ya sea en términos de la proporción de alelos correctamente imputados o en función de la correlación entre las dosis alélicas imputadas y observadas (Badke *et al.*, 2013). Sin embargo, el primer enfoque es una estimación sesgada por MAF. Por lo tanto, para obtener una medida de exactitud de la imputación insesgada, en este capítulo se utilizó la correlación entre la dosis alélica observada y la imputada mediante el enfoque de Badke *et al.* (2013).

2.3.2. Desarrollo de estudios de asociación genómica basados en modelos GBLUP

A los efectos de realizar el análisis conjunto de asociación, se estimaron los componentes de varianza y los valores de cría siguiendo un modelo animal centrado para la evaluación genómica cuya ecuación es la siguiente:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{a} + \mathbf{e} \quad [2.1]$$

El vector de registros \mathbf{y} contiene los fenotipos evaluados; \mathbf{X} es la matriz de incidencia que relaciona los registros al vector de efectos fijos $\boldsymbol{\beta}$; \mathbf{a} es el vector aleatorio de valores de cría, cuya distribución es $\mathbf{a} \sim N(\mathbf{0}, \mathbf{G}\sigma_a^2)$ y \mathbf{G} es la matriz de relaciones genómicas ($n \times n$), siendo n el número de animales genotipados. La matriz \mathbf{G} , calculada como $\mathbf{G} = \mathbf{Z}\mathbf{Z}'$, es escalada y por tanto, análoga a la matriz de relaciones de parentesco \mathbf{A} . Por su parte \mathbf{Z} ($n \times m$, con m igual al número de SNP disponibles dentro de cada población después de la edición), contiene las dosis alélicas normalizadas (el conteo del alelo menor, en nuestro caso el alelo “B”) menos su valor esperado dividido por la desviación estándar esperada. Por ejemplo, el elemento de la matriz \mathbf{Z} relacionado con el animal i y el SNP j se calculó del modo siguiente:

$$\mathbf{z}_{ij} = \frac{M_{ij} - 2p_j}{\sqrt{2p_j(1-p_j)}} \quad [2.2]$$

donde M_{ij} es el ij -ésimo elemento de la matriz M con dimensión $n \times m$. Esta matriz que contiene genotipos en el intervalo $[0, 2]$ (conteos del alelo “B”) es normalizada por la frecuencia del SNP j (p_j), a su vez calculada idealmente a partir de la generación F_0 . Asimismo, e es el vector de efectos residuales, con $e \sim N(0, I\sigma_e^2)$. Los escalares σ_a^2 y σ_e^2 representan las varianzas genética aditiva y residual, respectivamente.

2.3.3. Estimación de efectos de SNP y su respectiva varianza.

En trabajos previos (Strandén y Garrick, 2009; Badke *et al.*, 2014; Gualdrón Duarte *et al.*, 2014) se mostró que el modelo [2.1] es equivalente al siguiente donde se centran los marcadores:

$$y = X\beta + Zg + e \quad [2.3]$$

Además de los elementos definidos previamente en [2.1], el vector g corresponde a los efectos aleatorios de SNP. En este contexto, los efectos de SNP pueden predecirse empleando la siguiente transformación lineal de los valores de cría predichos \hat{a} :

$$\hat{g} = Z'G^{-1}\hat{a} \quad [2.4]$$

Las matrices Z y G han sido definidas previamente. Además, Gualdrón Duarte *et al.* (2014) demostraron que la varianza de las estimaciones de los efectos de SNP se puede calcular como sigue:

$$\text{Var}(\hat{g}) = Z'G^{-1}Z\sigma_a^2 - Z'G^{-1}C^{aa}G^{-1}Z \quad [2.5]$$

donde σ_a^2 es la varianza genética y C^{aa} es la parte de la inversa de la matriz de los coeficientes en las ecuaciones de modelo mixto asociada con los valores de cría.

2.3.4. p -valores para significancia de los efectos de SNP.

A fin de identificar asociaciones significativas para los caracteres de interés, se calcularon los p -valores a partir de la siguiente expresión:

$$p\text{-valor}_{ij} = 2 \left[1 - \Phi \left(\left| \frac{\hat{g}_{ij}}{\sqrt{\text{Var}(\hat{g}_{ij})}} \right| \right) \right] \quad [2.6]$$

donde $p\text{-valor}_{ij}$ es el p -valor asociado con el j -ésimo SNP en la i -ésima población, y $\Phi(\bullet)$ es la distribución acumulada normal estándar (Gualdrón Duarte *et al.*, 2014).

2.3.5. Análisis de la estructura genética de la población.

Como se mencionó en la sección 2.1., la estructura genética hace referencia a la presencia de individuos en una población con composición genética diferente (Price *et al.*, 2010). En las poblaciones animales, es posible observar dos tipos de estructura genética. En primer lugar, puede existir estructura genética dentro de cada población. Adicionalmente, es factible observar estructura genética entre poblaciones, considerando la heterogeneidad en el origen racial de los ancestros de cada población. Un enfoque comúnmente empleado para tener en cuenta la estructura genética de la población consiste en incluir los componentes principales (obtenidos a partir de la descomposición de la matriz de relaciones genómicas) como efectos fijos en el modelo de evaluación (Price *et al.*, 2010). Sin embargo, si se incluye en el modelo la matriz \mathbf{G} con una definición adecuada del efecto aleatorio de animal es posible tener en cuenta toda la estructura genética, haciendo innecesario el ajuste de componentes principales (Lans *et al.*, 2012). A modo de comprobación durante la investigación actual se compararon los efectos de SNP estimados a partir de modelos GBLUP que incluían componentes principales como efectos fijos, respecto de modelos que ignoraban dichas fuentes de variación. En ambos casos se retuvo el efecto aditivo de animal en el modelo, con una matriz de (co)varianzas proporcional a \mathbf{G} , siguiendo el enfoque de Lans *et al.* (2012).

Para determinar el número de componentes principales a incluir, se factorizó \mathbf{G} empleando la descomposición en los autovalores tal que $\mathbf{G} = \mathbf{U}\mathbf{D}\mathbf{U}'$, donde \mathbf{U} ($n \times n$, con n igual al número de animales genotipados) es la matriz de autovectores de \mathbf{G} , mientras que \mathbf{D} es una matriz diagonal con los autovalores. En consecuencia, se incluyeron como efectos fijos en el modelo aquellos autovectores que explicaron una fracción importante de la varianza genómica (en este caso, los dos primeros componentes principales), dando lugar a la siguiente ecuación del modelo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}_1\alpha_1 + \mathbf{U}_2\alpha_2 + \mathbf{a} + \mathbf{e} \quad [2.7]$$

En adición a aquellos elementos definidos en [2.1], los escalares α_i ($i = 1, 2$) corresponden a los coeficientes para los dos primeros componentes principales \mathbf{U}_1 y \mathbf{U}_2 . Finalmente, se estimaron los componentes de varianza y los efectos de SNP para los modelos [2.7] y [2.1].

2.3.6. Análisis de asociación conjunto.

Se ajustaron distintos modelos JA empleando en todos los casos el modelo animal centrado [2.1]. En todos los enfoques, \mathbf{X} fue la matriz de incidencia que relaciona los registros al vector de efectos fijos $\boldsymbol{\beta}$ construida en bloques para cada población, con el objeto de ajustar los mismos efectos fijos de los estudios de asociación a nivel poblacional (efecto de grupo de contemporáneos en las poblaciones Comercial y MSUPRP; grupo de contemporáneos, sexo y edad al sacrificio en población MARC). Previo a la predicción de los efectos de SNP, se estimaron las varianzas de los efectos

de SNP y los p -valores para identificar asociaciones entre los SNP y el carácter evaluado.

2.3.6.1. *Estimación de componentes de varianza asumiendo una matriz de relaciones genómicas \mathbf{G} diagonal en bloques y homogeneidad de varianzas.*

Dado que las tres poblaciones evaluadas son independientes, es razonable asumir que las relaciones aditivas entre poblaciones no son informativas sobre la estructura genética o los efectos de QTL. Por lo tanto, se ajustó un modelo de análisis conjunto siguiendo el modelo [2.1] con las siguientes especificaciones: \mathbf{a} , vector aleatorio de valores de cría, sigue una distribución tal que $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$. Asimismo, \mathbf{G} es la matriz de relaciones genómicas construida de modo diagonal en bloques, los cuales contienen las matrices genómicas específicas de cada población. Los bloques no diagonales son iguales a cero. Entonces:

$$\text{Var}(\mathbf{a}) = \begin{bmatrix} \mathbf{G}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{G}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{G}_i \end{bmatrix} \sigma_a^2 \quad [2.8]$$

con $i = 1, \dots, n$ y n siendo el número de poblaciones. El vector \mathbf{e} de efectos residuales se distribuye de modo tal que $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$. Los escalares σ_a^2 y σ_e^2 representan las varianzas genética aditiva y residual, respectivamente, las que fueron asumidas homogéneas entre poblaciones. Consecuentemente, las estimaciones de los efectos de SNP de cada población fueron obtenidas mediante la siguiente formulación:

$$\begin{aligned} \hat{\mathbf{g}} = \mathbf{Z}'\mathbf{G}^{-1}\hat{\mathbf{a}} &= \begin{bmatrix} \mathbf{Z}'_1 & 0 & \cdots & 0 \\ 0 & \mathbf{Z}'_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{Z}'_i \end{bmatrix} \begin{bmatrix} \mathbf{G}_1^{-1} & 0 & \cdots & 0 \\ 0 & \mathbf{G}_2^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{G}_i^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \\ \vdots \\ \hat{\mathbf{a}}_i \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Z}'_1\mathbf{G}_1^{-1} & 0 & \cdots & 0 \\ 0 & \mathbf{Z}'_2\mathbf{G}_2^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{Z}'_i\mathbf{G}_i^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \\ \vdots \\ \hat{\mathbf{a}}_i \end{bmatrix} \quad [2.9] \\ &= \begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \\ \vdots \\ \hat{\mathbf{g}}_i \end{bmatrix} \end{aligned}$$

2.3.6.2. Estimación de componentes de varianza asumiendo varianza genética heterogénea pero varianza residual homogénea.

De modo similar al enfoque anterior, se ajustó un JA siguiendo el modelo [2.1] con la matriz \mathbf{G} diagonal en bloques. Aquellos bloques diagonales fueron iguales a las matrices genómicas específicas de cada población y los no diagonales iguales a cero. Sin embargo, se consideró heterocedasticidad entre las varianzas genéticas de las tres poblaciones, de modo tal que:

$$\text{Var}(\mathbf{a}) = \begin{bmatrix} \mathbf{G}_1\sigma_{a_1}^2 & 0 & \cdots & 0 \\ 0 & \mathbf{G}_2\sigma_{a_2}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{G}_i\sigma_{a_i}^2 \end{bmatrix} \quad [2.10]$$

con $i = 1, \dots, n$ y n siendo el número de poblaciones. Los valores de cría se distribuyeron entonces como $\mathbf{a} \sim N(0, \bigoplus_{i=1}^n \mathbf{G}_i\sigma_{a_i}^2)$, con n correspondiendo al número de poblaciones y donde \bigoplus es el operador suma directa y \mathbf{G}_i y $\sigma_{a_i}^2$ son la matriz de relaciones genómicas y la varianza genética estimadas para la i -ésima población, respectivamente. Similar al modelo anterior, se asumió homogeneidad de varianza residual entre poblaciones: $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$. Dada la estructura de \mathbf{G} y las varianzas aditivas, los efectos de SNP fueron predichos por ponderaciones genéticas distintas dentro de cada población.

2.3.6.3. Estimación de componentes de varianza asumiendo varianzas genéticas y residuales heterogéneas.

Se ajustó además un modelo asumiendo \mathbf{G} diagonal en bloques y heterocedasticidad de las varianzas genéticas ($\mathbf{a} \sim N(0, \bigoplus_{i=1}^n \mathbf{G}_i\sigma_{a_i}^2)$) y ambientales ($\mathbf{e} \sim N(0, \bigoplus_{i=1}^n \mathbf{I}_i\sigma_{e_i}^2)$) entre poblaciones, con n siendo el número de poblaciones. La estructura de dichas covarianzas fue entonces la siguiente:

$$\text{Var}(\mathbf{a}) = \begin{bmatrix} \mathbf{G}_1\sigma_{a_1}^2 & 0 & \cdots & 0 \\ 0 & \mathbf{G}_2\sigma_{a_2}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{G}_i\sigma_{a_i}^2 \end{bmatrix} \quad \text{y} \quad \text{Var}(\mathbf{e}) = \begin{bmatrix} \mathbf{I}_1\sigma_{e_1}^2 & 0 & \cdots & 0 \\ 0 & \mathbf{I}_2\sigma_{e_2}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{I}_i\sigma_{e_i}^2 \end{bmatrix} \quad [2.11]$$

Al igual que en los JA anteriores, el modelo JA [2.11] ponderó diferencialmente los efectos de SNP de cada población.

2.3.6.4. Estimación de componentes de varianza a partir de la combinación de genotipos y fenotipos provenientes de cada población.

A diferencia de los anteriores JA, ahora se combinaron fenotipos y genotipos de todas las poblaciones asumiendo que todos los animales pertenecían a única población (Bravata y Olkin, 2001). Por lo tanto, se construyó una matriz de incidencia X común para todos los individuos, que incluyó como efecto fijo el grupo de contemporáneos (efecto ajustado en todas las poblaciones). Además, se creó un archivo único de genotipos, manteniendo aquellos animales con genotipos y fenotipos completos para los caracteres evaluados, así como los marcadores en común después de filtrar cada conjunto de datos, cómo se explicó en la sección Materiales. Posteriormente, se construyó la matriz Z considerando SNP en común para las tres poblaciones, teniendo en cuenta las frecuencias alélicas conjuntas sobre todos los conjuntos de datos y se calculó la matriz de relaciones genómicas G general para todos los individuos como $G = ZZ'$. Se asumió homogeneidad en las varianzas genética, $a \sim N(0, G\sigma_a^2)$ y residual $e \sim N(0, I\sigma_e^2)$, con estimaciones únicas de σ_a^2 y σ_e^2 entre poblaciones. Este modelo dio lugar a efectos únicos de SNP estimados con información de las tres poblaciones.

2.4. Resultados.

2.4.1. Exactitud en la imputación de genotipos.

Las Figuras 2.4.1.1. a y b respectivamente presentan, los resultados relacionados con la exactitud de la imputación ajustada por la posición de cada SNP dentro del cromosoma y por MAF.

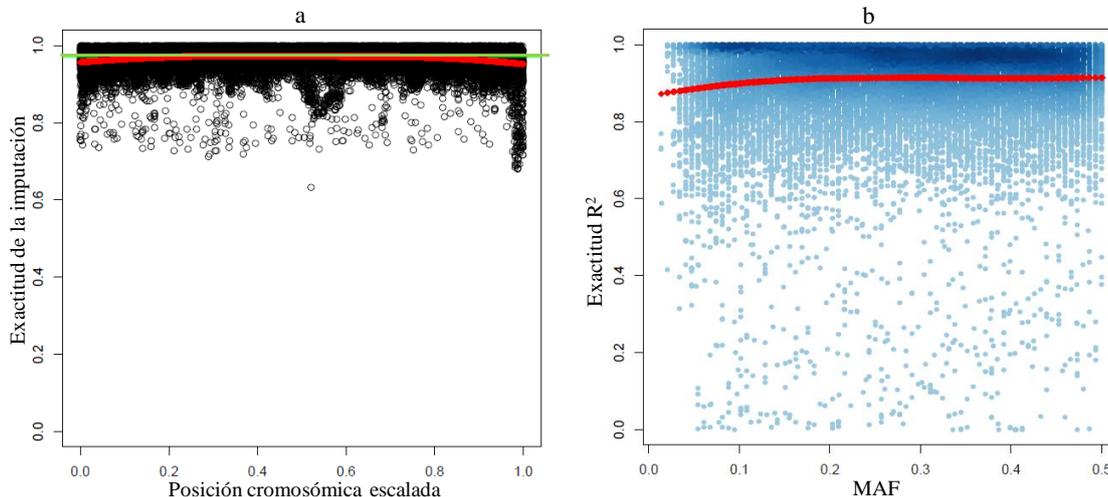


Figura 2.4.1.1. Gráficos de exactitud de la imputación de genotipos. Exactitud de la imputación para la población Comercial: a. Ajustada por posición del SNP dentro del cromosoma (línea verde representa el promedio de la exactitud de la imputación); b. Correlación entre dosis alélica observada e imputada ajustada por Frecuencia del alelo menor (MAF). La línea roja representa la media ponderada estimada bajo la técnica de regresión LOESS (regresión polinomial por segmentos; Cleveland *et al.*, 1992).

Se observa en la Figura 2.4.1.1.a que el promedio en la exactitud de la imputación fue 0.9683 (línea verde). Para los SNP ubicados en los extremos del cromosoma, la exactitud de la imputación fue cercana a 0.952, valor inferior al observado en aquellos SNP ubicados en el centro del cromosoma, el cual fue aproximadamente igual a 0.973. Asimismo, la exactitud en la imputación fue levemente mayor en aquellos SNP con MAF superior a 0.2 (Figura 2.4.1.1.b).

2.4.2. Variación fenotípica entre poblaciones.

La Tabla 2.4.2.1 presenta estadísticos descriptivos para el carácter relacionado con el color rojo en el musculo *longissimus dorsi*, o CIE a*, entre poblaciones. Para las poblaciones MARC y MSUPRP, el número de registros disponible fue relativamente pequeño, a diferencia de la población comercial la cual presentó el mayor tamaño muestral. Los valores medios oscilaron entre 6.75 a 17.26, siendo menores en la población MARC que en las poblaciones Comercial y MSUPRP.

Tabla 2.4.2.1. Estadística descriptiva para CIE a* a través de las poblaciones evaluadas.

	Población		
	Comercial	MARC	MSUPRP
No. Registros	1780	704	874
Media (Desvío Estándar)	14.49 (1.495)	6.746 (1.428)	17.26 (1.827)
Valores mínimos-máximos	9.238 – 19.360	2.525 – 10.960	13.23– 23.55

2.4.3. Estimación de componentes de varianza y heredabilidad

La Tabla 2.4.3.1 presenta las estimaciones de los parámetros genéticos y las heredabilidades para CIE a* obtenidos a partir del análisis de asociación a nivel poblacional y de los análisis de asociación conjuntos (**JA**) descritos en la sección 2.2.6.

Tabla 2.4.3.1. Componentes de varianza y heredabilidades para CIE a* a través de las poblaciones evaluadas y bajo análisis conjunto de asociación.

Análisis	Varianza Genética		Varianza Residual	h^2
GWA_com	0.8990		1.1030	0.4490
GWA_MARC	0.1310		0.6880	0.1600
GWA_MSU	0.5520		0.3630	0.6030
JA_1	0.7813		0.7539	0.5089
JA_2	Com	1.6051	0.6419	0.7143
	MARC	0.1652		0.2047
	MSU	0.3446		0.3493
JA_3	Com	0.8995	1.1029	0.4492
	MARC	0.1311	0.6882	0.1600
	MSU	0.5523	0.3634	0.6031
JA_4	0.8354		0.7651	0.5219

GWA_com: GWA población Comercial; **GWA_MARC:** GWA población MARC; **GWA_MSU:** GWA población MSUPRP; **JA_1:** Análisis de asociación conjunto asumiendo una matriz de relaciones genómicas G diagonal en bloques y homogeneidad de varianzas; **JA_2:** asumiendo varianza genética heterogénea pero varianza residual homogénea; **JA_3:** asumiendo varianzas genéticas y residuales heterogéneas; **JA_4:** combinación de genotipos y fenotipos provenientes de cada población.

En primer lugar, las varianzas genéticas se estimaron dentro del rango 0.131 (GWA población MARC) a 1.6051 (población Comercial JA bajo heterocedasticidad de varianza genética). Asimismo, las varianzas residuales oscilaron entre 0.363 (análisis de asociación población MSUPRP y análisis conjunto bajo heterocedasticidad de varianza genética en la misma población) y 1.103 (análisis de asociación población Comercial y análisis conjunto bajo heterocedasticidad de varianza genética en la misma población). Además, las estimaciones de heredabilidad tuvieron un rango amplio entre 0.160 a 0.7143.

Comparando los resultados de cada uno de los modelos ajustados y considerando que bajo algunos enfoques se obtuvieron estimaciones únicas entre poblaciones, o específicas para cada conjunto de datos, se observa que las estimaciones de varianza genética, residual y heredabilidad generadas en cada análisis de asociación dentro de población fueron consistentes con los valores obtenidos a partir del JA asumiendo heterocedasticidad de ambas varianzas. No obstante, asumir heterogeneidad sólo a nivel genético produjo estimaciones de σ_a^2 y de heredabilidad mayores en las poblaciones Comercial y MARC en comparación con las estimaciones para la población MSUPRP. Las estimaciones del JA asumiendo σ_a^2 heterogénea y σ_e^2 homogénea fueron menores que las obtenidas con el JA bajo heterocedasticidad completa aditiva y residual.

Asimismo, comparado con los modelos con estimaciones únicas de varianza y heredabilidad, los análisis JA combinando todas las poblaciones mostraron

componentes de varianza y la heredabilidad más altos que los JA donde se asumía una matriz de relaciones genómicas G en bloques pero homogeneidad de varianza genética.

2.4.4. Análisis de la estructura genética de la población.

Siguiendo el enfoque de Lans *et al.* (2012) y considerando tanto los GWAS poblacionales como el JA basado en la combinación de todos los animales en un único conjunto de datos, se factorizó la matriz G de cada modelo con el fin de determinar el número de componentes principales necesarios para explicar la variabilidad debida a estructura genética en la población. La Figura 2.4.4.1 muestra los dos primeros componentes principales que explicaron una fracción importante de la varianza genética para CIE a* bajo GWAS poblacionales y JA.

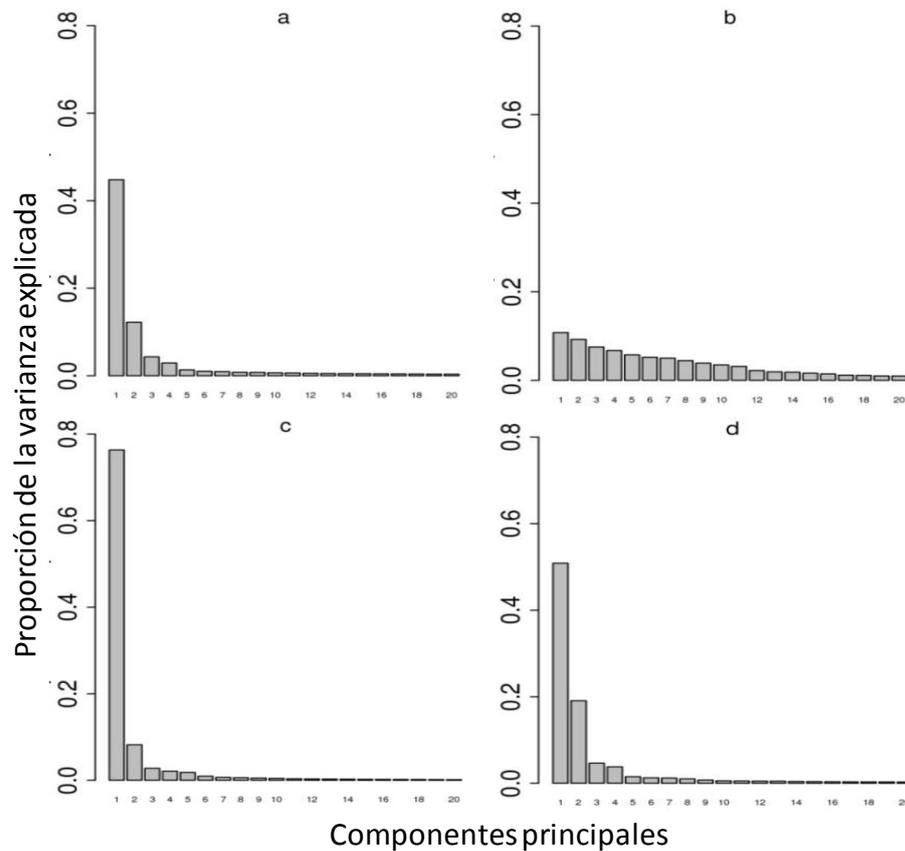


Figura 2.4.4.1. Proporción de la varianza explicada por los componentes principales. Las figuras corresponden a la proporción de varianza genética explicada por los 20 primeros autovectores, obtenidos después de la descomposición en autovalores de la matriz G para: a. población Comercial; b. Población MARC; c. Población MSUPRP; d. Análisis conjunto de asociación.

El porcentaje de varianza genética explicada para el carácter CIE a* por los dos primeros componentes osciló entre 0.2 y 0.84. Específicamente, el porcentaje fue mayor en las poblaciones Comercial (0.45 y 0.12 para el primer y segundo componente,

respectivamente) y MSUPRP (0.76 y 0.08 para el primer y segundo componente respectivamente), en comparación con la población MARC (0.11 y 0.09 para el primer y segundo componente, respectivamente). En el caso del JA, la proporción de variación explicada fue similar a la población Comercial (0.51 y 0.19). Teniendo en cuenta estos resultados, se estimaron los componentes de varianza, los efectos SNP y los p -valores para asociación a partir del modelo presentado en [2.7] para cada una de las poblaciones, así como en el JA donde se combinaron todas las poblaciones. La Figura 2.4.4.2 presenta los gráficos Q-Q de comparación de p -valores incorporando o no componentes principales a cada uno de los modelos evaluados.

Como se observa en la Figura 2.4.4.2, los p -valores obtenidos a partir de modelos incluyendo o excluyendo componentes principales estuvieron altamente correlacionados en cada análisis ($R^2 = 0.997, 0.984, 0.999$ para GWA en las poblaciones Comercial, MARC y MSUPRP, respectivamente, y 0.997 para JA). Estos resultados muestran que, si bien los componentes principales explicaron en general un alto porcentaje de la variación genética del carácter evaluado, la incorporación de componentes principales no produjo cambio sustancial alguno en los estadísticos de búsqueda de asociaciones significativas. Por esta razón, los resultados presentados en las siguientes secciones corresponden a modelos GBLUP que no incluyen componentes principales.

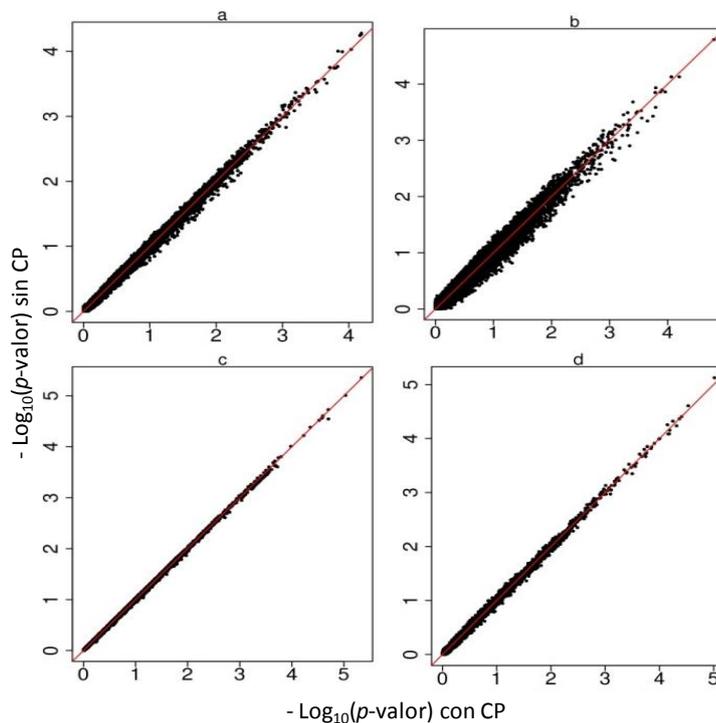


Figura 2.4.4.2. Comparación de p -valores entre modelos incluyendo y omitiendo componentes principales. Gráficos Q-Q para comparación de p -valores después de incluir (eje x) y omitir (eje y) componentes principales en: a. GWAS población comercial; b. GWAS población MARC; c. GWAS población MSUPRP; d. Análisis conjunto de asociación a partir de la combinación de poblaciones.

2.4.5. Análisis de asociación poblacional para CIE a*.

Las Figuras 2.4.5.1.a, b y c muestran los gráficos Manhattan (“Manhattan plots”) para las mediciones de CIE a* en las poblaciones Comercial, MARC y MSUPRP, respectivamente. Los picos que se observan se encontraron en el cromosoma 6 al combinar todas las poblaciones. Más allá de lo observado en el cromosoma 12 en la población MSUPRP, ningún pico alcanzó el umbral de significancia a nivel de todo el genoma. Específicamente, se consideró la corrección de Bonferroni para inferencia múltiple de 1.094×10^{-6} , 1.1359×10^{-6} y 1.2325×10^{-6} en las poblaciones Comercial, MARC y MSUPRP, respectivamente.

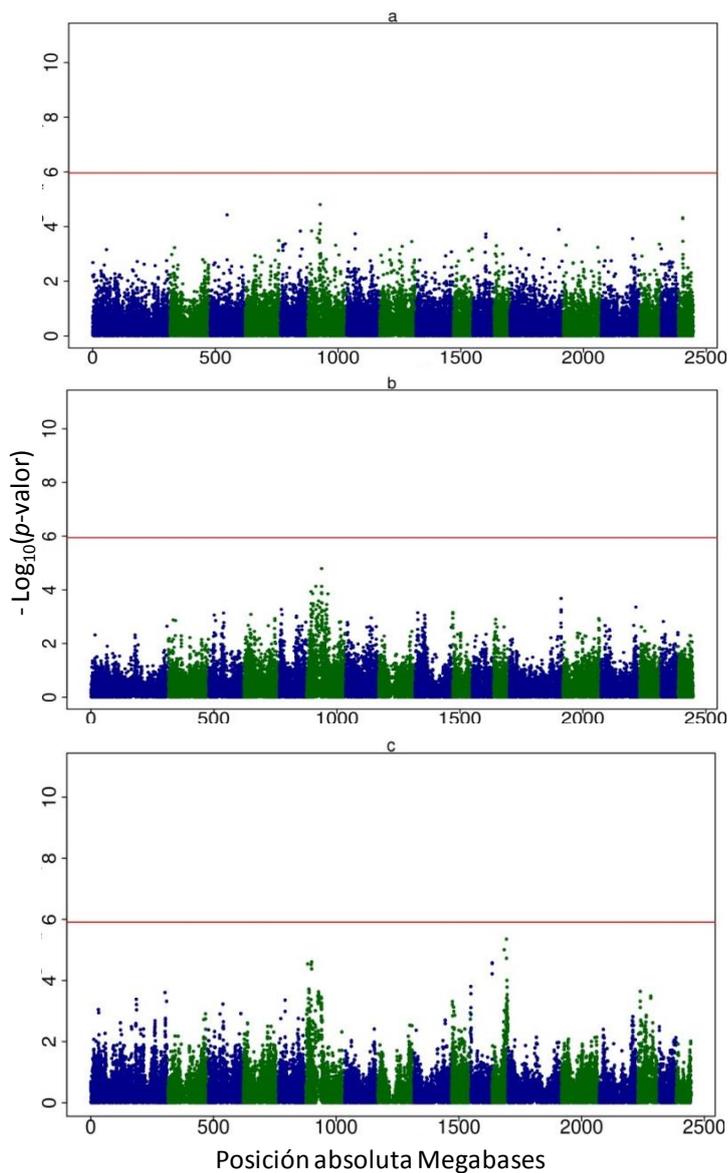


Figura 2.4.5.1. Gráficos Manhattan para CIE a* para todas las poblaciones. Asociaciones de SNP con CIE a* en la población: a. Comercial; b. MARC; c. MSUPRP. $-\text{Log}_{10}(p\text{-valor})$ en eje y versus posición absoluta del SNP en Megabases (eje x). La línea horizontal define el umbral de significancia $p < 0.05$ a nivel del genoma completo.

2.4.6. Análisis de asociación conjunto.

Dado que algunos modelos JA produjeron estimaciones de los efectos de SNP específicos para cada población, se compararon los p -valores obtenidos a partir del JA y aquellos obtenidos a partir de cada GWAS poblacional (Figuras 2.4.6.1 - 2.4.6.3).

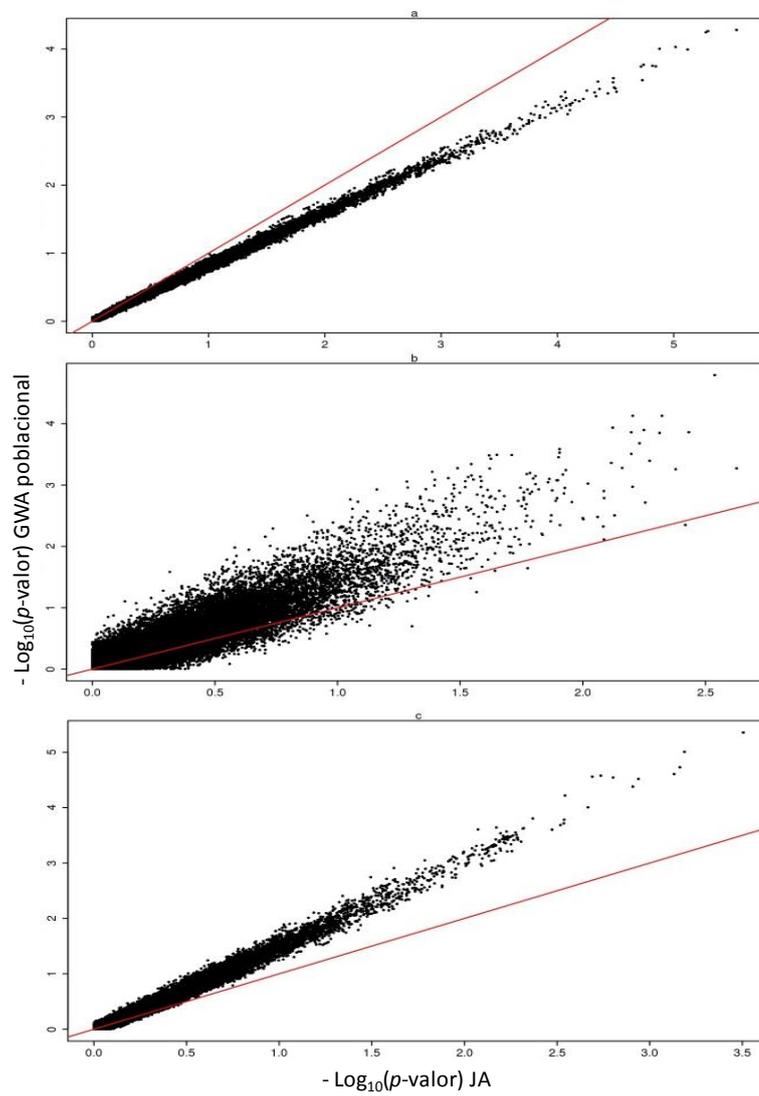


Figura 2.4.6.1. Comparación de p -valores entre GWA poblacional y JA asumiendo homocedasticidad en componentes de varianza. Gráficos Q-Q para comparación de p -valores obtenidos a partir de JA (eje x) versus GWA en población: a. Comercial (eje y); b. MARC (eje y); c. MSUPRP (eje y).

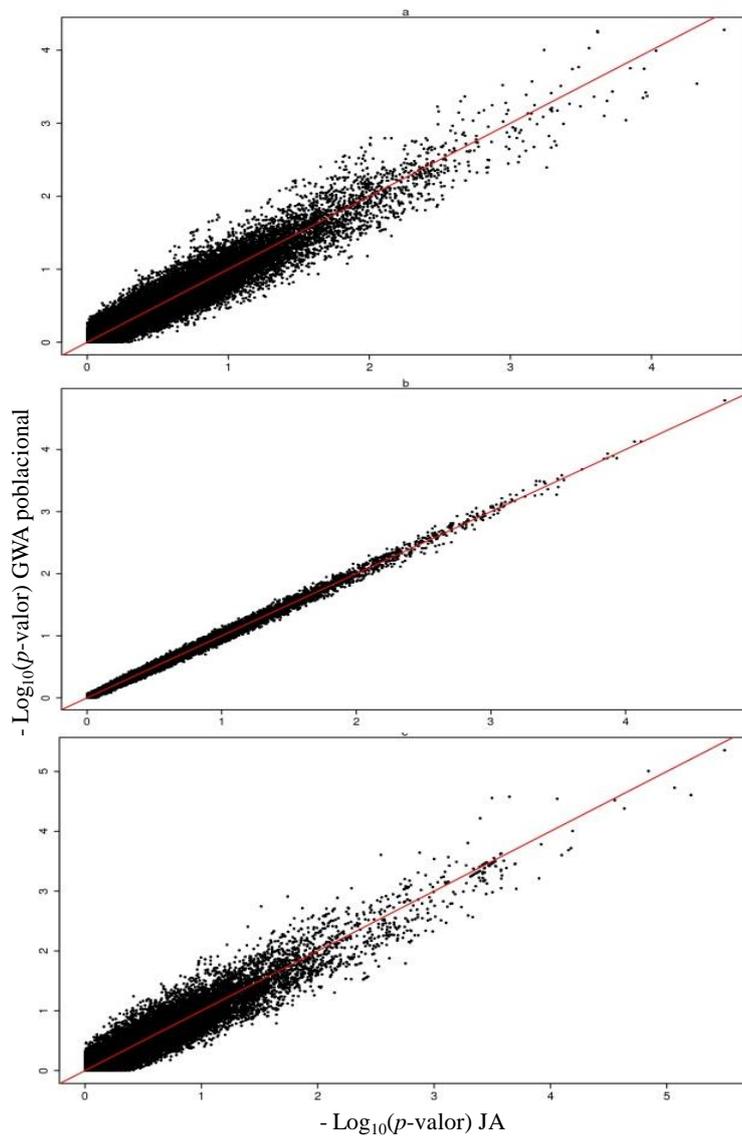


Figura 2.4.6.2. Comparación de p -valores entre GWA poblacional y JA asumiendo heterocedasticidad en varianza genética. Gráficos Q-Q para comparación de p -valores obtenidos a partir de JA (eje x) versus GWA en población: a. Comercial (eje y); b. MARC (eje y); c. MSUPRP (eje y).

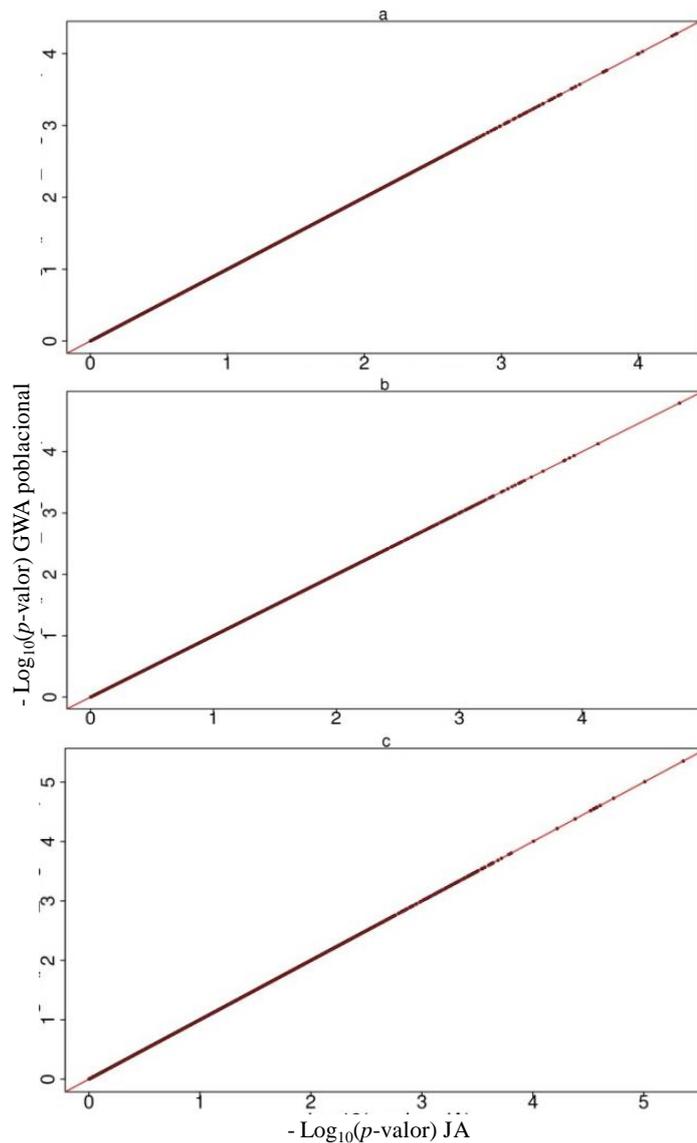


Figura 2.4.6.3. Comparación de p -valores entre GWA poblacional y JA asumiendo heterocedasticidad en componentes de varianza. Gráficos Q-Q para comparación de p -valores obtenidos a partir de JA (eje x) versus GWA en población: a. Comercial (eje y); b. MARC (eje y); c. MSUPRP (eje y).

El JA que consideró la matriz de relaciones genómicas como diagonal en bloques y homocedasticidad en los componentes de varianza mostró p -valores más bajos que el GWA para la población Comercial, pero más altos que en las otras dos poblaciones (Figura 2.4.6.1.). No obstante, se obtuvieron altos valores de correlación entre $-\text{Log}_{10}$ de los p -valores a partir de cada enfoque, siendo iguales a 0.9976, 0.8945 y 0.99 para las poblaciones Comercial, MARC y MSUPRP, respectivamente. En el caso del JA asumiendo heterocedasticidad de varianza genética y homocedasticidad en la varianza residual (Figura 2.4.6.2.), los p -valores en las poblaciones Comercial y MSUPRP tuvieron una alta dispersión alrededor de la línea 1-1, lo cual se vio reflejado en valores de correlación iguales a 0.953 y 0.938, para las poblaciones Comercial y MSUPRP,

respectivamente. No obstante, en la población MARC los p -valores calculados mediante JA y GWAS poblacional fueron muy similares, con una correlación de 0.997 entre ambos. Adicionalmente, JA considerando heterogeneidad de varianzas aditiva y residual generó resultados virtualmente idénticos a aquellos obtenidos a partir de los GWAS para cada población (Figura 2.4.6.3). Esto se confirma en los valores de correlación de 1, 0.99 y 1 obtenidos en cada una de las poblaciones evaluadas.

La Figura 2.4.6.4 presenta el gráfico Manhattan para el carácter CIE a* empleando el JA que produjo estimaciones únicas de los efectos SNP sobre las poblaciones.

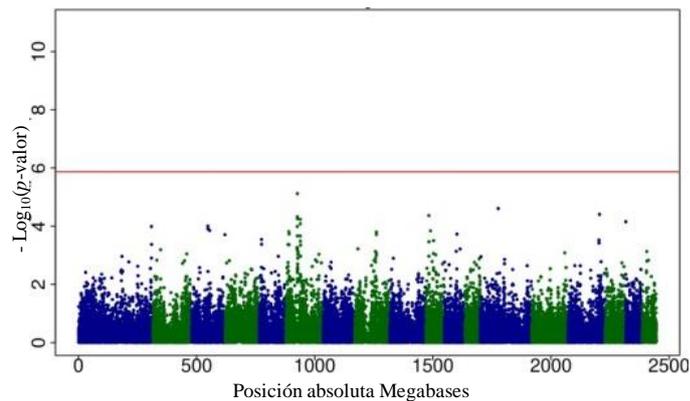


Figura 2.4.6.4. Gráfico de Manhattan para CIE a* en análisis conjunto (JA). Gráfico Manhattan para asociaciones de SNP con CIE a* considerando JA ($-\text{Log}_{10}(p\text{-valor})$ en eje y vs. posición absoluta del SNP en Megabases (eje x). La línea horizontal define el umbral de significancia $p < 0.05$ a nivel del genoma entero.

Considerando las correlaciones entre los p -valores obtenidos en los GWAS poblacionales y los calculados a partir de JA, así como el gráfico de Manhattan presentado en la Figura 2.4.6.4, se observó que ningún JA identificó asociaciones significativas para CIE a*, al igual que los análisis de asociación a nivel de poblacional individual. Es decir, el análisis conjunto de las poblaciones, sea asumiendo homocedasticidad en los componentes de varianza como heterogeneidad entre las distintas poblaciones, no condujo a aumentos en la potencia de detección de asociaciones significativas para CIE a*.

2.5. Discusión

En el presente capítulo, se mostró como desarrollar evaluaciones genómicas basadas en metodología GBLUP, y se compararon los resultados obtenidos a partir de dicho enfoque con aquellos obtenidos a partir de un análisis conjunto de asociación (JA). Dicho procedimiento fue ilustrado utilizando el carácter CIE a* debido a su importancia económica (Cannon *et al.*, 1996; Ovilo *et al.*, 2002) y debido a la presencia de heterocedasticidad al considerar todas las poblaciones (Tabla 2.4.3.1), la cual sienta precedente en relación a la importancia de modelar apropiadamente dicha

heterogeneidad en estudios de asociación genómica. Asimismo, se muestran los resultados relacionados con la imputación de genotipos, información necesaria para el desarrollo de GWAS poblacionales así como de JA.

En primer lugar, la exactitud en la imputación de los genotipos fue alta (0.969), resultado similar al informado también en cerdos por Badke *et al.* (2012) y Gualdrón Duarte *et al.* (2013). Dicha exactitud dependerá, entre otros factores, del tamaño y composición del panel de referencia, así como de la diversidad y relaciones genéticas entre los individuos genotipados en baja y alta densidad (Marchini y Howie, 2010; Hickey *et al.*, 2012). Sin embargo, además de la importancia de obtener exactitudes altas en la imputación debido a su efecto posterior sobre la exactitud de las predicciones genómicas para individuos genotipados en baja densidad (Dassonneville *et al.*, 2011; Chen *et al.*, 2014), la importancia de la imputación en el contexto de esta tesis está relacionada con el aumento en el número de SNP que pueden ser testeados para asociación con el carácter de interés. Esto permitirá aumentar la potencia de los GWAS y JA para identificar variantes causales y facilitará el posterior desarrollo de meta-análisis de asociación (Marchini y Howie, 2010).

Para considerar la heterogeneidad en la interacción SNP \times población entre los conjuntos de datos evaluados para el carácter CIE a*, se estimaron las varianzas genéticas y residuales, así como las heredabilidades (Tabla 2.4.3.1) considerando GWAS poblacionales y diferentes enfoques JA. Al comparar entre poblaciones, los valores de heredabilidad tuvieron un rango amplio, oscilando entre 0.16 y 0.60 para MARC y MSUPRP, respectivamente. Este resultado fue idéntico al obtenido a partir del JA que asumió heterocedasticidad para las varianzas genéticas y residuales. El resultado implica que dicho enfoque de JA permite considerar apropiadamente la interacción SNP \times población, resultado en concordancia con el trabajo de Möhring y Piepho (2009). En el caso del JA asumiendo varianza genética heterogénea y residual homogénea, las estimaciones de la variabilidad genética y la heredabilidad fueron mayores en las poblaciones Comercial y MARC en comparación con los GWA dentro de población, resultado opuesto al observado para la población MSUPRP. Con lo cual este enfoque de JA, desarrollado con el objetivo de capturar la posible heterogeneidad a nivel genético entre poblaciones, estaría sobre-estimando el porcentaje de variación explicado por el componente aditivo en las poblaciones con el mayor tamaño muestral. Por su parte, el ajuste de JA asumiendo una matriz de relaciones genómicas diagonal en bloques y homogeneidad de varianzas, produjo estimaciones de los componentes de varianza y las heredabilidades similares a las obtenidas a partir del JA basado en la combinación de genotipos y fenotipos provenientes de cada población. Este resultado implica que la construcción de una matriz de relaciones genómicas diagonal en bloques no es suficiente para representar la interacción SNP \times población, sugiriendo la necesidad de definir componentes de varianza específicos para cada conjunto de datos con el fin de representar la interacción apropiadamente. Además, la semejanza entre los resultados obtenidos a partir del JA basado en la combinación de las poblaciones en un único conjunto de datos y los observados en el caso del JA asumiendo una matriz de relaciones genómicas diagonal en bloques y homogeneidad de componentes de varianza, indican que el último enfoque es el menos deseable si se busca capturar la heterogeneidad en la interacción SNP \times población. En este caso, la interacción estará

incluida en la varianza residual (Möhrling y Piepho, 2009), lo cual concuerda con las estimaciones altas obtenidas para dicho componente de varianza en los dos enfoques JA mencionados.

Una contribución importante de este capítulo es la correcta definición de la metodología para la implementación de estudios de asociación siguiendo el enfoque GBLUP. En este capítulo, dicho modelo fue empleado para desarrollar GWAS a nivel poblacional así como diferentes enfoques de JA. Además, como se verá en el siguiente capítulo, los GWAS específicos dentro de población constituyen el primer paso en el desarrollo del MA, y por tanto, es relevante considerar toda la variación sistemática dentro de población. Al respecto, se siguió el enfoque propuesto por Lans *et al.* (2012) para incluir como covariables los autovectores obtenidos a partir de la matriz \mathbf{G} . Sin embargo, no se encontraron diferencias entre los p -valores para los efectos de SNP estimados a partir de los modelos incluyendo ([2.7]) y excluyendo ([2.1]) los autovectores de la matriz \mathbf{G} como covariables fijas (Figura 2.4.4.2). Con lo cual, se eliminaron los autovectores de \mathbf{G} y se definió el empleo del modelo [2.1] para todos los análisis.

Aunque se incluyeron apropiadamente las fuentes de variación dentro y entre poblaciones, los análisis de asociación a nivel poblacional no detectaron SNP relacionados significativamente con CIE a^* , dado que los picos de asociación observados no alcanzaron el umbral de significancia establecido. Esta es una situación común en muchos GWAS, donde a pesar de la existencia de algunos picos, estos no alcanzan el umbral de significancia luego de realizarse los ajustes por pruebas de hipótesis múltiples (Hang *et al.*, 2009; Roberts *et al.*, 2010). No obstante, dado que la corrección de Bonferroni es fácil de implementar y mucho menos intensiva en términos computacionales que otras metodologías como el testeo por permutación (Cordell y Clayton, 2005; Lam *et al.*, 2009), se usó dicha corrección para definir el umbral de significancia a nivel del genoma en los GWAS poblacionales y JA.

Comparando los resultados obtenidos a partir de los GWAS con aquellos obtenidos a partir de diferentes enfoques de análisis conjunto de asociación (JA), así como las ventajas y desventajas metodológicas de uno u otro enfoque, surgen varios aspectos a tener en cuenta. En primer lugar, una dificultad inicial en la implementación de los JA es la construcción de una matriz de relaciones genómicas. Si se modelan las relaciones entre poblaciones, deben calcularse las frecuencias alélicas de la población base, es decir, aquellas que existían antes de la división de las poblaciones. Desde luego, dichas estimaciones no están disponibles. Una alternativa sería el uso de las frecuencias alélicas específicas dentro de población. Un segundo desafío es la presencia de heterocedasticidad entre poblaciones. Por ejemplo y como se mencionó previamente, cuando se modeló la varianza dentro de una población específica al mismo tiempo que se asumió una covarianza igual a cero entre poblaciones (Reverter *et al.*, 2004; Möhrling y Piepho, 2009), el análisis conjunto generó los mismos resultados que los GWAS a nivel poblacional (Figura 2.4.6.3). Sin embargo, este modelo produjo estimaciones de los efectos de SNP y pruebas estadísticas específicas dentro de población. En consecuencia, surge el interrogante respecto de combinar los resultados en un estadístico único para cada SNP entre todas las poblaciones.

Finalmente, los resultados de este capítulo muestran la importancia de la correcta definición de los modelos a emplear en estudios de asociación genómica, tanto en relación al ajuste por existencia de estructura genética dentro de la población, como a la consideración apropiada la heterogeneidad en la interacción SNP \times población y de los componentes de varianza entre poblaciones.

Capítulo 3

Meta-análisis de asociación a partir de modelos de predicción genómica.

Meta-análisis de asociación a partir de modelos de predicción genómica

3.1. Introducción

Como se mencionó en el capítulo anterior, una limitación de los análisis de asociación sobre el genoma completo, o GWAS, es la escasa potencia para detectar variantes con efecto reducido, con lo cual se requiere un número mayor de individuos con el fin de observar asociaciones significativas entre los SNP y el carácter de interés (Minozzi *et al.*, 2012). En consecuencia, el desarrollo de un *análisis conjunto de asociación* (JA, del inglés “Joint Analysis”) permite aumentar el tamaño muestral a partir de la combinación de datos provenientes de diferentes poblaciones. Ocurre que en algunas ocasiones no se cuenta con los datos originales, condición necesaria para realizar JA. Además, la diferencia en las escalas de medición del fenotipo en las distintas poblaciones, así como la heterocedasticidad en los componentes de varianza, hacen que el JA sea una metodología de compleja implementación (Minozzi *et al.*, 2012). Por lo tanto, una alternativa razonable es desarrollar un *meta-análisis de estudios de asociación* (MA-GWA), metodología que combina en un estadístico único los resultados obtenidos a partir de diferentes estudios de asociación, considerando la estructura genética de la población así como covariables específicas dentro de cada población (Willer *et al.*, 2010; Evangelou y Ioannidis, 2013). Mencionado inicialmente por Glass (1976), el MA-GWA genera estimaciones más exactas de los efectos de SNP derivados a partir de valores de cría genómicos específicos para cada población mejorando el poder de detección de las asociaciones consistentemente significativas en varias poblaciones (Minozzi *et al.*, 2012).

En el contexto de genética humana y de otros organismos, el MA es una metodología utilizada ampliamente (Begum *et al.*, 2012; Evangelou y Ioannidis, 2013), a diferencia del mejoramiento genético animal donde su difusión es reciente. Por ejemplo, Wood *et al.* (2006) desarrollaron un MA basado en modelos jerárquicos bayesianos con el fin de evaluar la consistencia de las asociaciones entre el gen de Tiroglobulina (TG5) y el marmóreo de la carne. En adición, Silva *et al.* (2010) realizaron un MA a partir de resultados de análisis de ligamiento independientes, con el objetivo de encontrar QTL para características de res, grasa y caracteres de crecimiento. De modo similar, Porto Neto *et al.* (2010) combinaron en un MA los resultados obtenidos a partir de GWAS y estudios de expresión génica de resistencia a garrapatas. Ciertamente, los trabajos mencionados son antecedentes de la aplicación del MA en producción animal, mostrando su utilidad al momento de cuantificar la variación genética que ocurre en caracteres complejos y también, la posibilidad que ofrecen de tener en cuenta la estructura genética existente en las poblaciones evaluadas.

Teniendo en cuenta que el MA se centra en el análisis de asociación de poblaciones independientes, no debe soslayarse el ajuste y correcta especificación del modelo utilizado en los GWAS. Es así que es relevante el concepto de desequilibrio gamético o,

más conocido, de ligamiento (LD por su definición en inglés “*Linkage disequilibrium*”), definido como la asociación no aleatoria entre alelos en dos loci polimórficos (Crow y Kimura, 1970). En general, la selección genómica se basa en la existencia de suficiente LD entre marcadores y QTL tal que las fases de los alelos de los SNP junto con los alelos del QTL persistan entre distintas generaciones (de Roos *et al.*, 2008). La persistencia de la fase del LD entre dos poblaciones puede explicar porque los marcadores en LD que fueron descubiertos en una población pueden no confirmarse en otra población (de Roos *et al.*, 2008). Por lo tanto, es importante conocer la persistencia de la fase de LD entre poblaciones, para poder evaluar la posibilidad de aprovechar la existencia de LD e incrementar la potencia de los análisis de asociación (Briscoe *et al.*, 1994).

Dado el importante empleo de la metodología GBLUP en estudios GWA dentro del mejoramiento genético animal (Wang *et al.*, 2012, 2014; Gualdrón Duarte *et al.*, 2014), particularmente para considerar la estructura genética de las poblaciones, es posible desarrollar e implementar MA-GWAS para diferentes poblaciones a partir de evaluaciones genómicas GBLUP. Esta alternativa requiere, entre otros aspectos importantes, de una definición adecuada de los esquemas de ponderación para los efectos de cada marcador en las distintas poblaciones evaluadas. Con lo cual, el objetivo del presente capítulo se centra en la descripción de los métodos para la implementación del MA-GWAS, combinando resultados de evaluaciones GBLUP independientes, y considerando la estructura genética de la población así como la heterogeneidad en los componentes de varianza. En primer lugar, se comparará el alcance de LD y la persistencia de la fase de LD entre diferentes poblaciones. Además, dado que en el capítulo anterior se presentaron los métodos relacionados con el desarrollo de GWAS empleando modelos mixtos y GBLUP, en este capítulo se describe el proceso para implementar MA-GWAS, mostrando cómo obtener los valores de ponderación adecuados para integrar los resultados de estudios múltiples de asociación, en un estadístico único de meta-análisis. En este caso, se demuestra la equivalencia entre las pruebas estadísticas derivadas a partir de modelos que asumen efectos de SNP fijos, en relación con aquellas que asumen dichos efectos como aleatorios. Posteriormente se aplica esta metodología desarrollando un MA-GWAS en un conjunto de datos reales obtenidos a partir de tres poblaciones de cerdos, y se comparan los resultados con aquellos obtenidos en el capítulo anterior empleando JA, y teniendo en cuenta la combinación de datos de todas las poblaciones.

3.2. Materiales.

Al igual que lo realizado en el capítulo anterior y para facilitar la posterior comparación de resultados obtenidos a partir del análisis de asociación conjunto (JA) y del meta-análisis de GWAS (MA-GWAS), se utilizaron datos correspondientes a mediciones objetivas del color rojo de la carne o CIE a* (CIE International, 1976) en la superficie del musculo *longissimus dorsi* después del sacrificio. Dichos registros se obtuvieron a partir de tres poblaciones de cerdos (población experimental de cerdos de Michigan State University – MSUPRP, East Lansing, Michigan, USA; población del centro de investigación Meat Animal Research Center – MARC, Clay Center,

Nebraska, USA; población comercial), las cuales fueron descritas ampliamente en la sección 2.2 del capítulo anterior. Como se indicó en dicha sección, en las tres poblaciones se descartaron aquellos datos provenientes de individuos con más de 10% de genotipos faltantes así como marcadores con MAF < 0.05 y más de 10% de datos faltantes (Badke *et al.*, 2012). El proceso de edición dio lugar a un total de 4054 individuos y un número específico de marcadores para cada población (Tabla 2.2.1).

3.3. Métodos

3.3.1. Estimación del desequilibrio de ligamiento y persistencia de la fase entre poblaciones.

Siguiendo el enfoque de Badke *et al.* (2012), el desequilibrio de ligamiento o LD para todos los pares de SNP fue estimado usando la siguiente expresión:

$$r_{ij}^2 = \frac{(p_{ij} - p_i p_j)^2}{p_i(1-p_i) \times p_j(1-p_j)} \quad [3.1]$$

donde p_i y p_j son las frecuencias alélicas marginales en el i -ésimo y j -ésimo SNP respectivamente y p_{ij} es la frecuencia del haplotipo de los dos marcadores. En este caso, se agruparon los SNP por pares, considerando su distancia física, en intervalos de 100 Kb y en el rango 0 Mb - 10 Mb.

Por otra parte, para estimar la persistencia de la fase de LD entre poblaciones, se incluyeron solamente marcadores con un valor de frecuencia del alelo B superior a 0.05. A partir de estos SNP, la persistencia de la fase fue estimada como:

$$R_{k,k'} = \frac{\sum_{(i,j) \in p} (r_{ij(k)} - \bar{r}_{(k)})(r_{ij(k')} - \bar{r}_{(k')})}{S_k S_{k'}} \quad [3.2]$$

donde $R_{k,k'}$ es la correlación de la fase entre $r_{ij(k)}$ en la población k y $r_{ij(k')}$ en la población k' , $S(k)$ y $S(k')$ son la desviación estándar de $r_{ij(k)}$ y $r_{ij(k')}$ respectivamente, y $\bar{r}_{(k)}$ y $\bar{r}_{(k')}$ son el promedio r_{ij} entre el i -ésimo y j -ésimo SNP, dentro del intervalo p para las poblaciones k y k' , respectivamente.

De este modo, los gráficos resultantes de las estimaciones de LD dentro de población y para la persistencia de la fase de LD entre poblaciones se presentarán en la sección de resultados del presente capítulo.

3.3.2. Desarrollo de estudios de asociación genómica basados en modelos GBLUP.

Se estimaron los componentes de varianza y los valores de cría para los animales genotipados de manera similar a lo hecho en el capítulo 2, siguiendo un modelo animal centrado con ecuación igual a:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{a} + \mathbf{e} \quad [3.3]$$

En [3.3] \mathbf{y} es el vector que contiene los fenotipos de los animales evaluados; \mathbf{X} es la matriz de incidencia que relaciona los registros al vector de efectos fijos $\boldsymbol{\beta}$; \mathbf{a} es el vector aleatorio de valores de cría, que sigue una distribución $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$, siendo \mathbf{G} la matriz de relaciones genómicas ($n \times n$; n el número de animales genotipados) calculada por $\mathbf{G} = \mathbf{Z}\mathbf{Z}'$. La matriz \mathbf{Z} contiene las dosis alélicas normalizadas según [2.2]. Como se mencionó previamente, la matriz \mathbf{G} es escalada y por lo tanto, análoga a la matriz de relaciones de parentesco \mathbf{A} . Asimismo, \mathbf{e} es el vector de efectos residuales, tal que $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$. Los escalares σ_a^2 y σ_e^2 representan las varianzas genética aditiva y residual, respectivamente.

El modelo presentado en [3.3] permite obtener los valores de cría para los individuos evaluados, estimaciones que son utilizadas para calcular los efectos de SNP a partir de la transformación lineal $\hat{\mathbf{g}} = \mathbf{Z}'\mathbf{G}^{-1}\hat{\mathbf{a}}$ (ecuación [2.4]) y, posteriormente, obtener la varianza como $\text{Var}(\hat{\mathbf{g}}) = \mathbf{Z}'\mathbf{G}^{-1}\mathbf{Z}\sigma_a^2 - \mathbf{Z}'\mathbf{G}^{-1}\mathbf{C}^{aa}\mathbf{G}^{-1}\mathbf{Z}$ (ecuación [2.5]). En esta expresión, σ_a^2 corresponde a la varianza genética y \mathbf{C}^{aa} es la porción de la inversa de la matriz de coeficientes de las ecuaciones de modelos mixtos asociada con los valores de cría. De este modo, es posible calcular p -valores para identificar los efectos de SNP significativos, siguiendo el enfoque presentado por Gualdrón Duarte *et al.* (2014) (ecuación [2.6]).

3.3.3. Análisis de la estructura genética dentro de población.

Considerando que los GWAS a nivel poblacional constituyen la base del desarrollo del MA-GWAS, es importante modelar apropiadamente la estructura genética dentro de cada población previo a la combinación en estadísticos únicos entre los conjuntos de datos evaluados, con el fin de disminuir y/o evitar la presencia de falsas asociaciones SNP-QTL. Sin embargo, como se demostró en la sección 2.4.4 del capítulo anterior, la incorporación de componentes principales para la corrección por estructura genética dentro de cada población no resultó en cambios relevantes en los valores de las pruebas estadísticas realizadas (Figura 2.4.4.2). Por esta razón, no se incluyeron componentes principales en los análisis de asociación a nivel poblacional.

3.3.4. Equivalencia entre test estadísticos basados en modelos animales centrados y test basados en modelos asumiendo efectos de SNP fijos.

Si bien el modelo [3.3] asume que los efectos de SNP son aleatorios, se puede demostrar algebraicamente que dicho enfoque es equivalente a un modelo que ajusta un SNP a la vez considerado fijo, mientras que tiene en cuenta los efectos poligénicos aleatorios empleando la matriz de relaciones genómicas \mathbf{G} . Consecuentemente, los p -valores y las pruebas estadísticas calculadas a partir de la expresión [2.6] son idénticos a aquellos que se obtienen a partir del algoritmo de asociación eficiente de modelos mixtos EMMAX (del inglés: *Efficient Mixed-Model Association Expedited*; Kang *et al.*, 2010; Zhang *et al.*, 2010). De gran difusión, el algoritmo EMMAX es una estrategia eficiente en términos computacionales y con propiedades estadísticas favorables al momento de estimar efectos fijos en modelos mixtos (Price *et al.*, 2010; Wu *et al.*, 2011). Entonces, para demostrar la equivalencia antes mencionada, considérese el siguiente modelo animal centrado:

$$\mathbf{y} = \mathbf{a} + \mathbf{e} \quad [3.4]$$

donde $\mathbf{a} \sim N(\mathbf{0}, \mathbf{Z} \mathbf{Z}' \sigma_u^2)$ y $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I} \sigma_e^2)$. En este caso, las soluciones para los efectos aleatorios de animal son iguales a:

$$\hat{\mathbf{a}} = \left(\mathbf{I} \sigma_e^{-2} + (\mathbf{Z} \mathbf{Z}')^{-1} \sigma_u^{-2} \right)^{-1} \mathbf{y} \sigma_e^{-2} \quad [3.5]$$

Strandén y Garrick (2009), Badke *et al.* (2014) y Gualdrón Duarte *et al.* (2014), observaron que el modelo presentado en [3.4] es equivalente al siguiente modelo de SNP centrado:

$$\mathbf{y} = \mathbf{Z} \mathbf{g} + \mathbf{e} \quad [3.6]$$

con $\mathbf{g} \sim N(\mathbf{0}, \mathbf{I} \sigma_u^2)$ y $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I} \sigma_e^2)$. Los efectos de SNP son estimados a partir de la expresión:

$$\hat{\mathbf{g}} = \left((\mathbf{Z} \mathbf{Z}')^{-1} \sigma_e^{-2} + \mathbf{I} \sigma_u^{-2} \right)^{-1} \mathbf{Z}' \mathbf{y} \sigma_e^{-2} \quad [3.7]$$

Teniendo en cuenta el modelo [3.6], los efectos de SNP pueden expresarse a partir de una transformación lineal de los valores de cría, es decir, $\hat{\mathbf{g}} = \mathbf{Z}' (\mathbf{Z} \mathbf{Z}')^{-1} \hat{\mathbf{a}}$. Además, la varianza del error de predicción de $\hat{\mathbf{g}}$ está definida como:

$$\text{Var}(\hat{\mathbf{g}} - \mathbf{g}) = \left((\mathbf{Z} \mathbf{Z}')^{-1} \sigma_e^{-2} + \mathbf{I} \sigma_u^{-2} \right)^{-1} \quad [3.8]$$

donde $\text{Var}(\hat{\mathbf{g}}) = \mathbf{I}\sigma_u^2 - \left((\mathbf{Z}\mathbf{Z}')^{-1} \sigma_e^{-2} + \mathbf{I}\sigma_u^{-2} \right)^{-1}$. En este contexto, el test estadístico para el i -ésimo efecto de SNP en [3.6] es igual a:

$$tg_i = \frac{\hat{g}_i}{\sqrt{\text{Var}(\hat{g}_i)}} \quad [3.9]$$

Ahora bien, el modelo EMMAX (Kang *et al.*, 2010; Zhang *et al.*, 2010) consiste en:

$$\mathbf{y} = \mathbf{z}_i \mathbf{b}_i + \mathbf{a} + \mathbf{e} \quad [3.10]$$

con $\mathbf{a} \sim N(\mathbf{0}, \mathbf{Z}\mathbf{Z}'\sigma_u^2)$ y $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$. Equivalentemente, el modelo [3.10] puede escribirse de manera tal que $\mathbf{y} = \mathbf{z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}$, donde $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V})$ y $\mathbf{V} = \mathbf{Z}\mathbf{Z}'\sigma_u^2 + \mathbf{I}\sigma_e^2$. En este modelo, la solución para el i -ésimo efecto de SNP es como sigue:

$$\hat{\mathbf{b}}_i = \left(\mathbf{z}_i' \mathbf{V}^{-1} \mathbf{z}_i \right)^{-1} \mathbf{z}_i' \mathbf{V}^{-1} \mathbf{y} = \frac{\mathbf{z}_i' \mathbf{V}^{-1} \mathbf{y}}{\mathbf{z}_i' \mathbf{V}^{-1} \mathbf{z}_i} \quad [3.11]$$

donde $\text{Var}(\hat{\mathbf{b}}_i) = \left(\mathbf{z}_i' \mathbf{V}^{-1} \mathbf{z}_i \right)^{-1}$. En este caso, el test estadístico obtenido a partir de la

ecuación [3.11] es igual a $tb_i = \frac{\hat{b}_i}{\sqrt{\text{Var}(\hat{b}_i)}}$. Para demostrar que el estadístico tg_i ,

calculado a partir de [3.9] es una solución factible en términos computacionales al estadístico que se obtiene a partir del modelo [3.10], debemos demostrar que

$$\text{Var}(\hat{g}_i) = \frac{(\sigma_u^2)^2}{\text{Var}(\hat{b}_i)} \text{ y que } \hat{g}_i = \frac{\sigma_u^2}{\text{Var}(\hat{b}_i)} \hat{b}_i.$$

a. *Demostración de* $\text{Var}(\hat{g}_i) = \frac{(\sigma_u^2)^2}{\text{Var}(\hat{b}_i)}$

En primer lugar, en la expresión $\mathbf{I}\sigma_u^2 - \left(\mathbf{Z}\mathbf{Z}'\sigma_e^{-2} + \mathbf{I}\sigma_u^{-2} \right)^{-1}$ cuyos elementos diagonales corresponden a $\text{Var}(\hat{g}_i)$, el segundo término puede simplificarse por medio de la identidad de Woodbury. Esta fórmula establece que, para matrices \mathbf{A} , \mathbf{U} y \mathbf{C} conformables, la inversa de la suma de matrices es igual a:

$$(\mathbf{A} + \mathbf{UCU}')^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{U}'\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{U}'\mathbf{A}^{-1} \quad [3.12]$$

A los efectos de emplear la identidad [3.12] en la simplificación de la expresión para el cálculo de la $\text{Var}(\hat{\mathbf{g}})$, sea $\mathbf{A} = \mathbf{I} \sigma_u^{-2}$, $\mathbf{U} = \mathbf{Z}'$, $\mathbf{U}' = \mathbf{Z}$ y $\mathbf{C} = \mathbf{I} \sigma_e^2$. Entonces, empleando [3.12] la varianza de los efectos de SNP es igual a:

$$\left(\mathbf{Z} \mathbf{Z}' \sigma_e^{-2} + \mathbf{I} \sigma_u^{-2} \right)^{-1} = \mathbf{I} \sigma_u^2 - \sigma_u^2 \mathbf{Z}' \left(\mathbf{I} \sigma_e^2 + \mathbf{Z} \mathbf{Z}' \sigma_u^2 \right)^{-1} \mathbf{Z} \sigma_u^2 \quad [3.13]$$

Nótese que $\mathbf{V} = \mathbf{Z} \mathbf{Z}' \sigma_u^2 + \mathbf{I} \sigma_e^2$ es el término central en la derecha de la expresión [3.13], tal que al reemplazar se obtiene:

$$\left(\mathbf{Z} \mathbf{Z}' \sigma_e^{-2} + \mathbf{I} \sigma_u^{-2} \right)^{-1} = \mathbf{I} \sigma_u^2 - \left(\sigma_u^2 \right)^2 \mathbf{Z}' \mathbf{V}^{-1} \mathbf{Z} \quad [3.14]$$

De este modo, la matriz de covarianzas de los efectos de SNP estimados bajo el modelo de marcador centrado presentado en la ecuación [3.6] es igual a:

$$\begin{aligned} \mathbf{I} \sigma_u^2 - \left(\mathbf{Z} \mathbf{Z}' \sigma_e^{-2} + \mathbf{I} \sigma_u^{-2} \right)^{-1} &= \mathbf{I} \sigma_u^2 - \mathbf{I} \sigma_u^2 + \left(\sigma_u^2 \right)^2 \mathbf{Z}' \mathbf{V}^{-1} \mathbf{Z} \\ &= \left(\sigma_u^2 \right)^2 \mathbf{Z}' \mathbf{V}^{-1} \mathbf{Z} \end{aligned} \quad [3.15]$$

Asimismo, nótese que el elemento diagonal en [3.15] es igual a $\left(\sigma_u^2 \right)^2 z_i' \mathbf{V}^{-1} z_i$, a su vez igual a $\frac{\left(\sigma_u^2 \right)^2}{\text{Var}(\hat{b}_i)}$, y por lo tanto:

$$\text{Var}(\hat{g}_i) = \frac{\left(\sigma_u^2 \right)^2}{\text{Var}(\hat{b}_i)} \quad [3.16]$$

verificando lo que se quería demostrar.

b. *Demostración de* $\hat{g}_i = \frac{\sigma_u^2}{\text{Var}(\hat{b}_i)} \hat{b}_i$

Para demostrar esta proposición trabajaremos el término de la derecha en la igualdad de manera tal que:

$$\begin{aligned}
\frac{\sigma_u^2}{\text{Var}(\hat{b}_i)} \hat{b}_i &= \sigma_u^2 \frac{z_i' \mathbf{V}^{-1} \mathbf{y}}{z_i' \mathbf{V}^{-1} z_i} = \sigma_u^2 z_i' \mathbf{V}^{-1} \mathbf{y} \\
&= \sigma_u^2 z_i' (\mathbf{Z} \mathbf{Z}' \sigma_u^2 + \mathbf{I} \sigma_e^2) \mathbf{y} \\
&= z_i' \left(\mathbf{Z} \mathbf{Z}' (\sigma_u^2)^2 + \mathbf{I} \left(\frac{\sigma_e^2}{\sigma_u^2} \right) \right)^{-1} \mathbf{y}
\end{aligned} \tag{3.17}$$

Un modo de simplificar esta expresión es multiplicar por la matriz identidad, e igual al producto de una matriz no-singular y su inversa, previo al primer paréntesis. Recordando que la matriz de relaciones genómicas \mathbf{G} es función de $\mathbf{G} = \mathbf{Z} \mathbf{Z}'$, tenemos entonces que:

$$\begin{aligned}
\frac{\sigma_u^2}{\text{Var}(\hat{b}_i)} \hat{b}_i &= z_i' \mathbf{G}^{-1} \mathbf{G} \left(\mathbf{Z} \mathbf{Z}' + \mathbf{I} \left(\frac{\sigma_e^2}{\sigma_u^2} \right) \right)^{-1} \mathbf{y} \\
&= z_i' \mathbf{G}^{-1} (\mathbf{G}^{-1})^{-1} \left(\mathbf{Z} \mathbf{Z}' + \mathbf{I} \left(\frac{\sigma_e^2}{\sigma_u^2} \right) \right)^{-1} \mathbf{y}
\end{aligned} \tag{3.18}$$

Si se aplica en la expresión [3.18] la regla de la inversa de un producto y se opera en la expresión resultante, se tiene que:

$$\begin{aligned}
\frac{\sigma_u^2}{\text{Var}(\hat{b}_i)} \hat{b}_i &= z_i' \mathbf{G}^{-1} (\mathbf{G}^{-1})^{-1} \left(\left(\mathbf{Z} \mathbf{Z}' + \mathbf{I} \left(\frac{\sigma_e^2}{\sigma_u^2} \right) \right) \mathbf{G}^{-1} \right)^{-1} \mathbf{y} \\
&= z_i' \mathbf{G}^{-1} \left(\left(\mathbf{Z} \mathbf{Z}' \mathbf{G}^{-1} + \mathbf{G}^{-1} \left(\frac{\sigma_e^2}{\sigma_u^2} \right) \right) \mathbf{G}^{-1} \right)^{-1} \mathbf{y} \\
&= z_i' \mathbf{G}^{-1} \left(\mathbf{I} + \mathbf{G}^{-1} \left(\frac{\sigma_e^2}{\sigma_u^2} \right) \right)^{-1} \mathbf{y}
\end{aligned} \tag{3.19}$$

Sin embargo, nótese que:

$$\left(\mathbf{I} + \mathbf{G}^{-1} \left(\frac{\sigma_e^2}{\sigma_u^2} \right) \right)^{-1} \mathbf{y} = \left(\mathbf{Z} \mathbf{Z}' \sigma_e^{-2} + \mathbf{I} \sigma_u^{-2} \right)^{-1} \mathbf{y} \sigma_e^{-2} = \hat{\mathbf{a}} \tag{3.20}$$

Y, por lo tanto:

$$\frac{\sigma_u^2}{\text{Var}(\hat{b}_i)} \hat{b}_i = z_i' \mathbf{G}^{-1} \hat{\mathbf{a}} \tag{3.21}$$

El término de la derecha en la expresión [3.21] permite estimar de forma indirecta los efectos de SNP \hat{g}_i (Strandén y Garrick, 2009; Hayes y Goddard, 2010; Wang *et al.*, 2012; Wang *et al.*, 2014; Gualdrón Duarte *et al.*, 2014) tal como se quería demostrar. Finalmente, la equivalencia entre tg_i y tb_i sigue a partir de la demostración de las dos proposiciones, es decir:

$$tg_i = \frac{\hat{g}_i}{\sqrt{\text{Var}(\hat{g}_i)}} = \frac{\frac{\sigma_u^2}{\text{Var}(\hat{b}_i)} \hat{b}_i}{\sqrt{\frac{(\sigma_u^2)^2}{\text{Var}(\hat{b}_i)}}} = \frac{\hat{b}_i}{\sqrt{\text{Var}(\hat{b}_i)}} = tb_i \quad [3.22]$$

Por lo tanto, tg_i es una solución razonable en términos computacionales al modelo de efectos fijos de SNP como EMMAX. Esta demostración es un resultado importante a los propósitos de esta tesis, dado que la varianza para el modelo de efectos fijos de SNP puede expresarse como:

$$\text{Var}(\hat{b}_i) = \frac{(\sigma_a^2)^2}{\text{Var}(\hat{g}_i)} \quad [3.23]$$

3.3.5. Meta-análisis de estudios de asociación genómica (MA-GWAS).

Estadísticamente, el meta-análisis consiste en la estimación del promedio de los efectos obtenidos a partir de estudios independientes, ponderando dichos efectos de manera apropiada (Evangelou y Ioannidis, 2013). Uno de los enfoques más comunes para calcular las ponderaciones es el estimador de Hedges y Vevea (Hedges y Vevea, 1998), en las que son función de la inversa de la varianza estimada de cada efecto. Según Hedges (1983) y Hedges y Olkin (1985), el enfoque de Hedges y Vevea (1998) es la metodología óptima para ponderar un conjunto de q efectos independientes, cuando la estimación se realiza empleando modelos donde los efectos son considerados fijos.

Alternativamente, es posible ponderar los efectos estimados por medio del estadístico de Hunter y Schmidt (1990). En este caso, se ponderan los efectos considerando el tamaño muestral de cada conjunto de datos como una aproximación a la ponderación óptima previamente definida. Para implementar un MA-GWAS, se combinaron entonces en un estadístico único (en este caso estadísticos z) los resultados calculados a partir de k estudios de asociación, empleando las dos alternativas de ponderación antes mencionadas.

3.3.5.1. Cálculo de estadísticos z (z -scores).

Independientemente del esquema de ponderación utilizado, los efectos de SNP estimados a partir de [2.4] son estandarizados, mediante la división de los efectos por la raíz cuadrada de su respectiva varianza, la cual es obtenida empleando [2.5]. En consecuencia, se tiene que:

$$z_{ij} = \frac{\hat{g}_{ij}}{\sqrt{\text{Var}(\hat{g}_{ij})}} \quad [3.24]$$

donde z_{ij} es el estadístico z correspondiente al j -ésimo SNP en la i -ésima población. Una ventaja del estadístico z en [3.24] es que permite considerar el signo del efecto, además de facilitar la incorporación de las ponderaciones (Evangelou e Ioannidis, 2013).

3.3.5.2. Esquemas de ponderación.

3.3.5.2.1. Ponderación empleando la inversa de la varianza de los efectos estimados de SNP.

Como se mencionó, el criterio óptimo de ponderación para combinar pruebas estadísticas obtenidas a partir de un modelo de efectos fijos es función de $\text{Var}(\hat{b}_i)$. Asimismo, se mostró en el apartado 3.3.4 que el test estadístico obtenido a partir de un modelo asumiendo efectos de SNP como aleatorios (tg_i) es equivalente a la prueba de efectos fijos (tb_i) pero con numerador y denominador diferentes. Con lo cual, para ponderar de modo óptimo los efectos de SNP estimados, es necesario calcular la varianza de los efectos fijos estimados a partir de la varianza de los efectos aleatorios en nuestro modelo, empleando la expresión [3.23]. De este modo, la ponderación para el criterio de inversa de la varianza es calculada del modo siguiente:

$$w_{ij} = \frac{1}{\text{Var}(\hat{b}_{ij})} \quad [3.25]$$

Willer *et al.* (2010) observaron que el enfoque que emplea la inversa de la varianza requiere que los efectos estimados y sus varianzas se encuentren expresados en las mismas unidades en todas las poblaciones analizadas. Por lo tanto, cuando el carácter es medido en escalas diferentes en las distintas poblaciones, es necesario transformar todas las mediciones a una escala común.

3.3.5.2.2. Ponderación por tamaño muestral.

Como se mencionó previamente, el tamaño muestral de la i -ésima población (N_i) constituye un esquema de ponderación alternativo para el efecto estimado del SNP j en

la población i (\hat{g}_{ij}), es decir, $w_{ij} = N_i$. En este caso, la población con el mayor número de registros tendrá a su vez mayor influencia en el MA-GWAS. Nótese que este esquema de ponderación es independiente de la escala de medición y por tanto, puede ser usado en el desarrollo de MA involucrando fenotipos expresados en unidades diferentes. Asimismo, nótese que bajo homogeneidad de varianzas a través de poblaciones, los dos esquemas de ponderación descritos deben generar resultados similares.

3.3.5.3. Combinación de estadísticos z .

Una vez calculados los estadísticos z según [3.24], deben ser ponderarlos según el esquema seleccionado combinándolos en un estadístico z único para todas las poblaciones. Entonces, el estimador del efecto del j -ésimo SNP ponderado es igual a:

$$z_j^* = \sum_{i=1}^k \left[z_{ij} \left(\frac{\sqrt{w_{ij}}}{\sqrt{\sum_{i=1}^k w_{ij}}} \right) \right] \quad [3.26]$$

donde z_{ij} es el estadístico z para el j -ésimo SNP dentro de la i -ésima población, calculado mediante [3.24]. El escalar w_{ij} es la inversa de la varianza y consecuentemente una cantidad no negativa, o ponderación correspondiente al tamaño muestral. El test estadístico resultante es empleado luego en el cálculo de p -valores mediante [2.6], de modo de probar asociaciones significativas consistentes sobre todas las poblaciones.

3.3.6. Comparación con análisis conjunto de asociación.

Para comparar los resultados del MA-GWAS con aquellos obtenidos a partir de un análisis conjunto de asociación (JA), se consideraron los modelos descritos e implementados en el capítulo anterior, específicamente el enfoque de JA basado en la combinación de todas las poblaciones en una base de datos única, asumiendo por lo tanto, homogeneidad de varianza aditiva y residual entre poblaciones (sección 2.4.6.4). Esto se debió a que el objetivo del presente capítulo es identificar aquellos SNP con efectos consistentes entre todas las poblaciones, en lugar de considerar la interacción SNP \times población.

3.4. Resultados.

3.4.1. Desequilibrio de ligamiento y persistencia de la fase a través de poblaciones.

Como medida de relación entre las tres poblaciones de cerdos, la Figura 3.4.1.1 presenta las estimaciones de LD en términos de la caída del valor del coeficiente de

correlación r^2 con respecto a la distancia entre marcadores, así como la persistencia de la fase. Los gráficos se construyeron a partir de los genotipos pertenecientes a los cromosomas 7 y 14, utilizando los códigos disponibles en la interfaz LDestV1.0 (https://www.msu.edu/~steibelj/JP_files/LD_estimate.html) y desarrollados por Badke *et al.* (2012). Asimismo, la Tabla 3.4.1.1 muestra el promedio de r^2 para marcadores adyacentes, ubicados a distancias de 0.5Mb, 1Mb y 5 Mb.

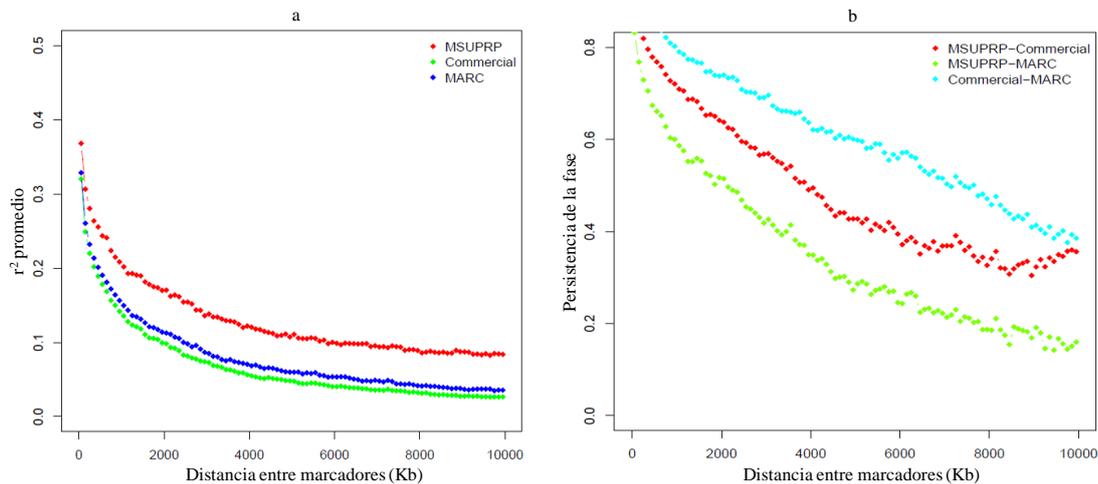


Figura 3.4.1.1. Estimación de LD y persistencia de la fase entre poblaciones. a. Estimación de LD por distancia de SNP adyacentes expresada en Kilobases; b. Persistencia de la fase entre poblaciones.

Tabla 3.4.1.1. Estimaciones promedio de LD en términos de correlación r^2 para distancias de 0.5Mb, 1Mb, 5Mb y 10Mb entre marcadores en las tres poblaciones de cerdos evaluadas.

Población	Distancia entre marcadores			
	0.5Mb	1Mb	5Mb	10Mb
Comercial	0.1838	0.1389	0.0480	0.0266
MARC	0.1960	0.1532	0.0599	0.0357
MSUPRP	0.2497	0.2053	0.1092	0.0839

En primer lugar, las estimaciones poblacionales de la caída en el valor de r^2 en relación a la distancia entre marcadores adyacentes (Figura 3.4.1.1.a y Tabla 3.4.1.1) indican que pares de marcadores disponibles en la población MSUPRP tuvieron mayor r^2 que el observado en las otras dos poblaciones. Específicamente, para SNP ubicados a 1Mb de distancia, el valor de r^2 en la población MSUPRP (0.2053) fue aproximadamente, 1.5 veces mayor que el observado en las otras dos poblaciones (0.1532 y 0.1389 en MARC y Comercial respectivamente). Esta diferencia es acentuada a mayores distancias entre pares de SNP, donde para marcadores ubicados a 10Mb, el valor de r^2 para la población MSUPRP fue entre 2.3 y 3 veces superior al estimado para las poblaciones MARC y Comercial, respectivamente.

Siguiendo el enfoque descrito por Badke *et al.* (2012), se calculó la persistencia de la fase para pares de SNP agrupados en categorías de distancias entre 0Mb-10Mb entre pares de poblaciones (Figura 3.4.1.1.b). Los valores para distancias de 0Mb a 0.1Mb, 0.9Mb a 1Mb, 4.9Mb a 5Mb y de 9.9Mb a 10Mb se incluyen en la Tabla 3.4.1.2.

Tabla 3.4.1.2. Comparación entre pares de poblaciones de la correlación de la fase para varias distancias entre pares de marcadores.

Distancia	Poblaciones comparadas		
	MSUPRP-Comercial	MSUPRP-MARC	Comercial-MARC
0_0.1Mb	0.8904	0.8327	0.9267
0.9_1Mb	0.7212	0.6006	0.8031
4.9_5Mb	0.4280	0.2732	0.6013
9.9-10Mb	0.3565	0.1599	0.3857

De acuerdo con la Figura 3.4.1.1.b, la persistencia de la fase disminuyó cuando aumentó la distancia entre marcadores, patrón similar al observado para la caída de LD dentro de población. Además, la disminución en la persistencia con aumentos en la distancia entre SNP fue más drástica para la comparación entre las poblaciones MSUPRP y MARC en comparación con la observada para la persistencia entre las poblaciones Comercial y MARC. Específicamente, la correlación de la fase entre SNP ubicados a una distancia de 1Mb osciló entre 0.60 para la comparación entre MARC con MSUPRP y 0.80 para la población Comercial con MARC. Para una distancia mayor, como el caso de SNP con una distancia promedio entre 9.9Mb y 10Mb, la correlación de la fase disminuyó a 0.16 para MARC con MSUPRP y a 0.39 para Comercial con la población MARC. Es decir, la caída en la correlación fue mayor en el caso de la comparación MSUPRP-MARC en comparación con la observada para las poblaciones Comercial con MARC. Estos resultados sugieren una relación más cercana entre las poblaciones Comercial y MARC que la existente entre las otras posibles comparaciones.

3.4.2. Variación fenotípica, componentes de varianza y análisis de asociación a nivel poblacional.

Dado que, al igual que el capítulo anterior, el carácter de calidad de carne considerado para ejemplificar el desarrollo de MA-GWAS corresponde a color rojo del músculo *longissimus dorsi* o CIE a*, los estadísticos descriptivos así como los parámetros genéticos y heredabilidades para todas las poblaciones se encuentran detallados en las Tablas 2.4.2.1 y 2.4.3.1. Los componentes de varianza, así como los valores de cría para los animales genotipados se estimaron empleando el modelo [3.3], dando lugar a estimaciones de heredabilidad de 0.160, 0.449 y 0.603 para las poblaciones MARC, Comercial y MSUPRP, respectivamente.

Asimismo, en las Figuras 2.4.5.1a, b y c se presentaron los gráficos de Manhattan para visualizar asociaciones significativas en CIE a* de las poblaciones Comercial, MARC y MSUPRP, respectivamente. Como se mencionó en dicha sección, se observaron algunos picos para CIE a* en los cromosomas 6 en todas las poblaciones y en el cromosoma 12 para MSUPRP. No obstante, ninguna de estas regiones alcanzó el umbral de significancia definido por la corrección de Bonferroni para inferencia múltiple (1.094×10^{-6} , 1.1359×10^{-6} y 1.2325×10^{-6} en las poblaciones Comercial, MARC y MSUPRP, respectivamente).

3.4.3. Meta-análisis de estudios de asociación genómica (MA-GWAS).

A partir de la combinación de las pruebas estadísticas poblacionales para el carácter CIE a*, la Figura 3.4.3.1 presenta los gráficos Manhattan para MA-GWAS, empleando ponderaciones que son función de la inversa de la varianza y del tamaño muestral.

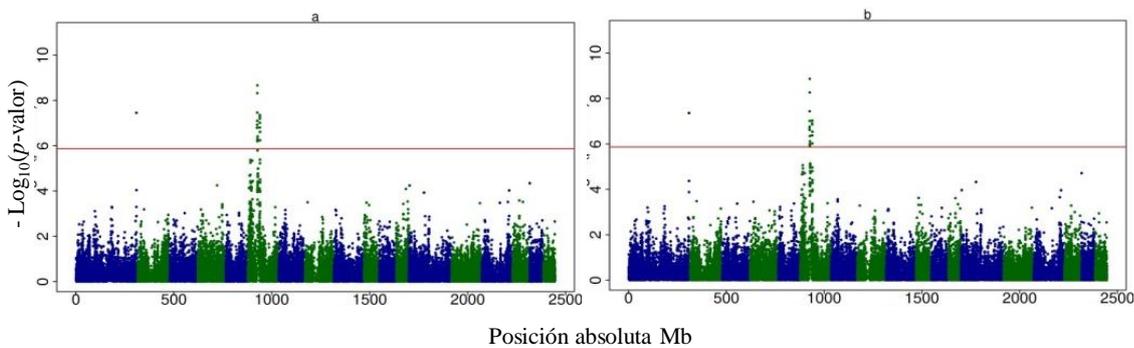


Figura 3.4.3.1. Gráficos Manhattan para CIE a* en MA-GWAS. Gráficos Manhattan para asociaciones de SNP con CIE a* considerando: a. MA ponderado por inversa de la varianza; b. MA ponderado por tamaño muestral. En los dos gráficos, $-\text{Log}_{10}(p\text{-valor})$ en eje y vs. posición absoluta del SNP en Megabases (eje x). La línea horizontal define el umbral de significancia $p < 0.05$ a nivel del genoma completo.

Adicionalmente, la Tabla 3.4.3.1 describe información detallada de los SNP asociados con CIE a* bajo los dos enfoques de ponderación en el MA-GWAS.

Tabla 3.4.3.1. Asociaciones significativas con CIE a* bajo meta-análisis. Asociaciones significativas obtenidas bajo meta-análisis de GWAS considerando ponderaciones empleando la inversa de la varianza de los efectos estimados de SNP y el tamaño muestral. Los picos de asociación se resaltan en negrita.

SNP_ID		Crm ³	Posición ⁴	Efecto ⁵		p-valor	
MA_Ivar ¹	MA_N ²			MA_Ivar	MA_N	MA_Ivar	MA_N
ALGA0103022	ALGA0103022	1	308984948	5.5136	5.4760	3.5165E-08	4.3493E-08
H3GA0053864	H3GA0053864	6	48585961	-5.2387	-5.1740	1.6175E-07	2.2917E-07
ALGA0102689	ALGA0102689	6	48717238	-5.2962	-5.2227	1.1821E-07	1.7632E-07
ASGA0104037	ASGA0104037	6	48792292	-5.2434	-5.1660	1.5762E-07	2.3910E-07
ASGA0089838	ASGA0089838	6	49146524	5.2434	5.1651	1.5766E-07	2.4027E-07
ASGA0093393	ASGA0093393	6	49168322	5.5173	5.5082	3.4421E-08	3.6258E-08
MARC0019764	MARC0019764	6	49351202	5.3696	5.3313	7.8902E-08	9.7520E-08
MARC0015928	MARC0015928	6	49538608	-5.8556	-5.8340	4.7530E-09	5.4098E-09
DIAS0000492	DIAS0000492	6	49802217	-5.9866	-6.0610	2.1433E-09	1.3527E-09
MARC0011519	MARC0011519	6	49817264	-5.0028	-4.8676	5.6509E-07	1.1296E-06
H3GA0056470	H3GA0056470	6	50006716	5.0260	4.8875	5.0088E-07	1.0214E-06
DIAS0004447	DIAS0004447	6	50037571	-5.0007	-4.8651	5.7132E-07	1.1439E-06
DIAS0003231	DIAS0003231	6	50065951	-4.9809	-4.8420	6.3291E-07	1.2854E-06
----	H3GA0017949	6	50532885	---	-5.0488	----	4.4449E-07
ASGA0103898	ASGA0103898	6	50867656	5.0747	4.9429	3.8821E-07	7.6981E-07
ASGA0095526	ASGA0095526	6	62229667	-5.3335	-5.3321	9.6337E-08	9.7089E-08
ASGA0089485	ASGA0089485	6	62369260	5.4454	5.2694	5.1690E-08	1.3685E-07
ASGA0105081	ASGA0105081	6	62384457	-5.4091	-5.2788	6.3352E-08	1.3003E-07
H3GA0052873	H3GA0052873	6	62399306	-5.4745	-5.3388	4.3869E-08	9.3572E-08
ASGA0095271	ASGA0095271	6	62584617	-5.2449	-5.0592	1.5638E-07	4.2106E-07
ALGA0106983	ALGA0106983	6	63059577	4.9996	4.9017	5.7450E-07	9.5022E-07
ALGA0124037	ALGA0124037	6	63146455	5.1614	5.1309	2.4509E-07	2.8830E-07

¹ Meta-análisis considerando ponderación por inversa de la varianza.

² Meta-análisis considerando ponderación por tamaño muestral.

³ Cromosoma (Sus scrofa genome build 10.2).

⁴ Posición del SNP en pares de bases.

⁵ Efecto fenotípico estimado por una copia del alelo B.

En primer lugar, los resultados de MA-GWAS obtenidos a partir de los dos esquemas de ponderación fueron bastante similares. Sin embargo, el MA bajo ponderación por tamaño muestral detectó un SNP adicional en el cromosoma 6 (H3GA0017949, p -valor $< 4.4449 \times 10^{-7}$) que no fue significativo bajo el enfoque de inversa de la varianza. A pesar de dicha diferencia, tanto la ponderación por inversa de la varianza como por tamaño muestral identificaron un QTL significativo en el cromosoma 1 ubicado a 308.9Mb (ALGA0103022, p -valores $< 4.349 \times 10^{-8}$ y 3.5164×10^{-8} para la inversa de la varianza y el tamaño muestral, respectivamente) y además, una región significativa en el cromosoma 6 ubicada entre 48.5Mb y 63.1Mb. En esta región, se detectó un pico de asociación a 49.8Mb siguiendo los dos esquemas de ponderación (DIAS0000492, p -valores $< 1.353 \times 10^{-9}$ y 2.1432×10^{-9} para la inversa de la varianza y el tamaño muestral, respectivamente). Comparando los p -valores

resultantes de ambos enfoques MA, la Figura 3.4.3.2 revela que los p -valores obtenidos a partir de la inversa de la varianza se encontraron altamente correlacionados con aquellos calculados ponderando por el tamaño muestral ($R^2 = 0.964$), principalmente en el caso de p -valores bajos ($-\log_{10}(p\text{-valor}) \geq 6$).

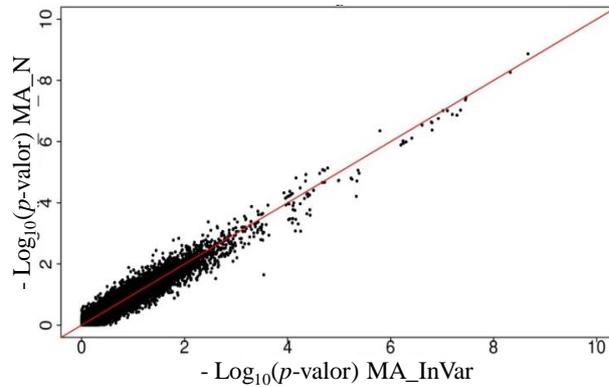


Figura 3.4.3.2. Comparación de p -valores obtenidos bajo MA. Gráficos Q-Q para comparación de p -valores obtenidos a partir de MA ponderado por inversa de la varianza (eje x) vs. MA ponderado por tamaño muestral (eje y).

3.4.4. Comparación meta-análisis de asociación y análisis de asociación conjunto.

Como se mencionara en la sección Métodos (sección 3.3.6), para comparar los resultados obtenidos a partir de MA-GWAS, se ajustó un JA que consideró homogeneidad de varianzas aditiva y residual y que no incorporó componentes principales como enfoque para corregir por estructura genética de la población. Esto último, teniendo en cuenta que no se observaron diferencias entre las estimaciones obtenidas a partir del JA incluyendo o no componentes principales (Figura 2.4.4.2.d). De este modo, se compararon los p -valores obtenidos a partir del JA con aquellos calculados a partir del MA-GWAS (Figura 3.4.4.1).

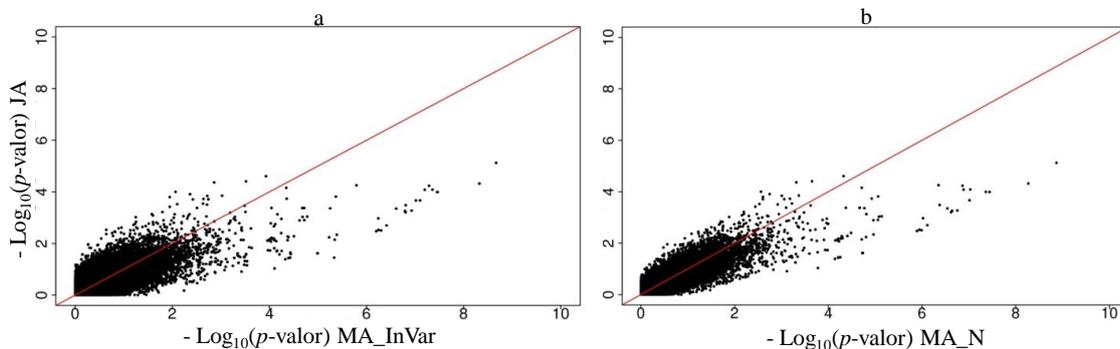


Figura 3.4.4.1. Comparación de p -valores obtenidos bajo el meta-análisis (MA) y el análisis conjunto de asociación (JA). Gráficos Q-Q para la comparación de p -valores

obtenidos a partir de: a. MA ponderado por la inversa de la varianza (eje x) vs. JA (eje y); b. MA ponderado por el amaño muestral (eje x) vs. JA (eje y).

De acuerdo a la Figura 3.4.4.1, el JA condujo a p -valores mayores que los dos enfoques de MA-GWA, con más dispersión alrededor de la línea 1-1 cuando la inversa de la varianza fue utilizada como esquema de ponderación ($R^2 = 0.726$), en comparación con el uso de ponderaciones calculadas con el tamaño muestral ($R^2 = 0.831$).

3.5. Discusión

En el presente capítulo se mostró como desarrollar un meta-análisis (MA) a partir de evaluaciones genómicas distintas e independientes, y se compararon los resultados obtenidos con aquellos que surgieron de un análisis conjunto de asociación (JA), enfoque descrito en el capítulo anterior. Al igual que en dicho capítulo, la metodología de MA fue ilustrada utilizando CIE a* debido a la importancia económica de los caracteres de color en la industria porcina (Cannon *et al.*, 1996; Ovilo *et al.*, 2002). En adición, como se mencionara en el capítulo precedente, la heterogeneidad en los componentes de varianza observada para dicho carácter puede ser representativa de futuros estudios de MA.

En primer lugar, considerando que la potencia de los análisis de asociación está relacionada con la existencia de desequilibrio gamético o LD entre los SNP y los QTL (Meuwissen *et al.*, 2001; Hayes *et al.*, 2009), se estimó la caída de LD en términos de r^2 así como la persistencia de la fase en las poblaciones evaluadas. Se observaron valores de r^2 promedio de 0.14 a 0.21 y de 0.05 a 0.10 para distancias entre SNP de 1Mb y 5Mb, respectivamente (Tabla 3.4.1.1). Estos valores fueron similares a los reportados por Badke *et al.* (2012) para diferentes razas de cerdos en las mismas distancias entre SNP adyacentes. Asimismo, los valores de r^2 para la población MSUPRP en distancias genómicas grandes (cercasas a 10Mb entre SNP) fueron mayores a los observados en las poblaciones Comercial y MARC. De acuerdo con Hayes *et al.* (2003), la diferencia en los valores de r^2 para distancias entre SNP extensas es una medida del tamaño efectivo reciente de la población. De este modo, los resultados observados en este capítulo sugieren un menor tamaño efectivo de la población MSUPRP en comparación con las poblaciones Comercial y MARC.

En el caso de la persistencia de la fase de LD, la cual para distancias genómicas grandes puede reflejar migración en la historia reciente de la población (de Roos *et al.*, 2008), se observó una disminución de dicha correlación cuando aumentó la distancia genómica. Para las diferentes comparaciones entre poblaciones, la persistencia de la fase entre las poblaciones Comercial y MARC fue mayor a la observada entre MSUPRP con las otras dos poblaciones (Tabla 3.4.1.2.). Para SNP con distancia entre pares de SNP de 100Kb, la persistencia de la fase osciló entre 0.83 (MSUPRP-MARC) y 0.92 (Comercial-MARC). Estos valores fueron superiores a los reportados por Badke *et al.* (2012) para pares de SNP a la misma distancia genómica, autores que observaron valores de persistencia de fase con rango de 0.67 a 0.78 entre diferentes razas de cerdos. Estos resultados son relevantes dado que el nivel de LD entre los marcadores y los

QTL, así como la correlación de la fase de LD para diferentes poblaciones, afectarán los resultados obtenidos a partir de estudios de asociación y de la selección genómica (Meuwissen *et al.*, 2001; Badke *et al.*, 2012). Específicamente, la correlación de la fase es un indicador de la exactitud con la cual se pueden predecir asociaciones entre los SNP y los QTL en una población usando asociaciones SNP-QTL de otra población (de Roos *et al.*, 2008). Además, en el contexto de esta tesis, si el objetivo recae en el aumento de la potencia en la detección de asociaciones significativas a partir de GWAS y si los animales de diferentes poblaciones están relacionados en términos de correlación de la fase gamética, es posible combinar individuos de distintas poblaciones para estimar los efectos de SNP a partir de la predicción de GEBV de ambas poblaciones (Goddard *et al.*, 2006).

Una de las contribuciones principales de la investigación aquí presentada es el cálculo apropiado de las ponderaciones para el MA a partir de modelos animales centrados GBLUP. Tanto en el presente capítulo (apartado 3.3.4) como en trabajos previos (Gualdrón Duarte *et al.*, 2014) se demostró la equivalencia numérica y analítica entre las pruebas estadísticas propuestas para los efectos de SNP, desarrolladas a partir de los modelos para obtener GBLUP, y una prueba obtenida a partir de un modelo que considera los efectos de SNP como fijos con un efecto poligénico aleatorio: EMMA (Kang *et al.*, 2010; Zhang *et al.*, 2010). Esta prueba es esencial para obtener las ponderaciones correctas para el enfoque de la inversa de la varianza propuesto en este capítulo, las cuales son obtenidos fácilmente a partir de la varianza de los efectos de SNP aleatorios, tal como se muestra en la ecuación [3.23].

Al igual que otros ejemplos de MA (Stankowich y Blumstein, 2005; Akanno *et al.*, 2013), en la presente investigación se consideraron ponderaciones empleando el tamaño muestral. Estas son sub-óptimas con respecto a aquellas que emplean la inversa de la varianza, dado que no incorporan exactitudes de imputación (Jiao *et al.*, 2011) y desconsideran heterocedasticidad específica dentro de población. La heterogeneidad de varianza es tenida en cuenta por las ponderaciones que son función de la inversa de la varianza (Hedges y Olkin, 1985; Marin-Martinez y Sanchez-Meca, 2010; Akanno *et al.*, 2013). Sin embargo, las ponderaciones empleando el tamaño muestral son más fáciles de implementar cuando la variable es medida en escalas distintas en las diferentes poblaciones (Willer *et al.*, 2010).

En el caso particular de los conjuntos de datos analizados, se observó una mayor potencia del MA en comparación con los análisis de asociación a nivel poblacional, resultado que es evidente considerando la detección de asociaciones significativas a nivel del genoma en el cromosoma 1 (p -valor $< 4.349 \times 10^{-8}$) y en el cromosoma 6 (p -valor $< 2.1433 \times 10^{-9}$), considerando el criterio de Bonferroni para definir el umbral de significancia a nivel del genoma. Para dicho carácter de color, la asociación observada en el cromosoma 1 (308.9Mb) estuvo alejada de la región reportada por Sanchez *et al.* (2014) ubicada entre las posiciones 8.7Mb y 12.8Mb para el mismo carácter. Además, los picos de asociación en el cromosoma 6 estuvieron ubicados a 0.7Mb del QTL reportado por Cherel *et al.* (2011) para CIE a*. Estos resultados muestran que el MA tiene potencial para detectar un mayor número de efectos SNP significativos que los GWAS poblacionales, sin requerir disponer de los datos originales y considerando

heterogeneidad de varianza, cuando el modelo para GWAS ha sido especificado apropiadamente. Además, se observaron semejanzas cuantitativas y cualitativas entre los resultados de los dos esquemas de ponderación para el MA, lo que se vio reflejado en una correlación entre $-\log_{10}(p\text{-valores})$ calculados a partir de ambos esquemas de ponderación ($R^2 = 0.964$; Figura 3.4.3.2). Todos los picos de asociación significativos resultantes del MA basado en la inversa de la varianza fueron observados también bajo el enfoque de tamaño muestral, el cual detectó un SNP adicional en el cromosoma 6 (H3GA0017949, $p\text{-valor} < 4.4449 \times 10^{-7}$). Con lo cual, la potencia de detección de ambos enfoques de ponderación fue prácticamente la misma. Este resultado concuerda con los de Akanno *et al.* (2013), quienes compararon ponderar por la inversa de la varianza y por el tamaño muestral, en un MA para caracteres de producción y reproducción en cerdos, obteniendo heredabilidades similares para ambos enfoques de ponderación.

Además de la definición apropiada de las ponderaciones en el MA, otro aspecto importante de este capítulo fue la definición de los modelos para los estudios de asociación dentro de cada población. La importancia radica en que los GWAS específicos dentro de cada población constituyen el primer paso en el desarrollo del MA, requiriendo considerar toda la variación sistemática dentro de población. Asimismo, se siguió el enfoque propuesto por Lans *et al.*, (2012) para incluir como covariables los autovectores obtenidos a partir de la matriz G . Sin embargo, como se mencionó en el capítulo anterior, no se encontraron diferencias entre los p -valores para los efectos de SNP estimados a partir de los modelos incluyendo ([2.7]) y excluyendo ([2.1]) los autovectores de la matriz G como covariables fijas (Figura 2.4.4.2). Con lo cual, se eliminaron los autovectores de G y se definió el uso del modelo [2.1] para todos los análisis.

Finalmente, se compararon los resultados de un MA con su alternativa natural: el análisis conjunto de asociación (JA). Dado que el objetivo de este capítulo es encontrar aquellos SNP con efectos consistentes entre todas las poblaciones (ignorando la interacción $\text{SNP} \times \text{población}$), en este capítulo se consideró el enfoque JA homocedástico donde todos los datos son combinados y analizados conjuntamente (Bravata y Olkin, 2001). En este capítulo se ajustaron JA con varianzas homogéneas entre poblaciones y no se lograron identificar SNP asociados significativamente con CIE a*. Aun cuando los p -valores calculados a partir del JA y del MA estuvieron altamente correlacionados ($R^2 = 0.8308$ y 0.7264 para MA basado en el tamaño muestral y en la inversa de la varianza, respectivamente), el JA produjo pruebas de hipótesis menos conservadoras que el MA, especialmente con aquellas extremas. Es decir, que JA produjo p -valores mayores que los obtenidos con MA (Figuras 3.4.4.1). De acuerdo con Zhou *et al.* (2011), las diferencias observadas están relacionadas con la habilidad del MA para considerar la heterogeneidad del tamaño de los efectos de SNP entre poblaciones, modelando apropiadamente las fuentes de variación importantes. No obstante, la concordancia entre JA y MA realizadas sobre la base del tamaño muestral ($R^2 = 0.8308$), comparada con la diferencia entre JA y MA empleando la inversa de la varianza ($R^2 = 0.7264$) es esperable, dado que al combinar los datos en un modelo único, el JA tiende a favorecer la población con mayor tamaño muestral, situación similar al MA ponderando por tamaño muestral (Walling *et al.*, 2000; Kim *et al.*, 2005).

Por lo tanto, los resultados de este capítulo proveen soporte y evidencia de las ventajas al emplear el MA para combinar evaluaciones genómicas en un solo estudio de asociación, permitiendo tener en cuenta la heterogeneidad genética existente entre poblaciones. El MA actúa entonces como alternativa al JA en el cual se modeló la varianza específica dentro de población y que produjo efectos de SNP particulares para cada población, requiriéndose luego la estimación de un estadístico único para cada SNP con respecto a todas las poblaciones.

Capítulo 4

Estudios de asociación y meta-análisis para caracteres de calidad de carne en cerdos.

Estudios de asociación y meta-análisis para caracteres de calidad de carne en cerdos.

4.1. Introducción

La calidad de carne de cerdo es un aspecto de importancia económica central en todos los segmentos de la producción e industria cárnica porcinas (Cassens, 2000; Rosenvold y Andersen, 2003). En este contexto, la selección genómica (Meuwissen *et al.*, 2001) así como los métodos de regresión sobre el genoma completo tales como GBLUP, permiten predecir valores de cría genómicos (GEBV) para caracteres de calidad de carne de cerdo, de carcasa y de composición (Wang *et al.*, 2012; Akanno *et al.*, 2014; Badke *et al.*, 2014). Conceptualmente, la metodología GBLUP asume que los efectos de SNP son independientes e idénticamente distribuidos, asignando la misma varianza a todos los marcadores (VanRaden, 2008; Habier *et al.*, 2010). En términos computacionales, los modelos GBLUP basados en algoritmos REML y BLUP se constituyen en la generalización del modelo animal, al que se agrega la información de marcadores, son sencillos de implementar y menos demandantes que otras metodologías de predicción genómica (Guo *et al.*, 2014).

Como se mencionó en el capítulo anterior, el desarrollo de evaluaciones genómicas facilita la implementación de estudios de asociación genómica o GWAS (Ma *et al.*, 2013; Rousseau *et al.*, 2013; Jung *et al.*, 2014; Stratz *et al.*, 2014). Los GWAS aprovechan la disponibilidad de un gran número de marcadores SNP, que se encuentran en desequilibrio gamético o de ligamiento con regiones o QTL y que albergan, posiblemente, genes candidatos para los caracteres complejos (Do *et al.*, 2013). Sin embargo, el poder de los GWAS es limitado debido a los tamaños de muestra pequeños que se suelen emplear para los análisis. Esto, dado el costo que implica aumentar el número de individuos evaluados principalmente en poblaciones comerciales y también, a causa de las complejas estructuras genéticas de población que se utilizan dentro de la mejora genética animal (Muñoz *et al.*, 2009; Minozzi *et al.*, 2012). Una alternativa para incrementar la potencia de los estudios de asociación y reducir el número de asociaciones falsamente positivas es desarrollar un meta-análisis de estudios de asociación (MA-GWAS), metodología que permite integrar los resultados obtenidos a partir de estudios de asociación individuales (Evangelou e Ioannidis, 2013). Existen algunos antecedentes de aplicación del MA en mejoramiento animal (Wood *et al.*, 2006; Salmi *et al.*, 2010; Silva *et al.*, 2010; Akanno *et al.*, 2013; Bolormaa *et al.*, 2014). Sin embargo, la implementación de dicha metodología a partir de resultados de GWAS independientes está en franco desarrollo.

Una vez obtenidas las asociaciones significativas entre los marcadores SNP y los QTL, ya sea a partir de GWAS o de MA-GWAS, es de interés identificar genes candidatos vinculados potencialmente con caracteres productivos, de calidad de carne, entre otros. De este modo, es posible describir los procesos biológicos y metabólicos relacionados con los caracteres bajo estudio. Usualmente, la búsqueda de dichos genes se basa en la definición de un intervalo alrededor del pico de asociación, es decir, un

número fijo de Megabases alrededor de dicho QTL (Neves *et al.*, 2011; Do *et al.*, 2013). Otro enfoque consiste en identificar genes candidatos ubicados cerca de los QTL reportados, priorizarlos de acuerdo a sus funciones biológicas y luego, identificar SNP posicionados alrededor de dichos genes usando por ejemplo, secuencias de mRNA (Karisa *et al.*, 2013). Sin embargo, dichos enfoques no consideran el desequilibrio de ligamiento existente entre los SNP incluidos en la región del gen candidato. Así, una alternativa consiste en especificar los límites de dicho intervalo, teniendo en cuenta el bloque de desequilibrio de ligamiento presente alrededor del pico de asociación. Sahana *et al.* (2013) observaron que la definición de un intervalo de 1Mb alrededor del pico de asociación puede capturar la relación entre SNP y genes causales, siempre y cuando los niveles de LD sean altos. Sin embargo, puede ocurrir que el intervalo definido ignore el verdadero alcance del LD en la región. Por lo tanto, el enfoque propuesto en el presente capítulo permite definir de forma más objetiva el intervalo para la búsqueda de genes que afectan posiblemente los caracteres de interés.

Aunque en mejoramiento genético animal se han desarrollado y aplicado diferentes GWAS, permitiendo identificar QTL para caracteres de calidad de carne en cerdo, el trabajo presentado en este capítulo es el primer MA-GWAS desarrollado para identificar variantes genéticas y mecanismos biológicos relacionados con caracteres de calidad de carne, combinando resultados a partir de evaluaciones genómicas basadas en modelos GBLUP. Este enfoque, como se mencionó en el capítulo 3, permite incrementar la potencia al detectar variantes que se manifiestan con un efecto pequeño o limitado en una población pero que, cuando se lo analiza conjuntamente en varias poblaciones, tiene un efecto de mayor importancia. Consecuentemente, el objetivo de este capítulo es implementar un MA-GWAS para caracteres de calidad de carne, combinando resultados de varias evaluaciones GBLUP independientes, siguiendo la metodología presentada en el capítulo 3 de la presente tesis. Asimismo, a partir de los resultados obtenidos en el MA-GWAS, se describe la metodología utilizada en la búsqueda de genes candidatos potenciales, considerando el alcance del LD en las regiones identificadas como asociadas significativamente con los caracteres de calidad de carne de cerdo.

4.2. Materiales.

4.2.1. Poblaciones animales.

Se analizaron fenotipos y genotipos para caracteres de calidad de carne de cerdo en las mismas tres sub-poblaciones descritas en el capítulo 2: (a) Población experimental de cerdos desarrollada en Michigan State University; (b) Población experimental del centro Meat Animal Research Center; (c) Población comercial; (ver apartado 2.2).

4.2.2. Información genotípica y fenotípica.

Además del color rojo de la carne, CIE a*, carácter que fuera utilizado para ejemplificar la metodología descrita en los capítulos 2 y 3, en esta investigación se analizaron registros de varios caracteres de calidad de carne y específicamente, las mediciones de pH 24 horas post-sacrificio (pHu), resistencia al corte (SF), porcentaje de pérdida de fluidos (PRL), porcentaje de grasa intramuscular (IMF), porcentaje de pérdidas por cocción (CKL), y mediciones adicionales de color relacionadas con el musculo *longissimus dorsi* tales como mediciones de luminosidad (CIE L*) y color amarillo (CIE b*). A continuación se da una breve descripción de cada uno de los caracteres evaluados.

En términos generales, el pH es una medida de la acidez del musculo, el cual afecta factores de calidad de carne como el color y la capacidad de retención de líquidos (Bendall, 1973). Específicamente, el *pH 24 horas post-sacrificio (pHu)* constituye un predictor adecuado de las características de calidad de carne comparado con las mediciones de pH en momentos cercanos al sacrificio. Además, la medida es usada en la industria como criterio de clasificación de reses por calidad antes del procesamiento (Boler *et al.*, 2010).

La *resistencia al corte (Shear Force - SF)* es un método de referencia para el análisis de la textura de la carne, dado que valores altos indican una mayor dureza. Este carácter puede ser medido mediante la técnica de *resistencia al corte en una porción (Slice Shear Force - SSF)*, o alternativamente por el método de la cuchilla *Warner-Bratzler (WBS)*. Si bien la metodología WBS tiene gran exactitud cuando los protocolos son ejecutados propiamente, Shackelford *et al.* (1999) reportaron variaciones importantes en la evaluación de resistencia al corte dentro y entre instituciones usando dicho enfoque. Por esta razón, y considerando las altas correlaciones entre WBS y SSF, Shackelford y Wheeler (2009) desarrollaron la técnica SSF, que consiste en una alternativa simplificada y más exacta que WBS como indicador de terneza. Como se describirá en la sección 4.4.3, en el caso de los conjuntos de datos evaluados, se contó con mediciones de resistencia al corte, sea usando la metodología SSF o por medio de WBS.

La calidad de la carne de cerdo se ve afectada también por el *porcentaje de grasa intramuscular (IMF)*. Un porcentaje óptimo de IMF garantiza sabor, terneza y jugosidad adecuados en la carne. Si bien los programas de mejoramiento se han enfocado en obtener carcasas de cerdo más magras, una disminución drástica en IMF conlleva efectos desfavorables en el sabor de la carne. No obstante, cuando IMF se encuentra en valores elevados puede ocasionar problemas de comercialización (Wood *et al.*, 1994; Fernandez *et al.*, 1999; Meisinger, 2002). Por esta razón, este carácter se incluye en los programas de mejoramiento genético en cerdos (Daszkiewicz *et al.*, 2003).

Otro carácter de calidad de gran relevancia económica es el *porcentaje de pérdida de fluidos* en la carne (*Purge loss – PRL*). Su importancia reside en el efecto que tiene sobre los procesos de empaque de la carne, ya que el producto obtenido en la planta de sacrificio se verá reducido drásticamente, afectando el rendimiento en la planta de

procesamiento (Ying *et al.*, 2014). Además, un alto porcentaje de pérdida de líquidos da lugar a una carne seca después de la cocción, disminuyendo la aceptación por parte del consumidor (Meisinger, 2002).

Al igual que PRL, el *porcentaje de pérdidas por cocción* (CKL) es un componente decisivo en los rendimientos de la carne obtenidos después de procesamiento. Este carácter, definido como la diferencia en peso de la carne antes y después de la cocción, expresada como un porcentaje del peso de la carne cruda (Nollet, 2012), constituye una de las variables más utilizadas al momento de estimar la capacidad de retención de líquidos en la carne (Hernandez y Dalle Zotte, 2010). Asimismo, está correlacionado negativamente con el pH y por tanto, es un parámetro indicador de la presencia de carne “Pálida, Suave y Exudativa” (Pale, Soft and Exudative – PSE; Woelfel *et al.*, 2002).

Finalmente, se evaluaron atributos de color de la carne y principalmente, las tres coordenadas de color fundamentales: L* (luminosidad), a* (color rojo) y b* (color amarillo). En el caso de la presente tesis, las mediciones se realizaron objetivamente siguiendo el protocolo dado por la Comisión Internacional de Iluminación (CIE International, 1976). De acuerdo con Ouhayoun y Dalle Zotte (1993), la variabilidad en las características de color del músculo se debe a factores como el tipo de músculo (metabolismo de la energía y propiedades contráctiles), pHu del músculo así como el contenido de mioglobina y la actividad realizada por el animal (Gondret *et al.*, 2009).

Tanto la información genotípica como los procesos de edición y filtrado de genotipos fueron idénticos a los descritos en la sección 2.2: se descartaron individuos con baja tasa de genotipado (< 90%), marcadores con MAF < 0.05 y con más del 10% de datos faltantes.

4.3. Metodología.

4.3.1. Análisis de asociación genómica para caracteres de calidad en carne de cerdo.

Para estimar los componentes de varianza y los valores de cría en los caracteres evaluados, y como continuación del enfoque presentado en el capítulo 2 en relación con los análisis de asociación genómica (GWAS), utilizaremos el siguiente modelo animal centrado:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{a} + \mathbf{e} \quad [4.1]$$

Como se describió en la sección 2.3.2, \mathbf{y} corresponde al vector de fenotipos; \mathbf{X} es la matriz de incidencia relacionando las observaciones al vector de efectos fijos $\boldsymbol{\beta}$; \mathbf{e} es el vector que contiene los efectos residuales, tal que $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ y σ_e^2 es la varianza residual; \mathbf{a} es el vector de valores de cría aleatorios, siendo $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$, σ_a^2 es la varianza genética y \mathbf{G} la matriz de relaciones genómicas ($n \times n$), con n igual al número

de animales. Dicha matriz se calcula como sigue: $\mathbf{G} = \mathbf{Z}\mathbf{Z}'$, y con los elementos de la matriz \mathbf{Z} ($n \times m$; m es el número de SNP) calculados según se observa en la ecuación [2.2].

4.3.2. Análisis de la estructura genética de la población.

Lans *et al.* (2012) observaron que la inclusión de componentes principales en el modelo presentado en [4.1] es innecesaria para considerar la estructura genética de la población. Esto se debe, fundamentalmente, a que la variación genética es tenida en cuenta una vez que la matriz de relaciones genómicas \mathbf{G} ha sido considerada en el modelo. Con lo cual, los componentes principales no fueron incorporados en los análisis de asociación a nivel poblacional desarrollados durante la realización del presente capítulo.

4.3.3. Test estadísticos por población.

Para cada SNP y carácter dentro de población, se estimaron los efectos de SNP $\hat{\mathbf{g}}$ partir de una transformación lineal de los valores de cría predichos $\hat{\mathbf{a}}: \hat{\mathbf{g}} = \mathbf{Z}'\mathbf{G}^{-1}\hat{\mathbf{a}}$ (expresión [2.4]), y con sus varianzas calculadas según la fórmula $\text{Var}(\hat{\mathbf{g}}) = \mathbf{Z}'\mathbf{G}^{-1}\mathbf{Z}\sigma_a^2 - \mathbf{Z}'\mathbf{G}^{-1}\mathbf{C}^{aa}\mathbf{G}^{-1}\mathbf{Z}$ ([2.5]; Gualdrón Duarte *et al.*, 2014). El escalar σ_a^2 es la varianza genética aditiva y \mathbf{C}^{aa} es la porción de la inversa de la matriz de los coeficientes en las ecuaciones de modelo mixto para [4.1]. Por lo tanto, los estadísticos z se calcularon estandarizando los efectos de SNP estimados, al dividirlos por la raíz cuadrada de su respectiva varianza:

$$z_{ij} = \frac{\hat{\mathbf{g}}_{ij}}{\sqrt{\text{Var}(\hat{\mathbf{g}}_{ij})}} \quad [4.2]$$

Los valores z_{ij} , $\hat{\mathbf{g}}_{ij}$ y $\text{Var}(\hat{\mathbf{g}}_{ij})$ corresponden al estadístico z , al efecto estimado y a su varianza, respectivamente, para el j -ésimo SNP en la i -ésima población. Teniendo en cuenta los estadísticos z , se obtuvieron p -valores para significancia de los efectos de SNP como sigue:

$$p\text{-valor}_{ij} = 2 \left[1 - \Phi \left(\left| \frac{\hat{\mathbf{g}}_{ij}}{\sqrt{\text{Var}(\hat{\mathbf{g}}_{ij})}} \right| \right) \right] \quad [4.3]$$

donde $p\text{-valor}_{ij}$ es el p -valor asociado con el j -ésimo SNP en la población i , y $\Phi(\cdot)$ es la distribución normal estándar acumulada (Gualdrón Duarte *et al.*, 2014).

4.3.4. Meta-análisis de estudios de asociación genómica (MA-GWAS).

Siguiendo el enfoque del capítulo 3, se combinaron los estadísticos z obtenidos a partir de GWAS de poblaciones independientes (expresión [4.2]) en un estadístico z único, considerando el estimador de Hedges y Vevea (1998). Tal como se describió en el apartado 3.3.5, en el estimador propuesto por Hedges y Vevea (1998) las ponderaciones son iguales a las inversas de las varianzas de cada efecto de SNP estimado. Como se discutió en el capítulo 3, el test estadístico presentado en [4.2] es equivalente a aquel test basado en el modelo de efectos fijos de SNP (ver apartado 3.3.4 de la presente tesis). Con lo cual, para ponderar en forma óptima los test estadísticos dados en la expresión [4.2], se obtuvo la varianza de los efectos estimados de SNP, $\text{Var}(\hat{b}_i)$, estimada a partir de la varianza de los efectos aleatorios $\text{Var}(\hat{g}_i)$ (ecuación [3.23]). Conjuntamente con [3.23], la ponderación para el criterio de la inversa de la varianza es igual a:

$$w_{ij} = \frac{1}{\text{Var}(\hat{b}_{ij})} \quad [4.4]$$

Posteriormente, se generaron los estadísticos z como una combinación ponderada de aquellos obtenidos dentro de población (z_{ij} en [4.2]), es decir:

$$z_j^* = \sum_{i=1}^k \left[z_{ij} \left(\frac{\sqrt{w_{ij}}}{\sqrt{\sum_{i=1}^k w_{ij}}} \right) \right] \quad [4.5]$$

Finalmente, se calcularon los p -valores para la identificación de asociaciones significativas en el meta-análisis, siguiendo la expresión [4.3].

4.3.5. Búsqueda de genes candidatos para caracteres de calidad de carne considerando bloques de desequilibrio de ligamiento.

Se definieron intervalos para la búsqueda de genes candidatos, mediante un enfoque similar al presentado por Do *et al.* (2014), a partir de la identificación de bloques de desequilibrio de ligamiento alrededor de los picos de QTL. El proceso fue el siguiente: una vez realizado el MA-GWA para los caracteres de calidad de carne considerando el esquema de ponderación por inversa de la varianza, se identificaron los picos (cuando se detectó solo un QTL) y regiones de asociación (cuando se observaron varios SNP asociados significativamente) para cada carácter. Posteriormente, se establecieron diferentes regiones dentro de cada cromosoma, a partir del agrupamiento de las asociaciones significativas detectadas. En el caso de regiones con varios SNP significativos, los límites para la construcción de los bloques fueron situados dentro del intervalo definido entre la posición mínima y máxima observada dentro de cada región. Por su parte, en aquellas regiones con un único QTL, se definió un intervalo de ± 1

Megabase alrededor de la posición reportada. Considerando estos intervalos y teniendo en cuenta los genotipos disponibles, se cuantificó el desequilibrio de ligamiento (LD) entre los SNP que integraban cada región cromosómica en términos de la medida de LD r^2 , la cual hace referencia al cuadrado del coeficiente de correlación entre dos loci, más formalmente:

$$r^2 = \frac{(p_{AB} - p_A p_B)^2}{p_A p_a p_B p_b} \quad [4.6]$$

Los valores p_A , p_a , p_B y p_b son las frecuencias alélicas en los SNP A y B respectivamente, y p_{AB} , p_{Ab} , p_{aB} y p_{ab} son las frecuencias para cada uno de los cuatro haplotipos (Pritchard y Przeworski, 2001). Asimismo, la estimación de LD se realizó usando el paquete SNP.plotter (Luna y Nicodemus, 2007). Brevemente, el programa SNP.plotter permite crear gráficos de p -valores para asociación de forma conjunta con gráficos de bloques de LD, sea utilizando datos de haplotipos o, como en el caso de la presente tesis, mediante información de los genotipos de marcadores.

Una vez obtenidos los gráficos, se identificaron los bloques de LD que contenían los picos de asociación de cada región. De esta manera, los bloques de LD permitieron definir y refinar los límites de cada intervalo, los cuales fueron considerados en la búsqueda de genes candidatos. Así, la disponibilidad tanto de las posiciones límite en cada región como del cromosoma de interés, permitió utilizar el paquete biomaRt (<http://www.biomart.org>), el cual provee acceso directo online a bases de datos de genes anotados e información del genoma en diferentes especies. En este trabajo, se realizó una búsqueda de genes anotados en el genoma del cerdo, usando la base de datos Sus scrofa Build 10.2 (http://ensembl.org/Sus_scrofa/Info/Index). A partir de la lista de genes anotados para cada región, se evaluó la funcionalidad y mecanismos de acción, y su relación con los caracteres evaluados. En algunos casos, se consideraron antecedentes de las vías de acción en humanos, dado que muchos genes no han sido caracterizados en cerdos (Do *et al.*, 2014).

4.4. Resultados.

4.4.1. Variación fenotípica entre poblaciones.

En primer lugar, la Tabla 4.4.1.1 presenta el número de registros, media y desvío estándar para los caracteres evaluados dentro de cada uno de los conjuntos de datos disponibles.

Tabla 4.4.1.1. Número de registros, media y desviación estándar para cada carácter dentro de población.

Carácter ¹	Comercial		MARC		MSUPRP		N Total ⁴
	N ²	Media (SD) ³	N	Media (SD)	N	Media (SD)	
pHu	1857	5.63 (0.166)	530	5.812 (0.168)	904	5.512 (0.139)	3291
SF	1892	16.81 (5.628)	1234	13.79 (3.435)	911	3.205 (0.683)	4037
PRL	1780	0.885 (0.763)	673	2.955 (1.259)	920	1.838 (1.175)	3373
IMF	700	2.154 (0.770)	1234	2.276 (1.047)	910	3.182 (1.406)	2844
CKL	1780	17.24 (2.257)	1234	20.18 (3.062)	912	22.73 (2.836)	3926
CIE L*	1780	57.67 (3.343)	704	56.13 (3.756)	874	53.79 (2.238)	3358
CIE a*	1780	14.49 (1.495)	704	6.746 (1.428)	874	17.26 (1.827)	3358
CIE b*	1780	21.03 (2.07)	704	12.92 (1.686)	874	9.107 (1.603)	3358

¹ Carácter: pHu = pH 24 horas post-sacrificio; SF = Resistencia al corte; PRL= Porcentaje de pérdida de fluidos; IMF = Porcentaje de grasa intramuscular; CKL = Porcentaje de pérdidas por cocción; CIE L*, a* and b* = Caracteres de color relacionados con medidas de luminosidad, color rojo y amarillo en la carne respectivamente.

² Número de registros (N) para cada carácter y para cada población.

³ Media y desviación estándar (DE) para cada carácter y población.

⁴ Número total de registros para cada carácter a través de los conjuntos de datos evaluados.

En términos generales, las poblaciones MARC y MSUPRP tuvieron el número más bajo de registros para todos los caracteres excepto para IMF, carácter para el que se observó el menor número de fenotipos en la población comercial. Al evaluar conjuntamente todas las poblaciones, se observó un número alto de datos para SF ($n = 4037$) en comparación con el número de fenotipos disponibles para IMF ($n = 2844$). Asimismo, se reportaron valores promedio homogéneos para pHu en las tres poblaciones, en comparación con los valores informados para los caracteres restantes. En este punto debe aclararse que el rango menor observado para SF en la población MSUPRP está relacionado con el método de medición utilizado, dado que en las poblaciones MARC y Comercial se empleó la metodología de resistencia al corte por porción (SSF) mientras que en la población MSUPRP la resistencia al corte fue medida por medio de la técnica Warner-Bratzler (WBS).

4.4.2. Variación genética entre poblaciones.

Con el fin de estimar los componentes de varianza para los caracteres de calidad de carne, se ajustó el modelo descrito en [4.1], incorporando para cada carácter y población efectos fijos diferentes. En el caso de la población Comercial, el único efecto fijo disponible fue el grupo de contemporáneos definido por el grupo de establecimiento al sacrificio. Por su parte, se dispuso de información relacionada con sexo, edad al

sacrificio y grupo de sacrificio para la población MARC, así como de efectos de sexo, edad al sacrificio, grupo contemporáneo de destete y peso de carcasa para la población MSUPRP. Con lo cual, se realizó un estudio preliminar para determinar aquellos efectos relacionados con la variación sistemática de cada carácter dentro de las poblaciones MARC y MSUPRP, información que se presenta en la Tabla 4.4.2.1.

Tabla 4.4.2.1. Efectos fijos considerados para cada carácter dentro de población.

Carácter ¹	Población		
	Comercial	MARC	MSUPRP
pHu		Edad al sacrificio, grupo de sacrificio	Grupo de destete
SF			Grupo de destete, peso de la carcasa
PRL	Grupo de establecimiento de sacrificio	Sexo,	Grupo de destete
CKL		Edad al sacrificio, grupo de sacrificio	Grupo de destete, sexo
IMF			
CIE L*			
CIE a*			Grupo de destete
CIE b*			

¹ Para definición de los caracteres ver Tabla 4.4.1.1.

Posteriormente, se estimaron las varianzas genéticas y residuales así como la heredabilidad para cada combinación carácter-población, resultados que se presentan en la Tabla 4.4.2.2. En general, las estimaciones de varianza genética oscilaron entre 0.003 (pHu en la población MSUPRP) y 7.584 (SF en la población Comercial). Para la varianza residual, con estimaciones entre 0.013 a 18.122, pHu mostró el valor más bajo en comparación con aquellos observados en los demás caracteres.

Adicionalmente, las estimaciones de heredabilidad oscilaron entre 0.16 y 0.60, valores que fueron observados para CIE a* en las poblaciones MARC y MSUPRP, respectivamente. Para los caracteres restantes, las heredabilidades reportadas fueron moderadas a altas, con pequeñas variaciones a través de los conjuntos de datos.

Tabla 4.4.2.2. Estimaciones de componentes de varianza y heredabilidades para cada carácter y conjunto de datos.

Caract ¹	Varianza genética			Varianza residual			Heredabilidad		
	Co ²	Ma ³	MSU ⁴	Co	Ma	MSU	Co	Ma	MSU
pHu	0.009	0.007	0.003	0.017	0.019	0.013	0.340	0.267	0.169
SF	7.584	3.329	0.112	18.122	7.499	0.286	0.295	0.307	0.282
PRL	0.184	0.528	0.289	0.313	0.801	0.797	0.370	0.397	0.266
IMF	0.275	0.445	0.917	0.324	0.361	0.794	0.459	0.552	0.536
CKL	1.483	1.270	2.230	2.929	2.148	5.363	0.336	0.372	0.294
CIE_L*	3.712	2.882	1.625	6.545	6.325	2.957	0.362	0.313	0.355
CIE_a*	0.900	0.131	0.552	1.103	0.688	0.363	0.449	0.160	0.603
CIE_b*	1.072	0.190	0.110	2.009	0.846	0.408	0.348	0.184	0.213

¹ Carácter: Ver Tabla 4.4.1.1. para descripción de los caracteres analizados.

² Población Comercial

³ Población MARC

⁴ Población MSUPRP.

4.4.3. Análisis de asociación poblacional y meta-análisis de GWA.

Además de los análisis de asociación a nivel poblacional, para cada carácter se realizó meta-análisis de asociación considerando ponderaciones por la inversa de la varianza de los efectos estimados de SNP. Debe recordarse que este enfoque requiere una escala de medición consistente sobre todos los conjuntos de datos. Sin embargo, como se mencionó en la sección 4.4.1, las unidades de medición para resistencia al corte (SF) en las poblaciones Comercial y MARC (resistencia al corte de una porción - SSF) fueron diferentes a aquellas utilizadas en la población MSUPRP (resistencia al corte por Warner-Bratzler - WBS). Por lo tanto, y a los efectos de combinar test estadísticos para SF abarcando las tres poblaciones, se transformaron las varianzas genéticas y los efectos estimados de SNP dentro de población, teniendo en cuenta la siguiente ecuación descrita en Shackelford y Wheeler (2009):

$$\text{WBS} = (0.1063 \times \text{SSF}) + 2.2718 \quad [4.7]$$

Posteriormente, se obtuvieron las ponderaciones bajo el enfoque de la inversa de la varianza de los efectos de SNP (expresión [4.4]), y se generaron los estadísticos z como una combinación ponderada de los estadísticos obtenidos dentro de cada población (ecuación [4.5]). De este modo, la Tabla 4.4.3.1 presenta las asociaciones más significativas para los caracteres de calidad de carne, sus respectivas regiones, el efecto

estimado y el p -valor, implementados por análisis de asociación poblacional como utilizando meta-análisis de GWAS.

Tabla 4.4.3.1. Asociaciones significativas para caracteres de calidad de carne usando GWAS a nivel poblacional y meta-análisis basado en ponderación por inversa de la varianza.

Car ¹	GWA ²	SNP_ID ³	Cr ⁴	Pico ⁵ (Mb)	Reg. Signif (Mb) y No. SNP ⁶	Ef ⁷	p -valor
pHu	Com	MARC0039273	15	133.9	133.6-134.1 (7 SNP)	6.275	3.503E-10
	MSU	MARC0047188	15	135.1	129.7-138.7 (10 SNP)	-6.658	2.774E-11
	MA	H3GA0052416	15	135.2	133.2-135.7 (12 SNP)	-6.652	2.897E-11
SF	Com	DRGA0003285	2	109.6	109.6	5.376	7.609E-08
		SIRI0001264	8	142.3	142.3	-5.157	2.506E-07
	MARC	H3GA0055977	2	5.4	5.4	5.149	2.616E-07
	MSU	M1GA0002229	2	2.9	2.9-5.4 (3 SNP)	-5.223	1.760E-07
		H3GA0052416	15	135.2	133.1-135.5 (5 SNP)	-5.645	1.654E-08
	MA	H3GA0055977	2	5.4	5.1-6.2 (11 SNP)	5.937	2.911E-09
		ASGA0011029	2	106.4	106.4	-5.132	2.868E-07
PRL	MSU	ALGA0087273	4	75.05	75.05	-5.428	5.711E-08
		MARC0093624	15	135.5	133.1-145.6 (34 SNP)	-7.706	1.310E-14
	MA	M1GA0020450	15	133.9	133.7-134.0 (5 SNP)	-5.984	2.178E-09
CKL	Com	ALGA0109178	6	0.013	0.013	6.087	1.150E-09
		M1GA0020450	15	133.9	133.6-134.0 (6 SNP)	-5.781	7.410E-09
	MSU	MARC0047188	15	135.1	133.1-137.5 (25 SNP)	7.812	5.551E-15
	MA	ALGA0109178	6	0.013	0.013	5.961	2.506E-09
		M1GA0020450	15	133.9	133.2-135.2 (11 SNP)	-7.311	2.647E-13
	IMF	Com	ALGA0109178	6	0.013	0.013	-9.271
	MA	ALGA0109178	6	0.013	0.013	-9.581	1.332E-15
CIE L*	Com	M1GA0020450	15	133.9	133.9	-5.685	1.310E-08
CIE a*	MA	ALGA0103022	1	308.9	308.9	5.514	3.516E-08
		DIAS0000492	6	49.8	48.5-63.1 (20 SNP)	-5.987	2.143E-09

¹ Carácter: Ver Tabla 4.4.1.1. para descripción de los caracteres analizados. ² Tipo de análisis: GWAS en población Comercial (Com), en población MARC (MARC) o en población MSUPRP (MSU); Meta-análisis considerando ponderación por inversa de la varianza (MA). ³ Nombre del SNP más significativo. ⁴ Cromosoma (Sus scrofa genome build 10.2). ⁵ Posición del SNP en el pico de asociación en Megabases. ⁶ Región asociada significativamente, expresada en Megabases y número de SNP significativos en la región. ⁷ Efecto fenotípico estimado por una copia del alelo B.

Asimismo, las Figuras 4.4.3.1. a 4.4.3.8 presentan los gráficos de Manhattan (“Manhattan plots”) para todos los caracteres evaluados, tanto a partir del GWA poblacional como del MA-GWA. En dichas figuras, los ejes x e y muestran la posición absoluta de los SNP en Megabases y el $-\text{Log}_{10}(p\text{-valor})$ respectivamente, considerando un umbral de significancia sobre todo el genoma tal que $p < 0.05$ (línea roja).

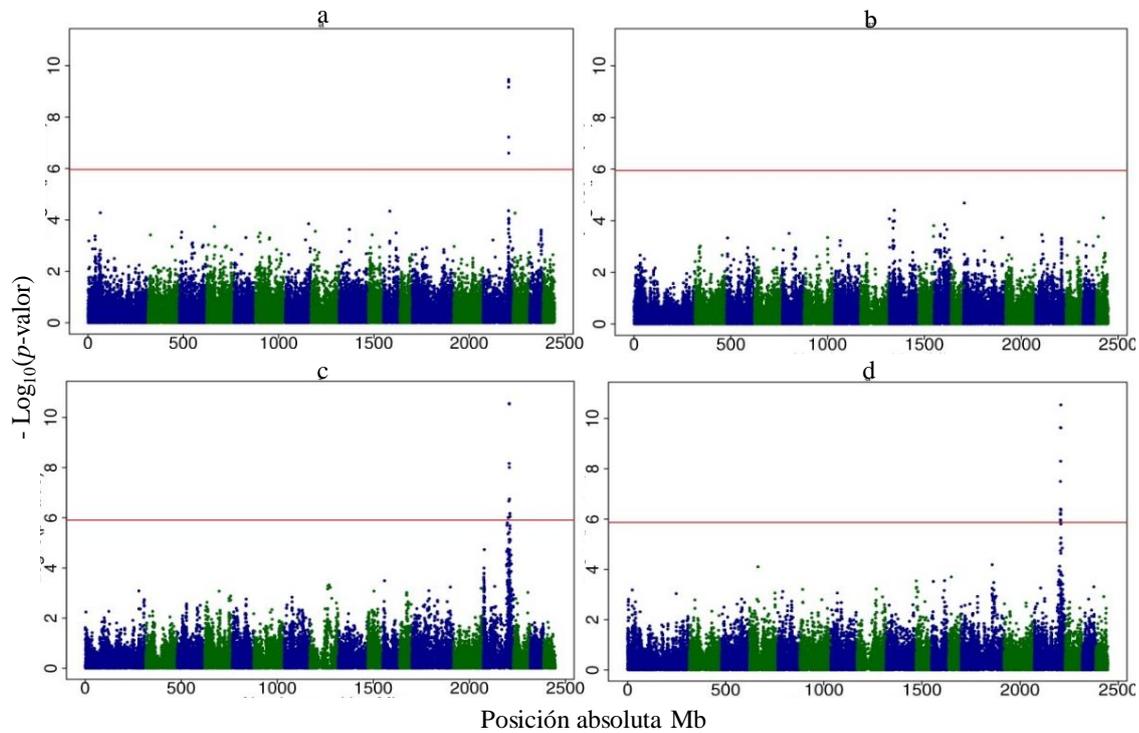


Figura 4.4.3.1. Gráfico Manhattan para pH 24 horas post-sacrificio considerando GWAS en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWAS.

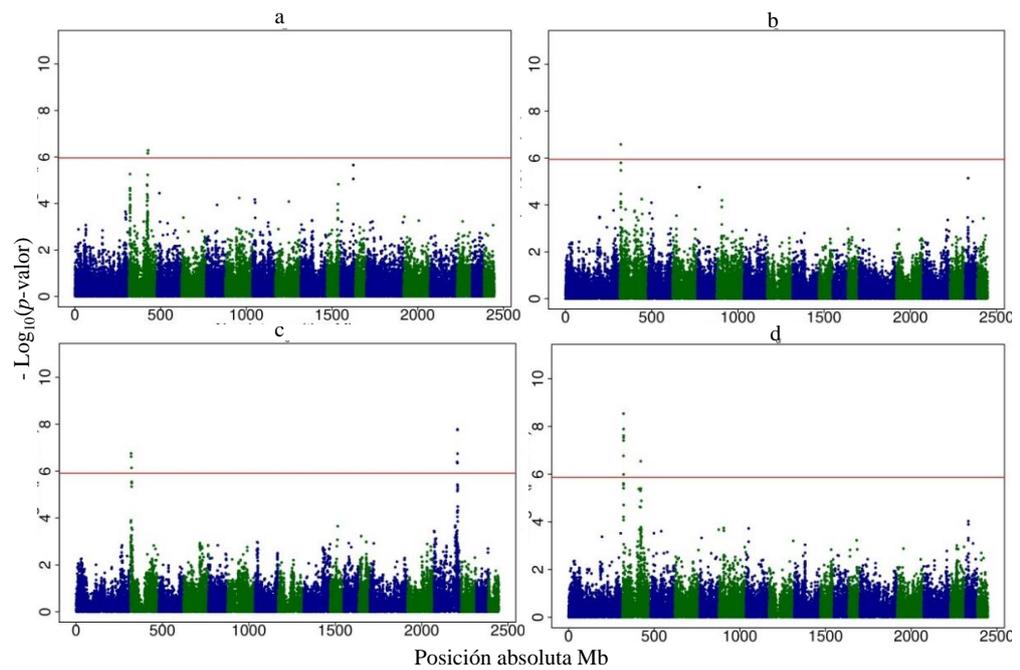


Figura 4.4.3.2. Gráficos de Manhattan para resistencia al corte considerando GWAS en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWAS.

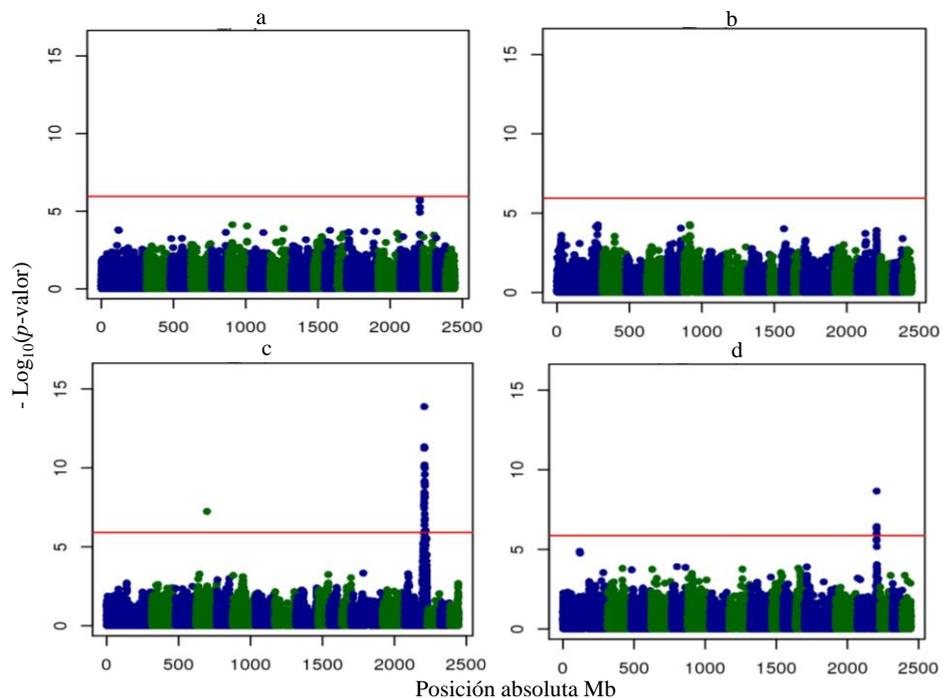


Figura 4.4.3.3. Gráficos de Manhattan para porcentaje de pérdida de líquidos a partir de GWAS en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWAS.

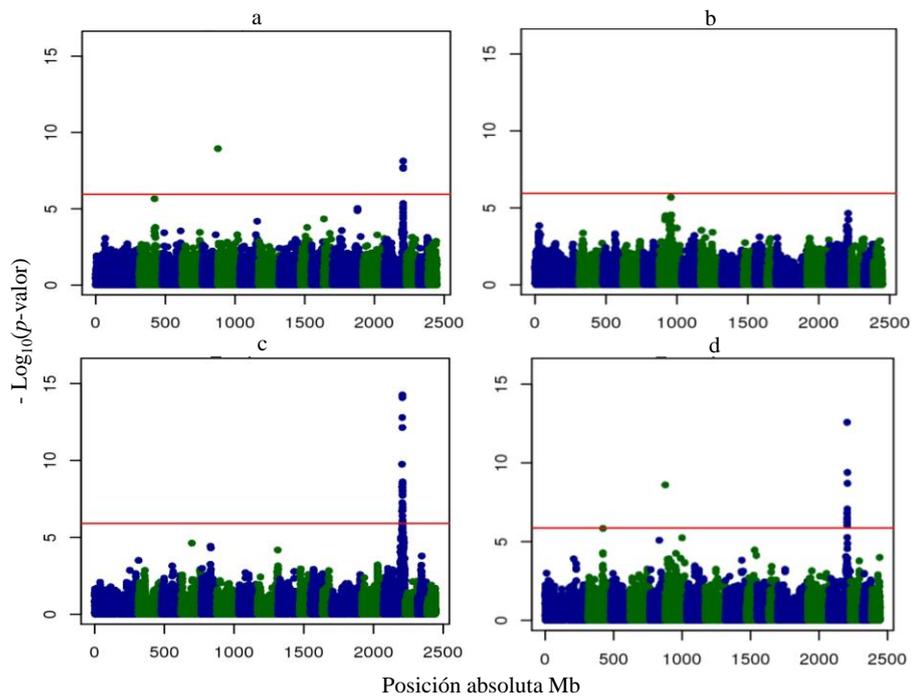


Figura 4.4.3.4. Gráficos de Manhattan para porcentaje de pérdidas por cocción a partir de GWAS en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWAS.

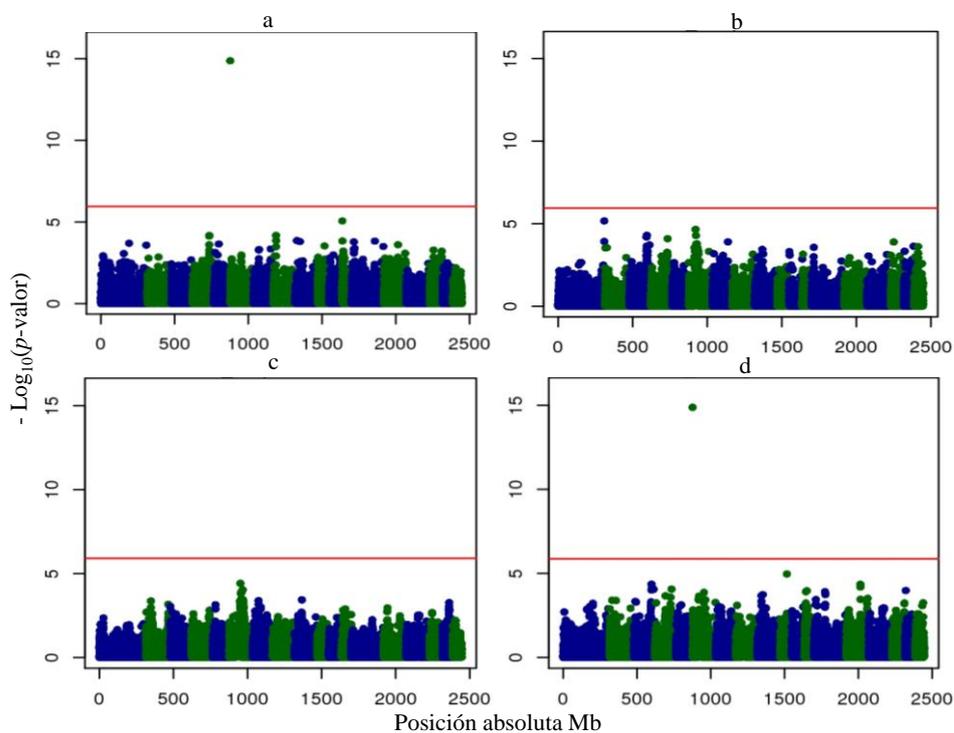


Figura 4.4.3.5. Gráficos de Manhattan para porcentaje de grasa intramuscular a partir de GWAS en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWAS.

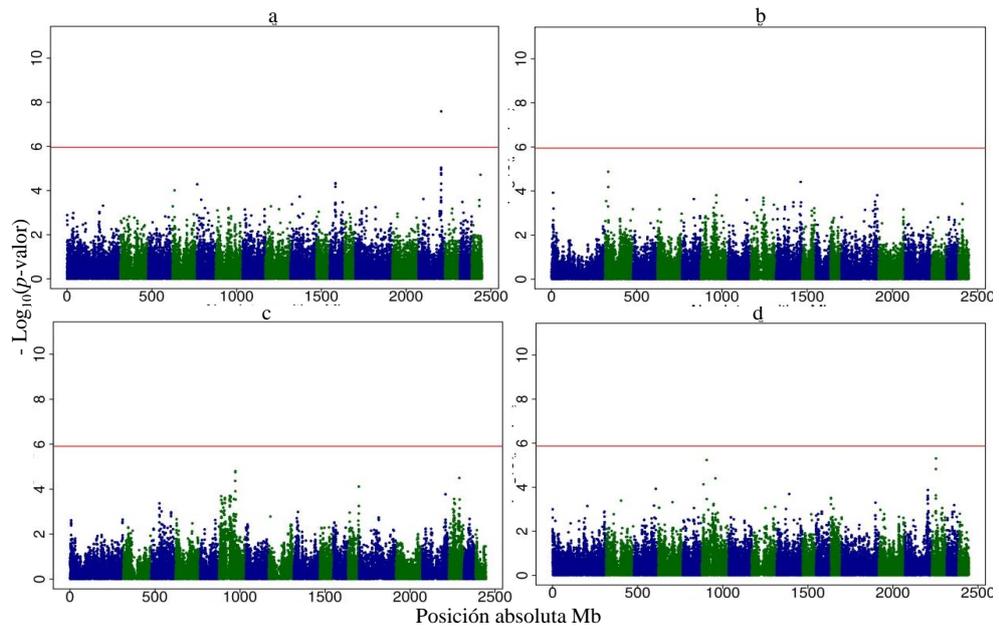


Figura 4.4.3.6. Gráficos de Manhattan para CIE L* (luminosidad) a partir de GWAS en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWAS.

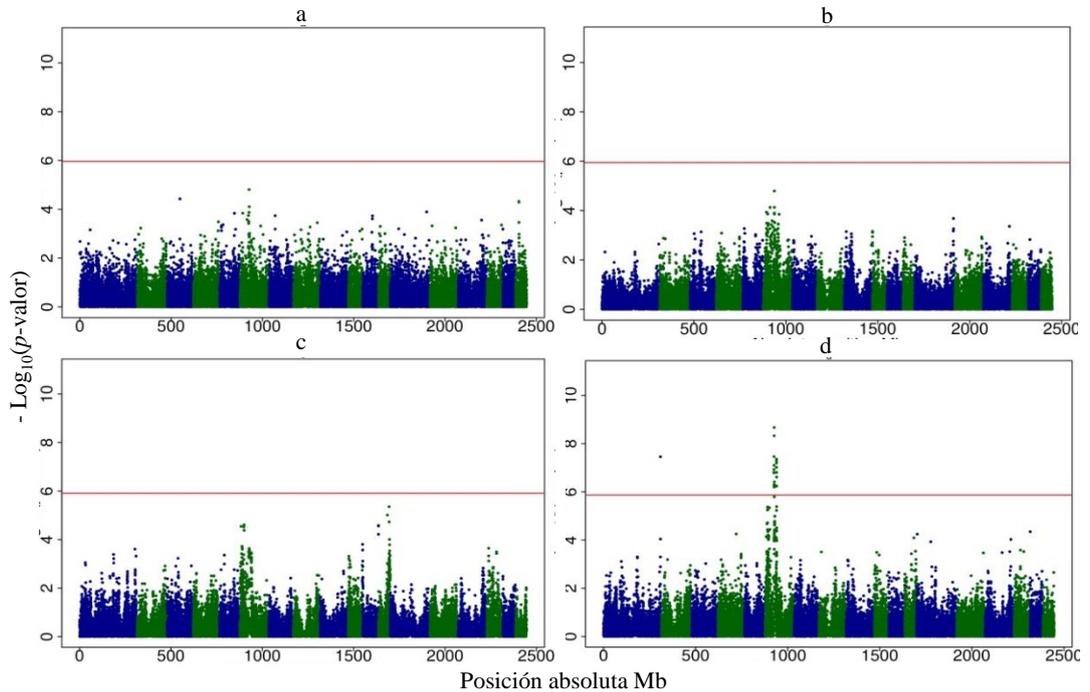


Figura 4.4.3.7. Gráficos de Manhattan para CIE a* (color rojo) a partir de GWAS en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWAS.

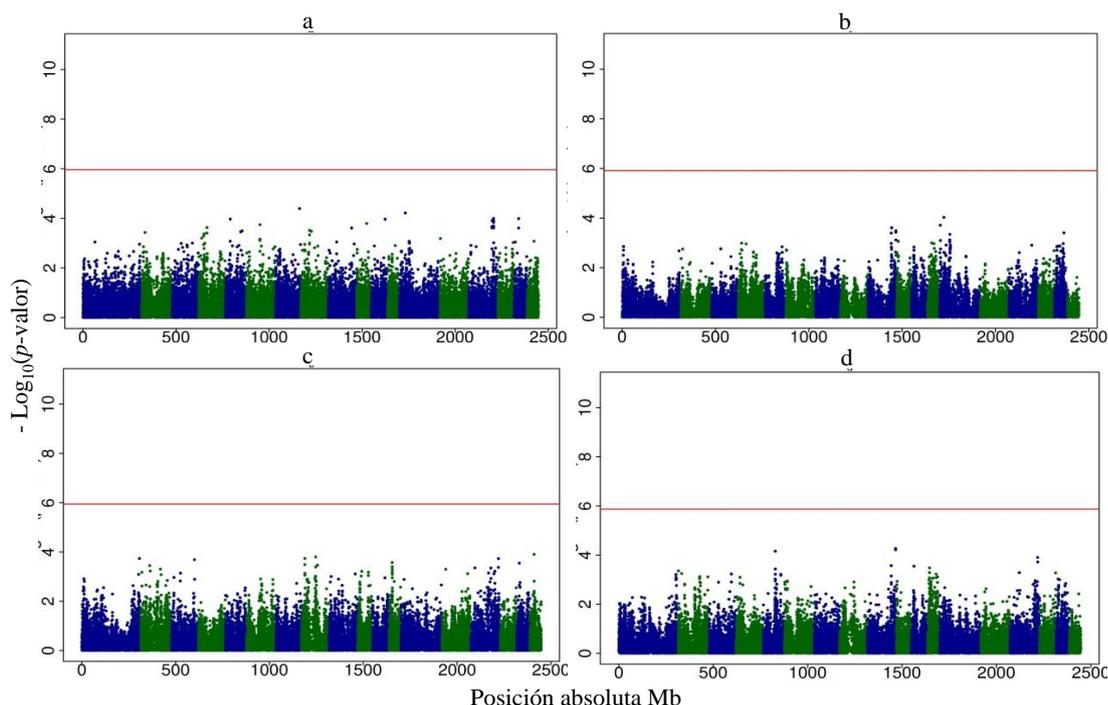


Figura 4.4.3.8. Gráficos de Manhattan para CIE b^* (color amarillo) a partir de GWAS en las poblaciones: a. Comercial; b. MARC; c. MSUPRP; d. MA-GWA.

En primer lugar, los análisis de asociación poblacional identificaron dos regiones genómicas significativas para pHu ubicadas en el cromosoma 15. La primera, situada entre 133.6Mb y 134.1Mb, abarcó siete SNP significativos y fue observada en la población Comercial. Dicha región estuvo incluida dentro de la segunda, la cual se halló definida por el intervalo 129.7Mb - 138.7Mb, comprendiendo 10 SNP e identificada en la población MSUPRP. En dichas poblaciones, los picos de asociación estuvieron ubicados a 133.9Mb (MARC0039273, p -valor $< 3.503 \times 10^{-10}$) y 135.1Mb (MARC0047188, p -valor $< 2.774 \times 10^{-11}$) respectivamente.

Para SF, se identificaron asociaciones significativas en los cromosomas 2, 8 y 15. En relación al cromosoma 2, se observaron dos SNP significativos en las poblaciones Comercial y MARC (uno en cada población) y una región significativa en la población MSUPRP. Esta última región estuvo comprendida entre 2.9Mb y 5.4Mb e integrada por tres SNP. En dicho cromosoma, las asociaciones observadas en las poblaciones MARC y MSUPRP se ubicaron cercanamente (H3GA0055977, 5.4Mb, p -valor 2.616×10^{-7} y M1GA0002229, 2.9Mb, p -valor $< 1.76 \times 10^{-7}$), en contraste con el QTL reportado en la población Comercial, el cual se encontró a más de 100Mb de los picos observados en las otras dos poblaciones (DRGA0003285, 109.6Mb, 7.609×10^{-8}). Por su parte, los picos significativos adicionales observados en el cromosoma 8 y cromosoma 15 fueron identificados a 142.3Mb (SIRI0001264, p -valor $< 2.506 \times 10^{-7}$) y a 135.2Mb (H3GA0052416, p -valor $< 1.654 \times 10^{-8}$) en las poblaciones Comercial y MSUPRP, respectivamente.

En el caso de PRL, se observaron asociaciones significativas únicamente en la población MSUPRP. Específicamente, un QTL en el cromosoma 4, ubicado a 75.05Mb (ALGA0087273, p -valor $< 5.711 \times 10^{-8}$) y adicionalmente, una región significativamente asociada en el cromosoma 15, con un pico de asociación a 135.5Mb (MARC0093624, p -valor $< 1.31 \times 10^{-14}$) y compuesta por 34 SNP.

Con respecto a CKL, el GWA poblacional identificó asociaciones significativas en las poblaciones Comercial y MSUPRP. En el primer conjunto de datos, se detectó un QTL significativo en el cromosoma 6 a 0.013Mb (ALGA0109178, p -valor $< 1.15 \times 10^{-9}$) y una región asociada en el cromosoma 15 integrada por seis SNP y con un pico significativo en 133.9Mb (M1GA0020450, p -valor $< 7.41 \times 10^{-9}$). Este SNP estuvo ubicado a 1.2Mb de distancia del pico identificado en la población MSUPRP (MARC0047188, p -valor $< 5.551 \times 10^{-15}$) en un conjunto de 15 SNP significativos para CKL. En este caso, la región observada en la población MSUPRP fue más amplia que la reportada para la población comercial, debido principalmente al alto LD observado en el conjunto de datos de la primera población. Además de la asociación reportada en el cromosoma 6 para CKL, el GWA poblacional permitió identificar una asociación significativa adicional entre dicho QTL y el carácter IMF (ALGA0109178, p -valor $< 1.332 \times 10^{-15}$).

En relación con los caracteres de color de la carne, se observó un QTL significativo para CIE L* en la población Comercial, situado en la posición 133.9Mb del cromosoma 15 (M1GA0020450, p -valor $< 1.31 \times 10^{-8}$). Comparativamente, los análisis de asociación poblacionales no detectaron QTL significativos para CIE a* y CIE b*.

Comparando los resultados del MA con aquellos obtenidos en los GWAS poblacionales, el MA replicó las asociaciones reportadas a nivel de población para pHu (cromosoma 15), SF (cromosoma 2), PRL (cromosoma 15), CKL (cromosomas 6 y 15) e IMF (cromosoma 6). Adicionalmente, el MA identificó nuevos QTL significativos para los caracteres SF y CIE a*. En el caso de SF, el MA-GWAS identificó un nuevo QTL en el cromosoma 2 ubicado en la posición 106.4Mb (ASGA0011029). Para CIE a*, el MA detectó un SNP significativo en el cromosoma 1, ubicado en la posición 308.9Mb (ALGA0103022, p -valor $< 3.52 \times 10^{-8}$) y también, una región significativa en el cromosoma 6 definida por el intervalo 48.5Mb – 63.1Mb. En relación a dicha región, y si bien a nivel poblacional se habían observado algunos picos para CIE a* en el cromosoma 6 (Figuras 2.4.5.1.a-c), ninguno de estos QTL alcanzó el nivel de significancia requerido considerando la corrección de Bonferroni ($0.05/\text{Número de pruebas}$). De este modo, el MA permitió incrementar la potencia de detección de asociación, a partir de la combinación de resultados de GWAS poblacionales.

Asimismo, la Tabla 4.4.3.1 muestra que la implementación del MA permitió refinar algunas de las regiones identificadas como significativas a nivel poblacional. Tal es el caso de PRL, con un intervalo inicial dado por 133.1Mb-145.6Mb en el cromosoma 15 y reducido por el MA a la región 133.7 y 134Mb. De modo similar, el MA permitió refinar la región de asociación significativa para pHu en el cromosoma 15, pasando de 129.7Mb - 138.7Mb a 133.2Mb-135.7Mb.

4.4.4. Búsqueda de genes candidatos para caracteres de calidad de carne.

A partir de las regiones identificadas como significativas por medio del MAGWAS (Tabla 4.4.3.1), se procedió a la búsqueda de genes candidatos para los caracteres de calidad analizados. Con el fin de facilitar la búsqueda, se agruparon las asociaciones por cromosoma, tal como se presenta en la Tabla 4.4.4.1.

Tabla 4.4.4.1. Asociaciones significativas identificadas por meta-análisis de asociación para caracteres de calidad agrupadas por cromosoma.

Cr ¹	Región ² (Mb)	Carácter
1	308.9	CIE a*
2	5.1-6.2	Resistencia al corte
2	106.4	Resistencia al corte
6	0.013	Porcentaje grasa intramuscular Porcentaje de pérdidas por cocción
6	48.5-63.1	CIE a*
15	133.2-135.7	pH 24 horas post-sacrificio Porcentaje de pérdida de líquidos Porcentaje de pérdidas por cocción

¹ Cromosoma (Sus scrofa genome build 10.2).

² Posición del SNP en el pico de asociación o región asociada significativamente, expresada en Megabases.

De acuerdo a la Tabla 4.4.4.1, se identificaron 6 regiones asociadas significativamente con pHu, SF, CKL, PRL, IMF y CIE a*. Para cada región, se estimó el LD y se construyeron bloques de LD siguiendo la metodología presentada en el apartado 4.2.5. En el caso de regiones comprendiendo varios SNP, los límites para la estimación de LD ocurrieron en los intervalos definidos en la Tabla 4.4.4.1: cromosoma 2: 5.1Mb-6.2Mb; cromosoma 6: 48.5Mb-63.1Mb; cromosoma 15: 133.2Mb-135.7Mb. Por su parte, en aquellas regiones con una sola asociación, los límites fueron definidos por el intervalo entre ± 1 Megabase alrededor de la posición reportada (cromosoma 1: 307.9Mb-309.9Mb; cromosoma 2: 105.4Mb-107.4Mb; cromosoma 6: 0Mb-1Mb). Así, empleando el programa SNP.plotter (Luna y Nicodemus, 2007), se obtuvieron los gráficos presentados en las Figuras 4.4.4.1 a 4.4.4.3.

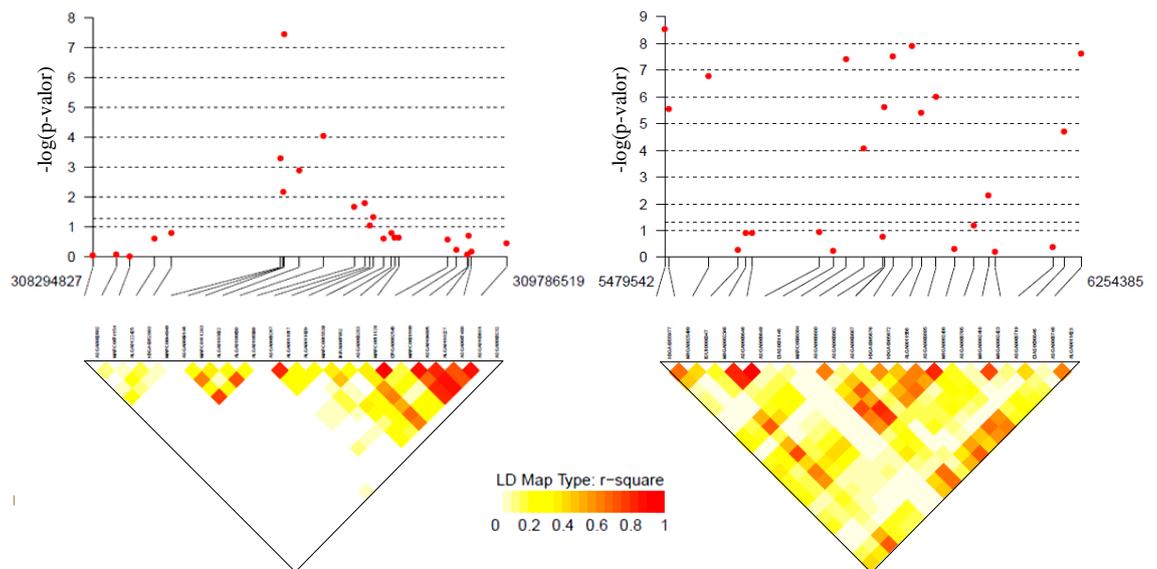


Figura 4.4.4.1. Gráficos de LD y p -valores para asociaciones con: a. CIE a* (Cromosoma 1: 307.9Mb-309.9Mb); b. SF (Cromosoma 2: 5.1Mb-6.2Mb).

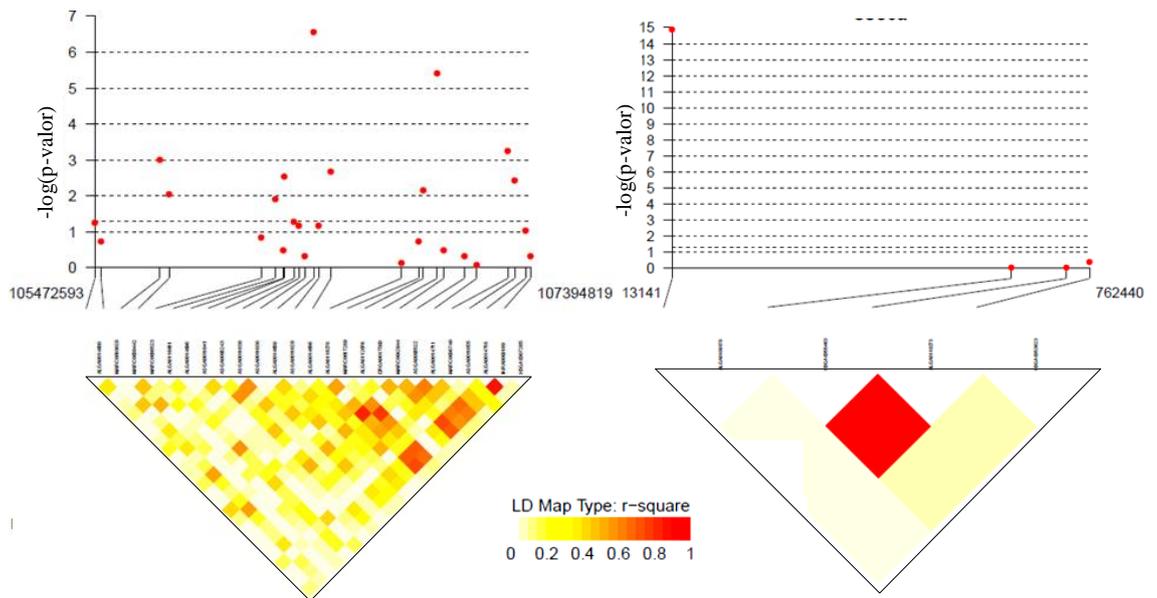


Figura 4.4.4.2. Gráficos de LD y p -valores para asociaciones con: a. SF (Cromosoma 2: 105.4Mb-107.4Mb); b. IMF y CKL (Cromosoma 6: 0Mb-1Mb).

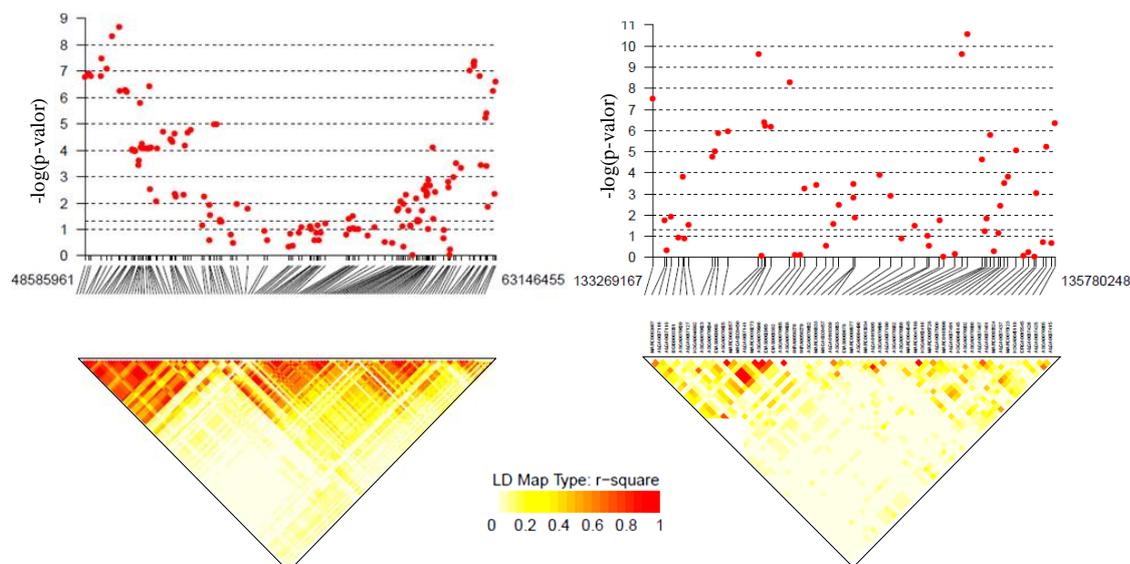


Figura 4.4.4.3. Gráficos de LD y p -valores para asociaciones con: a. CIE a* (Cromosoma 6: 48.5Mb-63.1Mb); b. pHu, PRL y CKL (Cromosoma 15: 133.2Mb-135.7Mb).

En primer lugar, en la región observada en el cromosoma 1 y asociada con el color rojo de la carne CIE a*, se identificaron tres bloques de LD en las regiones 308.2Mb-308.5Mb, 308.9Mb-309.1Mb y 309.2Mb-309.6Mb (Figura 4.4.4.1.a). Teniendo en cuenta la posición de la asociación más significativa en esta región, se consideró el segundo intervalo en la búsqueda de genes candidatos dentro de la base de datos Ensembl, proceso que retornó tres proteínas no caracterizadas, así como el gen *VNIR2* (vomeronasal 1 receptor 2).

Para el caso de las asociaciones encontradas en el cromosoma 2 respecto de SF, la primera región dio lugar a un bloque de LD que abarcó la totalidad del intervalo (5.4Mb-6.2Mb, Figura 4.4.4.1.b), con un total de 56 genes anotados. En la segunda, el bloque de LD estuvo definido por el intervalo 105.4Mb-107.3Mb (Figura 4.4.4.2.a), segmento en el cual se encontraron 10 genes anotados.

Asimismo, en la región del cromosoma 6 que se asoció significativamente con IMF y CKL se observó un único bloque de LD comprendido entre 0.013Mb y 0.76Mb (Figura 4.4.4.2.b) y se identificaron 22 genes candidatos. Por su parte, en el segundo segmento de dicho cromosoma y asociado con mediciones de CIE a*, se observaron dos bloques de LD dados por los intervalos 48.5Mb-52.3Mb y 52.7Mb-63.1Mb (Figura 4.4.4.3.a). La búsqueda de genes anotados se centró en el primer intervalo, proceso a partir del cual se identificaron 140 genes candidatos. Si bien en dicha figura se observa un pico en el segundo bloque de LD, la selección del primer intervalo se basó en la presencia del SNP más significativo. Debe aclararse que en dicha región, la presencia de un pico en el segundo bloque de LD está relacionado principalmente con el valor de LD, entre moderado y alto, observado en las poblaciones evaluadas: $r^2 = 0.1357, 0.1498$ y 0.2020 para SNP distanciados a 1Mb en las poblaciones Comercial, MARC y MSUPRP, respectivamente.

Finalmente en el cromosoma 15, donde se reportaron asociaciones significativas para pHu, PRL y CKL, se identificaron dos bloques de LD ubicados entre 133.2Mb-134.2Mb y 135Mb-135.7Mb (Figura 4.4.4.3.b). A diferencia de las asociaciones observadas en PRL y CKL que se situaron en el segundo intervalo, las asociaciones más significativas para pHu estuvieron en el primer bloque, con lo cual la búsqueda de genes candidatos en este cromosoma se realizó considerando todo el intervalo: entre 133Mb a 135.7Mb. De este modo, se obtuvieron 38 genes anotados por Ensembl.

Cada uno de los genes fue evaluado considerando sus procesos biológicos, vías metabólicas y su posible relación con cada carácter. Por consiguiente, la Tabla 4.4.4.2 presenta aquellos genes identificados como candidatos posibles para los caracteres evaluados así como una breve descripción de las características funcionales de cada gen. Algunos de ellos han sido reportados previamente por estar asociados con características fisiológicas musculares, o con caracteres de calidad de carne en cerdos u otras especies.

Tabla 4.4.4.2. Genes candidatos identificados para caracteres de calidad de carne.

Cr ¹	Región significativa		Gen candidato	
	Región ²	Caracter ³	Símbolo ⁴	Nombre ⁵
2	5.3-6.5	SF	<i>CAPN1</i>	Calpain-1 catalytic subunit
2	105.02-109.6	SF	<i>CAST</i>	Calpastatin
6	0-1	IMF, CKL	<i>ACSF3</i>	Acyl-CoA synthetase family member 3
6	47.3-52.7	CIE a*	<i>GYS1</i>	Glycogen synthase 1 (muscle)
			<i>FTL</i>	Ferritin light polypeptide.
15	131.9-135.7	pHu, PRL, CKL	<i>PRKAG3</i>	Protein kinase, AMP-activated, gamma 3 non-catalytic subunit

¹ Cromosoma

² Región significativa expresada en Megabases

³ Caracteres: PRL = Porcentaje de pérdida de fluidos; pHu = pH medido 24 horas post-sacrificio; SF = Resistencia al corte; CKL = Porcentaje de perdidas por cocción; IMF = Porcentaje de grasa intramuscular; CIE a* = Medida de color rojo en la carne.

⁴ Símbolo HGNC del gen candidato (<http://www.genenames.org/>)

⁵ Nombre del gen candidato (<http://www.genenames.org/>).

Para SF se identificaron como candidatos los genes *CAPN1* (Calpain-1 Catalytic Subunit, 2: 6.12Mb-6.15Mb) y *CAST* (Calpastatin; 2:106.4Mb-109.6Mb). Para CKL y IMF, el gen *ACSF3* (Acyl-CoA Synthetase Family Member 3 Mitochondrial; 6: 0.82Mb-0.84Mb) está ubicado cerca de los picos de asociación identificados para estos caracteres. Para CIE a*, se identificaron dos genes como candidatos para mediciones de color rojo CIE a*. Primero, el gen *GYS1* (Sus scrofa glycogen synthase 1-muscle; 6: 50.07Mb-50.09Mb) y adicionalmente, *FTL* (Sus scrofa ferritin, light polypeptide- light chain ferritin gene; 6: 50.09Mb). Finalmente, en el cromosoma 15, *PRKAG3* (Sus scrofa protein kinase, AMP-activated, gamma 3 non-catalytic subunit; 15: 133.8Mb), reportado ampliamente por tener efecto en diferentes caracteres de calidad de carne de cerdo (Milan et al., 2000; Ciobanu et al., 2001), es el gen candidato más importante para pHu, PRL y CKL en cerdos.

4.5. Discusión.

En este capítulo se presentaron los resultados obtenidos a partir de la aplicación de MA-GWAS para diferentes caracteres de calidad de carne en cerdos, cuya importancia económica ha sido discutida previamente (Cannon *et al.*, 1996; Ovílo *et al.*, 2002; Dube *et al.*, 2013). Para ello se empleó la metodología descrita en el capítulo 3 en cuanto al meta-análisis a partir de estudios de asociación genómica GBLUP. Asimismo, los resultados obtenidos a partir del MA-GWAS fueron utilizados para determinar asociaciones significativas no detectadas a nivel poblacional, y para identificar genes candidatos potencialmente relacionados con los caracteres evaluados.

En primer lugar, se analizaron las diferencias fenotípicas sobre todas las poblaciones, respecto de cada carácter. Si bien los valores medios obtenidos para la mayoría de ellos son similares a los reportados en la literatura (Cherel *et al.*, 2011; Ma *et al.*, 2013), se observaron diferencias para IMF, donde la población MSUPRP presentó el mayor promedio para dicho carácter. En este caso, los animales que integran la población MSUPRP están compuestos por cruces Duroc \times Pietrain, razas caracterizadas por poseer un mayor contenido de grasa intramuscular comparado con la raza Landrace (Cameron *et al.*, 1990; Edwards *et al.*, 2003; Trefan *et al.*, 2013). Además, teniendo en cuenta la disponibilidad de diferentes efectos fijos para cada carácter dentro de los conjuntos evaluados, los análisis estadísticos iniciales permitieron determinar los efectos significativos a ser incluidos en cada set de datos. En este caso, la edad al sacrificio tuvo efectos significativos en todos los caracteres tomados en la población MARC. Por su parte, el efecto de sexo y peso de res fueron incluidos en los modelos para IMF y CIE L* en MSUPRP y para todos los caracteres excepto pHu, en animales de la población MARC. Al respecto, Liu *et al.* (2007) observaron resultados similares para caracteres de composición de carcasa, donde la edad al sacrificio fue incluida junto con el efecto de sexo y peso de res en los modelos para pH, color de la carne, pérdida de fluidos, pérdidas por cocción y resistencia al corte.

Las varianzas genéticas y residuales, así como las heredabilidades para los caracteres evaluados son presentadas en la Tabla 4.4.2.2. En general, las estimaciones de heredabilidad para los caracteres evaluados tuvieron un rango amplio, con valores mínimos y máximos de 0.16 y 0.60 para CIE a* en MARC y MSUPRP respectivamente. En nuestros conjuntos de datos, el rango amplio de heredabilidades fue observado principalmente para CIE a*. Sin embargo, las estimaciones globales fueron moderadas a altas y similares a resultados reportados en la literatura. Por ejemplo, Gjerlaug-Enger *et al.* (2010) reportaron heredabilidades oscilando entre 0.23 y 0.33 para PRL, 0.19 a 0.27 para pHu, de 0.50 a 0.62 para IMF y entre 0.28 a 0.41, 0.43 a 0.46 y de 0.31 a 0.33 para CIE L*, CIE a* y CIE b* respectivamente, las cuales fueron relacionadas a estimaciones presentadas en este capítulo.

En cuanto a los resultados de los análisis de asociación y exceptuando las mediciones de color CIE a* y CIE b*, se identificaron asociaciones significativas para los demás caracteres de calidad de carne en diferentes regiones de los cromosomas 2, 4, 6, 8 y 15. En primer lugar, se observaron regiones significativas superpuestas en el

cromosoma 15 asociadas con pHu en dos de las poblaciones evaluadas (Comercial y MSUPRP). En estos casos, los picos de asociación estuvieron ubicados a 7Mb y 5Mb del QTL reportado para el mismo carácter por Uimari *et al.* (2013). Para SF, se identificaron tres SNP significativos en el cromosoma 2 (uno en cada población), uno en el cromosoma 8 (población Comercial) y uno en el cromosoma 15 (MSUPRP). El SNP observado en el cromosoma 2 en la población MARC estuvo ubicado a 500Kb del SNP reportado por Nonneman *et al.* (2013) para el mismo conjunto de datos. Además, el pico significativo adicional, identificado en la población Comercial en el cromosoma 2, se localizó a 17Mb de un QTL significativo reportado por Nonneman *et al.* (2011) para la población MARC y otras poblaciones de cerdos. En el caso de las regiones significativas observadas en la población MSUPRP (uno en el cromosoma 2 y otro en el 15), Choi *et al.* (2011) reportaron un QTL significativo en el cromosoma 15 para SF en la población MSUPRP, delimitado por dos marcadores microsatélites ubicados entre las posiciones 127.9Mb y 135.9Mb (intervalo mayor al observado en la presente investigación).

Para PRL se observó un SNP significativo en el cromosoma 4 ubicado a 6Mb del QTL reportado para este carácter por Ma *et al.* (2013). Asimismo, se detectó una región significativa en el cromosoma 15 para dicho carácter, segmento que incluyó el QTL reportado por Ryan *et al.* (2012) a 142Mb. Con respecto a CKL, los análisis de asociación a nivel poblacional identificaron un QTL significativo en el cromosoma 6 en la población Comercial y también, una región asociada en el cromosoma 15 tanto en la población Comercial como en MSUPRP. Estas regiones en el cromosoma 15 fueron similares a las informadas para el mismo carácter por Nonneman *et al.* (2013), quienes reportaron picos de asociación en las posiciones 133.4Mb y 133.8Mb.

Para IMF se detectó un QTL significativo al inicio del cromosoma 6 en la población Comercial, el cual estuvo asociado también con CKL. Dicho QTL estuvo ubicado a 9Mb del QTL reportado para IMF por Paszek *et al.* (2001) y muy distante (aproximadamente 20Mb) de la región significativa identificada por Fontanesi *et al.* (2010). Para las mediciones de CIE L*, el SNP significativo identificado en el cromosoma 15 en la población Comercial se localizó en una región similar al QTL reportado para el mismo carácter por Choi *et al.* (2011). Sin embargo, el QTL detectado por los autores se observó en la población MSUPRP, resultado debido posiblemente a la diferencia en la resolución utilizada en cada estudio, así como al modelo de análisis empleado.

Comparando los resultados obtenidos a partir de MA con aquellos observados en los GWAS a nivel poblacional, los cuadros “d” en las Figuras 4.4.3.1 a 4.4.3.8 presentan gráficos de Manhattan para MA de asociación de caracteres de calidad de carne. Picos significativos de asociación son resumidos en la Tabla 4.4.3.1. Primero, un QTL significativos en el cromosoma 1 en la posición 308.9Mb (ALGA0103022) fue asociado con CIE a*. En el cromosoma 2, se observaron asociaciones significativas con SF en la posición 5.4Mb (H3GA0055977) y 106.4Mb (ASGA0011029). En el caso del cromosoma 6, se identificó un QTL asociado con IMF y CKL a 0.013Mb (ALGA0109178) y además, una asociación significativa con CIE a* en la posición 49.8Mb (DIAS0000492). Finalmente, se identificó una región en el cromosoma 15

asociada con los caracteres pHu (H3GA0052416, 135.2Mb) y PRL y CKL (M1GA0020450, 133.9Mb para los dos caracteres). No se identificaron asociaciones significativas para CIE b* y CIE L* por medio del MA-GWA.

Para cada intervalo de QTL observado con el MA y definido de acuerdo a bloques de LD alrededor de los picos de QTL, se identificaron diferentes genes candidatos para los caracteres de calidad de carne evaluados (Tabla 4.4.4.2), excepto para la región en el cromosoma 1 asociada con CIE a* (pico de asociación en la posición 308.9Mb). En este caso, el intervalo inicial fue definido entre 307.9Mb y 309.9Mb (± 1 Mb alrededor del pico de asociación), el cual fue caracterizado por un alto LD entre SNP alrededor del pico y en los bloques de LD contiguos, pero con bajo LD con las regiones adyacentes. Con lo cual, se refinaron los intervalos genómicos considerando el nivel de LD alrededor de los picos significativos (308.9Mb-309.1Mb), y en esta región, el proceso de búsqueda permitió identificar un gen (*VNIR2*, vomeronasal 1 receptor 2, cromosoma 1: 308.9Mb) y tres proteínas no caracterizadas. Sin embargo, a nuestro conocimiento no se tiene evidencia en relación a la asociación biológica entre el gen *VNIR2* y el carácter CIE a*. Adicionalmente, el bloque de LD en la región fue pequeño (0.2Mb), lo cual podría limitar el intervalo de búsqueda y por lo tanto, disminuir la potencia para detectar genes adicionales posiblemente asociados con CIE a*. Estos resultados sugieren la importancia de especificar genes anotados en la región con el fin de determinar si existe una asociación entre las proteínas no caracterizadas encontradas en el intervalo o si hay diferentes genes que no han sido anotados para este cromosoma.

En el caso de SF y especialmente para la primera región en el cromosoma 2 (5.3Mb – 6.5Mb), el gen *CAPN1* (calpain-1 catalytic subunit, cromosoma 2: 6.12Mb-6.15Mb) puede ser mencionado como gen candidato en cerdos. En estudios previos, la región cercana al gen *CAPN1* se ha asociado con SF en cerdos (Nonneman *et al.*, 2013) y en el caso de ganado de carne, ha sido vinculada con caracteres de ternura en diferentes poblaciones (Page *et al.*, 2004; White *et al.*, 2005; McClure *et al.*, 2012). De acuerdo con Huff-Lonergan *et al.* (1996), la calpaína es una proteasa que tiene un papel importante en la proteólisis post-mortem del músculo mediante la degradación de las proteínas miofibrilares. En general, los mecanismos que controlan la ternura de la carne dependen de la interacción compleja de procesos celulares, degradación miofibrilar y actividad de las enzimas incluyendo catepsinas, el sistema calpaína/Calpastatina y la proteosoma (Bendixen, 2005). Con lo cual, se realizó la búsqueda de genes adicionales en la misma región que tuvieran posiblemente un efecto en la ternura de la carne. En primer lugar, surge el gen *CST6* (cystatin E/M, cromosoma 2: 5.395Mb-5.396Mb), el cual codifica para Cistatina M o E/M, proteína que actúa como un inhibidor endógeno de las proteasas de cisteína lisosomal, regulando y protegiendo las células contra la proteólisis no controlada, ejercida por las proteínas de cisteína Catepsina L y V (Turk y Bode, 1991; Cheng *et al.*, 2006; Zeeuwen *et al.*, 2009). Asimismo, la búsqueda permitió identificar el gen *CTSW* (catepsina W, cromosoma 2: 5.550Mb-5.554Mb), el cual codifica para Catepsina W. Se ha reportado que las catepsinas contribuyen a la proteólisis miofibrilar en el músculo *Longissimus dorsi* en cerdos, teniendo un papel importante en la degradación de proteína en el músculo (Ertbjerg *et al.*, 1999). Con lo cual, el gen *CTSW* emerge como gen candidato para SF. Otro de los genes identificados en la búsqueda para SF es el gen *SYVNI* (synovial apoptosis inhibitor 1, synoviolin,

cromosoma 2: 6.19Mb-6.2Mb), el cual codifica para una proteína involucrada en la degradación asociada al retículo endoplasmático y que usa el sistema proteosoma-ubiquitina para regular la proteólisis de proteínas “mal dobladas”. Estas proteínas fueron relacionadas con la hipertrofia de los condrocitos (células encontradas en el cartílago) en ratones, debido a su rol en el mejoramiento de la capacidad del retículo endoplasmático para el plegamiento de proteínas (Kaufman, 1999; Malhotra y Kaufman, 2007; Ron y Walter, 2007; Cameron *et al.*, 2011). De este modo, y aunque el gen *CAPNI* permanece como el gen candidato más relevante para SF en esta región del cromosoma 2, estos resultados sugieren la importancia de estudios adicionales que permitan determinar el papel de genes cercanos al gen *CAPNI* no reportados previamente, que codifican para proteínas relacionadas bioquímicamente y que pueden ser genes candidatos para SF.

En la segunda región asociada con SF y ubicada en el cromosoma 2 entre las posiciones 105.02Mb y 109.6Mb, el gen *CAST* (Calpastatina; cromosoma 2:106.9Mb-107.1Mb) es un gen candidato, considerando que los marcadores en el gen *CAST* fueron asociados con terneza de la carne en cerdos (Ciobanu *et al.*, 2004; Meyers y Beever, 2008; Lindholm-Perry *et al.*, 2009) y bovinos (Schenkel *et al.*, 2006; McClure *et al.*, 2012; Tait *et al.*, 2014). La Calpastatina es un inhibidor endógeno de μ -y *m*-calpaína, regulando la actividad de la calpaína post-mortem en el musculo (Koochmaraie, 1992). En cerdos, Lonergan *et al.* (2001) reportaron la asociación entre la elevada actividad de la Calpastatina y la degradación de las proteínas del musculo. Asimismo, Ciobanu *et al.* (2004) identificaron diferentes variantes del gen *CAST* en cerdos, reportando efectos significativos en la terneza y otros caracteres de calidad de carne de relevancia económica. Con lo cual, estos resultados replican reportes previos, aportando evidencia adicional de la asociación entre marcadores cercanos al gen *CAST* y la terneza en la carne de cerdo.

En la región asociada significativamente con IMF y CKL (cromosoma 6, 0-1Mb), el gen *ACSF3* (Acyl-CoA synthetase family member 3 mitochondrial; cromosoma 6: 0.82Mb-0.84Mb) es un gen candidato ubicado cerca del pico de asociación observado para dichos caracteres (0.013Mb). El gen *ACSF3*, que pertenece a la familia de enzimas que activan los ácidos grasos (Bovo *et al.*, 2015), presentó sobre-expresión en el musculo *Semimembranosus* en cerdos y estuvo relacionado con la vía de beta-oxidación de los ácidos grasos a nivel mitocondrial (Herault *et al.*, 2014). En relación al carácter CKL y a nuestro conocimiento, no se tienen resultados previos que soporten la asociación posible entre el gen *ACSF3* y dicho carácter. Sin embargo, Aaslyng *et al.* (2003) reportaron una relación entre IMF y CKL, donde carne con alto contenido de IMF tuvo bajo CKL a temperaturas de 60°C y 70°C (temperatura en el centro del corte evaluado), mientras que a 80°C, la carne con el contenido más bajo de IMF tuvo un CKL significativamente mayor. De este modo, los resultados obtenidos en el presente estudio para esta región en el cromosoma 6 son relevantes en diferentes aspectos: primero, la identificación del gen *ACSF3* dentro del intervalo genómico asociado significativamente con IMF provee soporte adicional en cuanto a la relación entre el gen *ACSF3* y el carácter IMF. Además, sirven como precedente de la posible asociación entre el gen *ACSF3* y el carácter CKL, la cual no ha sido reportada en estudios previos. Asimismo, la detección de un QTL significativo común para IMF y

CKL (es decir, se identificó el mismo SNP asociado significativamente con ambos caracteres en el MA-GWAS), reafirma la relación biológica entre IMF y CKL, caracteres relevantes en la calidad de carne en cerdos.

Para la medición de color rojo en la carne CIE a* se identificaron dos genes candidatos. Primero, el gen *GYS1* (glycogen synthase 1-muscle; cromosoma 6: 50.07Mb-50.09Mb), el cual codifica para la enzima glucógeno-sintasa en músculo esquelético (Wang *et al.*, 2012), enzima que está relacionada con el almacenamiento de glucosa. En este caso, la degradación anaeróbica del glucógeno controla la tasa y grado de la caída del pH en el musculo, afectando los atributos de color de la carne, incluyendo las mediciones de color (Scheffler *et al.*, 2011). Adicionalmente, el gen *FTL* (ferritin, light polypeptide; cromosoma 6: 50.09Mb) el cual, junto con el gen *FTH1* (ferritin, heavy polypeptide 1, cromosoma 2: 8,9Mb), integran el complejo Ferritina para el almacenamiento de hierro. El contenido de hierro en la carne de cerdo está altamente correlacionado con el color CIE a* y otras mediciones de color (Hermesch y Jones, 2012). Sin embargo, no se han reportado posibles asociaciones entre los genes *GYS1* y *FTL* con el carácter CIE a*. Por consiguiente, considerando las proteínas codificadas por los genes mencionados así como sus mecanismos de acción biológica, los resultados presentados en este capítulo sirven como precedente de la posible relación en los genes *GYS1* y *FTL* y las mediciones de color rojo en la carne de cerdo.

En el cromosoma 15, se identificó el gen *PRKAG3* (protein kinase, AMP-activated, gamma 3 non-catalytic subunit; cromosoma 15: 133.8Mb), el cual ha sido reportado ampliamente por su efecto en la variación en el contenido de glucógeno, perdidas de líquidos, perdidas por cocción y por consiguiente, en la calidad de la carne en cerdos (Milan *et al.*, 2000; Ciobanu *et al.*, 2001; Otto *et al.*, 2007; Rohrer *et al.*, 2012; Nonneman *et al.*, 2013). Con lo cual, el gen *PRKAG3* es el gen candidato más relevante para pHu, PRL y CKL en cerdos. Dicho gen ha sido asociado con variaciones en el contenido de glucógeno en musculo, así como con la producción de lactato y el pH en el musculo post-mortem y por lo tanto, con la habilidad de la carne para retener agua durante el corte, calentamiento y presión (capacidad de retención de agua o water holding capacity - WHC) en carne de cerdo fresca (Huff-Lonergan y Lonergan, 2005; Ryan *et al.*, 2012). Adicionalmente, existen reportes previos de la relación entre pHu, CKL y PRL. Específicamente, Hamm (1960) observó que la tasa y la magnitud de la caída del pH afectó la proteólisis de proteínas del citoesqueleto y la WHC en carne de cerdo. Además, Huff-Lonergan y Lonergan (2005) reportaron que valores altos de PRL y baja WHC estuvieron relacionados con bajo pHu, disminución que generó desnaturalización de muchas proteínas (perdida de funcionalidad y habilidad para retener agua), incluyendo aquellas involucradas en la retención de agua. En relación a CKL, Aaslyng *et al.* (2003) indicaron que WHC y pH influyeron en las perdidas por cocción, y particularmente, valores bajos de WHC y pH conllevaron a altos valores de CKL. Por lo tanto, considerando la asociación fisiológica entre pHu, PRL y CKL, la detección de una región común para aquellos caracteres en nuestras poblaciones es un resultado esperado. Consecuentemente, la identificación del gen *PRKAG3* sirve como evidencia adicional para la asociación entre dicho gen y caracteres de calidad de carne como pHu, PRL y CKL.

Por lo tanto, la implementación de MA-GWAS permitió dilucidar cuantitativamente los factores y posibles vías metabólicas en caracteres complejos de calidad de carne de cerdo, combinando resultados de GWAS individuales y considerando factores de ponderación sobre la base de la inversa de la varianza de los efectos estimados de SNP.

Capítulo 5
Discusión general

Discusión general

En esta tesis se presentaron métodos alternativos para el análisis de asociación entre marcadores y QTL en regiones genómicas, calculados en función de las predicciones GBLUP. Se describieron primero distintos enfoques para el desarrollo del análisis conjunto de asociación (JA), combinando fenotipos y genotipos provenientes de diferentes poblaciones y considerando distintas especificaciones de la heterogeneidad de varianzas entre poblaciones. Luego se presentó el meta-análisis (MA), combinando resultados de GWAS en las poblaciones independientes y empleando dos esquemas alternativos de ponderación de los efectos de los marcadores. Finalmente, se evaluó el potencial de las regiones que habían sido detectadas como significativas en MA para sugerir potenciales genes candidatos de los ERT analizados.

La investigación reafirma la utilidad de imputar genotipos, la importancia de considerar el desequilibrio gamético (LD) y la persistencia de la fase, cuando se desarrollan metodologías de análisis para SG. El valor de la imputación de genotipos está en aumentar la cantidad de información disponible para la SG y los GWAS (Gualdrón Duarte *et al.*, 2013; Chen *et al.*, 2014). Se encontraron valores de LD (en términos de correlación promedio) entre 0.19 a 0.25 para distancias entre marcadores de 500 Kb. Dichas estimaciones son similares a las reportadas para diferentes razas de cerdos por Badke *et al.* (2012). Su relevancia para la SG se debe a que Meuwissen *et al.* (2001) estimaron que el LD promedio para marcadores adyacentes debía ser igual o mayor a 0.2 para que la selección sea mínimamente eficiente. Fue observado un r^2 mayor en la población MSURP en comparación con las otras dos en todo el rango de distancias analizado. Fu *et al.* (2015) observaron que un LD tal que $r^2 \geq 0.2$ es compatible con la presencia de un número elevado de SNP asociados significativamente con determinado carácter, afirmación ratificada por la amplitud y número de SNP detectados significativos en la población MSURP en comparación con las otras dos poblaciones (Tabla 4.4.3.1). Las diferencias en los patrones de LD permiten evaluar el grado de relación entre las mismas, así como la confiabilidad en la predicción de los GEBV (de Roos *et al.*, 2008). La mayor persistencia de fase para SNP ubicados a una distancia de 100Kb entre las poblaciones Comercial y MARC (0.93) sugiere que ambos grupos comparten una mayor proporción del genoma. Por su parte, la persistencia de fase de MSURP con la población Comercial fue 0.89, mientras que con MARC fue igual a 0.83. Consecuentemente, es esperable que las estimaciones de los efectos de SNP sean similares para esas dos poblaciones.

Un aspecto que recibió considerable atención fue la evaluación del efecto de la estructura genética poblacional sobre los GWAS realizados, aspecto que de ser ignorado, dará lugar a un número mayor de falsas asociaciones positivas en los GWAS (Pritchard *et al.*, 2000). Así se analizó la incorporación de los componentes principales obtenidos por la descomposición de la matriz \mathbf{G} y se los ajustó como efectos fijos en los modelos de análisis para los GWAS (Figura 2.4.4.1). En adición, se incluyó un efecto aleatorio por animal con matriz de varianzas-covarianzas proporcional a \mathbf{G} . Sin embargo, no se encontraron diferencias entre los p -valores para los efectos de SNP derivados de los modelos que incluían (ecuación 2.7) o excluían (ecuación 2.1) los

autovectores de \mathbf{G} (Figura 2.4.4.2). Los resultados son coincidentes con los informados por Lans *et al.* (2012), quienes sugirieron que al incluir los autovectores en el modelo para calcular GBLUP se contabiliza información por duplicado, dado que el efecto del autovector es tenido en cuenta por la matriz de relaciones genómicas (Lans *et al.*, 2012). Al respecto, (MacLeod *et al.*, 2010) evaluaron el efecto de incluir u omitir el pedigree en el número de QTL detectados por medio de simulación, reportando un incremento sustancial en el número de falsos positivos cuando el efecto poligénico no fue tenido en cuenta. Es decir, el número de falsas asociaciones positivas fue mayor cuando no se incluyó información de pedigree. En el caso de no disponer de información de pedigree, como en el caso de esta tesis, la matriz de relaciones genómicas \mathbf{G} puede ser ajustada con el fin de considerar la presencia de estructura genética de la población (Lans *et al.*, 2012) y por tanto, reducir el número de falsas asociaciones para determinado carácter. En consecuencia, se realizaron los análisis sin incluir los autovectores de la matriz \mathbf{G} y se mantuvo la ecuación del modelo [2.1] en todos los casos.

Además de la dificultad en el acceso a la información principalmente a partir de poblaciones comerciales, uno de los mayores desafíos al implementar JA (Bravata y Olkin, 2001) es la especificación correcta del modelo. Con GWAS donde los efectos de SNP se calcularon como funciones de los GBLUP, es difícil encontrar la especificación óptima para estimar las relaciones genómicas en la matriz \mathbf{G} . Al considerar las relaciones entre poblaciones es necesario calcular las frecuencias alélicas de la población base. Como tales frecuencias son desconocidas o no se encuentran disponibles, es común utilizar las frecuencias alélicas específicas dentro de cada población. Además, es complicado especificar apropiadamente la heterocedasticidad entre poblaciones. Cuando se modelaron las varianzas específicas de cada población asumiendo una covarianza igual a cero entre poblaciones (Reverter *et al.*, 2004; Möhring y Piepho, 2009) el JA produjo los mismos resultados que los estudios GWAS dentro de cada población (Figura 2.4.6.3). Surge entonces la duda de cómo combinar apropiadamente los estimadores de los efectos SNP entre poblaciones para el JA. El paso siguiente en futuros JA será el desarrollo de ponderaciones para combinar las estimaciones de los efectos de SNP con el fin de obtener un estadístico único que integre la información de todas las poblaciones. En el otro extremo, si se utilizan las frecuencias alélicas obtenidas juntando los datos de distintas poblaciones para estimar las relaciones genómicas entre y dentro grupos, la matriz \mathbf{G} resultante ajusta un solo componente de varianza genómico. En la presente investigación este modelo JA con varianzas homogéneas entre poblaciones no identificó SNP asociados significativamente con CIE a* (Figura 2.4.6.4), carácter elegido para ilustrar la metodología, posiblemente porque esa especificación no explicaba de manera adecuada la estructura genética entre poblaciones.

Cuando los datos originales no se encuentran disponibles, una alternativa al JA empleada en la tesis son los métodos MA-GWA, los que son calculados con evaluaciones genómicas de poblaciones distintas e independientes. Se describió cómo ponderar dichas estimaciones empleando modelos animales centrados GBLUP. Strandén y Garrick (2009) y Gualdrón Duarte *et al.* (2014) mostraron la equivalencia numérica entre las pruebas estadísticas propuestas para efectos de SNP que son función de predicciones GBLUP, con las pruebas realizadas por el software EMMAX (Efficient

Mixed-Model Association eXpedited; Kang *et al.*, 2010; Zhang *et al.*, 2010) que prueba efectos de SNP como fijos e incluye un efecto poligénico aleatorio. La demostración analítica original desarrollada en la sección 3.3.4, demuestra que la prueba de hipótesis empleada en el MA-GWA corresponde a un test para los efectos de SNP fijos de uno a la vez, fácilmente calculable a partir de la varianza de los efectos de SNP aleatorios (expresión [3.23]). El resultado es relevante para el cálculo de ponderaciones apropiadas bajo el enfoque de la inversa de la varianza propuesto en esta tesis, fundamentalmente porque es dicho esquema de ponderación ha sido resaltado como el enfoque óptimo para desarrollar MA-GWAS (Cochran, 1954; Hedges y Olkin, 1985; Lipsey y Wilson, 2001; Zhou *et al.*, 2011). Al igual que en otros ejemplos de uso del MA (Stankowich y Blumstein, 2005; Akanno *et al.*, 2013), en esta tesis se consideraron también ponderaciones que son función del tamaño muestral de cada población. Comparado con el enfoque de la inversa de la varianza, las ponderaciones por tamaño muestral son menos óptimas porque no consideran la exactitud de imputación de genotipos (Jiao *et al.*, 2011), y no tienen en cuenta la heterocedasticidad específica dentro de población (Hedges y Olkin, 1985; Marin-Martinez y Sanchez-Meca, 2010; Akanno *et al.*, 2013). Sin embargo, el empleo del tamaño muestral es de fácil implementación cuando se mide el carácter de interés en escalas distintas en las diferentes poblaciones (Willer *et al.*, 2010). Un aspecto saliente de los análisis MA-GWAS de la tesis es la falta de ajuste por parte de los efectos de SNP específicos dentro de población. A pesar de la complejidad estructural de los estadísticos que contemplan signos opuestos, considerar la existencia de interacción entre SNP \times población en la prueba de hipótesis permitirá incrementar la potencia para identificar asociaciones significativas entre poblaciones. En tal sentido es trascendente para futuros MA-GWAS el considerar que los efectos de SNP pueden variar de acuerdo a la composición genética de la población. Una alternativa es considerar un enfoque de MA en que los efectos de interés son variables aleatorias, hecho que conduce a especificar naturalmente las ponderaciones modelando la heterogeneidad de los efectos de SNP entre poblaciones.

Asimismo, debe considerarse que para que un SNP sea significativo en las pruebas MA dentro de población (z -score), la dirección o signo del test debe ser consistente con los obtenidos en las demás poblaciones; de lo contrario, se cancelarán las pruebas estadísticas de signo opuesto entre poblaciones. Sin embargo, este supuesto es difícil de cumplir cuando se usan plataformas de SNP con densidades medias en poblaciones animales con bajo nivel de LD, dando lugar a una baja persistencia de la fase entre poblaciones (Badke *et al.*, 2012). Una alternativa de análisis podría ser combinar el valor absoluto de las pruebas de hipótesis poblacionales. No obstante, el z -score obtenido en ese caso no seguirá una distribución normal estándar bajo hipótesis nula, sino una combinación lineal de distribuciones normales “dobladas” (Folded Normal distribution; Leone *et al.*, 1961) en número igual al de poblaciones incluidas en el MA. A nuestro conocimiento, tal distribución tiene una forma compleja y la estimación de los parámetros constituye un desafío (Chakraborty y Chatterjee, 2013). No obstante, teniendo en cuenta el interés creciente en la utilidad de los resultados de estudios MA-GWAS en el mejoramiento genético animal (Zhang *et al.*, 2014), los futuros análisis deberán considerar las diferencias en el signo de las pruebas de hipótesis de asociación provenientes de distintas poblaciones.

Tomando como ejemplo las mediciones del color rojo en la carne CIE a*, se compararon los resultados del MA con los obtenidos a partir del JA. En primer lugar, se observó una mayor coincidencia entre JA y el MA calculado con el tamaño muestral ($R^2 = 0.8308$), respecto de JA y MA basado en la inversa de la varianza ($R^2 = 0.7264$). Este resultado es esperable porque, tal como sugirieron Walling *et al.* (2000) y Kim *et al.* (2005), al combinar los datos en JA se favorece a la población con el mayor tamaño muestral. Asimismo, aunque los *p*-valores obtenidos bajo los dos enfoques de ponderación para MA se hallaron altamente correlacionados con los resultados del JA, este último método de análisis produjo pruebas de hipótesis más conservadoras (mayores *p*-valores) que el MA, especialmente para valores extremos del estadístico (Figura 3.4.4.1). Estos resultados son similares a aquellos obtenidos por Walling *et al.* (2000) y Zhou *et al.* (2011), quienes reportaron test más significativos en el MA en comparación con el JA. Zhou *et al.* (2011) observaron que este tipo de diferencias observadas se relacionan con la habilidad del MA para considerar el efecto de la heterogeneidad de varianza sobre la magnitud de los efectos de SNP entre poblaciones. Además de cierta limitación del JA para modelar adecuadamente las fuentes de variación importantes dentro de cada población, implementar JA conlleva a una demanda computacional mayor comparado con MA-GWAS. Esto se debe a la cantidad de información a analizar, en especial al considerar la heterogeneidad de la varianza entre poblaciones. Möhring y Piepho (2009) reportaron problemas de convergencia al emplear JA debido a la complejidad de la estructura genética a analizar, la cantidad de componentes de varianza a estimar, el desbalance en los datos y a la interacción genotipo \times población.

El análisis MA-GWAS fue empleado con datos de caracteres de calidad de carne en cerdos pertenecientes a tres poblaciones distintas. Específicamente, se evaluaron el pH del músculo *longissimus dorsi* 24 horas post-sacrificio (pHu), mediciones objetivas de luminosidad (CIE L*), el color amarillo (CIE b*) y el color rojo (CIE a*), el porcentaje de grasa intramuscular (IMF) y pérdida de fluidos (PRL), y caracteres de procesamiento como el porcentaje de pérdidas por cocción (CKL) y la resistencia al corte (SF). Además de mostrar heredabilidades moderadas (Miar *et al.*, 2014), los caracteres de calidad de carne son importantes a nivel económico debido al aumento en la demanda de una carne de mejor calidad por parte del consumidor, así como de un producto con mayor rendimiento a nivel industrial (Dransfield *et al.*, 2005). En adición, la aplicación de GWAS y especialmente del MA en los datos ERT evidenció el valor de considerar la heterogeneidad de los componentes de varianza al analizar GWAS. Esto se observó en las estimaciones de las varianzas genética y residual, así como en las heredabilidades por carácter dentro de población (Tabla 4.4.2.2). Si bien dichas estimaciones presentaron un rango amplio, fueron generalmente consistentes con los reportadas en la literatura (Gjerlaug-Enger *et al.*, 2010; Li *et al.*, 2010; Miar *et al.*, 2014). En consecuencia, en MA-GWAS se observó mayor potencia para las pruebas de asociación al compararlas con la de GWAS en poblaciones individuales (Figuras 4.4.3.1. a 4.4.3.8). Comparando los resultados del MA con aquellos obtenidos en los GWAS poblacionales, se delimitaron con mayor precisión las regiones para PRL, pHu, SF y CKL en los cromosomas 2, 6 y 15. Estos resultados son similares a los obtenidos por Do *et al.* (2013) quienes observaron una reducción en las regiones significativas en MA, que contrastaron con la longitud de los segmentos significativos obtenidos a partir

de GWAS individuales para consumo de alimentos en cerdos. Asimismo, el MA identificó tres regiones significativas adicionales que no fueron detectadas en los GWAS poblacionales, incluyendo un SNP significativo para SF en el cromosoma 6 y dos QTL para CIE a* en el cromosoma 1 y cromosoma 6 (uno en cada cromosoma). Estos resultados muestran que MA es una metodología de utilidad para detectar un mayor número de SNP asociados significativamente en comparación con los GWAS en poblaciones individuales. La observación es consistente con Bolormaa *et al.* (2014) quienes en ganado de carne realizaron un MA multicarácter para crecimiento, reproducción y producción, identificando un mayor número de SNP significativos con respecto a los resultados de los GWAS poblacionales.

El nivel de significancia en las pruebas de hipótesis múltiples de los GWAS amerita discusión. Es común que en muchos GWAS haya “picos” en el valor del estadístico de prueba de hipótesis que no alcanzan el umbral de significancia, al considerar el genoma completo y ajustar por múltiples pruebas (Hang *et al.*, 2009; Roberts *et al.*, 2010). Una posible explicación es que el criterio de Bonferroni es extremadamente conservador (Hirschhorn y Daly, 2005). Si bien esta afirmación es correcta, existen otras posibles razones asociadas con la no detección de picos aparentes. Por ejemplo, considérese el caso de los GWAS poblacionales para CIE a*. En este ERT, el p -valor más bajo fue igual a 4.415×10^{-6} , con el GWAS para la población MSUPRP. Para que dicho p -valor sea significativo a un nivel de $p < 0.05$ a nivel de todo el genoma, el número efectivo de pruebas de hipótesis (es decir, de SNP) no debería ser mayor a 11325. Si la información en los SNP incluidos en el chip de 60K es igual a la disponible en un chip de 11K, solo estos SNP deberían ser considerados para la significancia a nivel del genoma. En esencia, el problema es similar al de los grados de libertad al ajustar efectos fijos en la estimación de componentes de varianza empleando REML o ML. Sin embargo, se ha reportado que un chip de 11K no permite obtener una exactitud de imputación del 100% respecto del chip de 60K (Weigel *et al.*, 2010; Badke *et al.*, 2012), lo que implica la existencia de un número mayor de marcadores efectivos que 11K. Consecuentemente, el número específico de SNP utilizado para implementar la corrección de Bonferroni no sería el principal factor de la potencia limitada de los GWAS bajo tamaños muestrales modestos (Wang *et al.*, 2005) pero que son frecuentemente utilizados en poblaciones animales (Zhang *et al.*, 2013; Finno *et al.*, 2014) posiblemente debido al costo de mantenimiento y genotipación de poblaciones grandes. No obstante, es conveniente que las asociaciones encontradas en futuros GWAS y MA-GWAS sean evaluadas empleando métodos alternativos, tales como la tasa de “descubrimientos falsos” (FDR, por sus siglas en inglés; False Discovery Rate), variantes recientemente desarrolladas de la corrección de Bonferroni (Fu *et al.*, 2014) o estadísticos alternativos tales como la proporción de falsos positivos (Fernando *et al.*, 2004).

El LD fue evaluado entre y dentro de población y se identificaron bloques de LD para buscar genes candidatos asociados con los ERT empleando el enfoque de Do *et al.* (2014). Las regiones evaluadas de manera significativa en el MA-GWAS para caracteres de calidad de carne, fueron refinadas por la información de los bloques de LD y se identificaron genes candidatos para SF en el cromosoma 2 (*CAPNI*, *CAST*) y para pHu, CKL y PRL en el cromosoma 15 (*PRKAG3*), asociaciones que fueron informadas

previamente (Milan, 2000; Ciobanu *et al.*, 2001; Ciobanu *et al.*, 2004; Meyers y Beever, 2008; Nonneman *et al.*, 2011; Nonneman *et al.*, 2013). Luego de considerar la cercanía de las asociaciones detectadas con el MA-GWA, los mecanismos biológicos y las vías metabólicas, se identificaron genes potencialmente candidatos para SF (*CST6*, *CTSW*, *SYVNI*), IMF y CKL (*ACSF3*) y para CIE a* (*GYS1* y *FTL*), que no habían sido reportados previamente. Estudios futuros confirmarán o rectificarán estas observaciones.

Finalmente, y si bien los antecedentes del empleo de teoría Bayesiana en MA son escasos (Han y Eskin, 2012), dichos métodos en modelos lineales incorporan naturalmente la información previa de los efectos de SNP para especificar distribuciones a priori conjugadas, tratando los efectos de SNP como variables aleatorias, hecho que puede ser relevante al momento de analizar las grandes cantidades de información producidas por los métodos de secuenciación de “próxima generación” (Next Generation Sequencing Methods). Asimismo, aunque existe una gran disponibilidad de programas para analizar MA (MetaQTL, Veyrieras *et al.*, 2007; MetABEL, Aulchenko *et al.*, 2007; METAL, Willer *et al.*, 2010; GWAMA, Mägi y Morris, 2010), existe potencial para desarrollar programas de utilidad en el ámbito del mejoramiento genético animal. Todo el software desarrollado en esta tesis, junto con los datos ilustrativos y el código en lenguaje R (RDC Team, 2013), para implementar el MA-GWAS se encuentran disponibles en <http://tinyurl.com/BLUPMA>. Es deseable resaltar la importancia de hacer disponible toda la información analizada en los GWAS poblacionales para facilitar el desarrollo de futuros meta-análisis que involucren los caracteres de interés aquí empleados.

Conclusiones

Conclusiones

Sobre la base de toda la investigación realizada, las principales conclusiones de esta tesis pueden ser sintetizadas como sigue:

1. Dada la amplia utilización de estudios GWA en el mejoramiento genético animal, en esta tesis se presentaron los métodos para el desarrollo de dicha metodología a nivel poblacional considerando modelos de predicción GBLUP, y resaltando la importancia de modelar la heterogeneidad de los componentes de varianza y la estructura genética dentro de población.
2. Con el fin de incrementar la potencia en la detección de asociaciones significativas, esta tesis presentó diferentes enfoques para el desarrollo de análisis conjuntos de asociación (JA), considerando estructuras de (co)varianza que permitan modelar la heterocedasticidad de los componentes genético y residual entre poblaciones.
3. De modo original, esta tesis presentó contribuciones teóricas y metodológicas sustanciales para la implementación de meta-análisis de GWAS (MA-GWAS), teniendo en cuenta la definición de ponderantes óptimos a partir de test obtenidos de modelos que consideran efectos de SNP como fijos, y considerando las diferencias en los componentes de varianza entre poblaciones.
4. Debido al aumento en la potencia y precisión del MA-GWAS en comparación con los GWAS poblacionales y el JA, se replicaron QTL reportados en estudios previos, se identificaron nuevas asociaciones significativas y se sugirieron nuevos genes candidatos para caracteres de calidad de carne en cerdo. Al respecto, se requiere investigación adicional para determinar el alcance y mecanismos biológicos de los genes candidatos reportados en esta tesis.

Bibliografía

- Aaslyng, M. D., C. Bejerholm, P. Ertbjerg, H. C. Bertram, and H. J. Andersen. 2003. Cooking loss and juiciness of pork in relation to raw meat quality and cooking procedure. *Food Qual. Prefer.* 14:277–288.
- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–52.
- Akanno, E. C., G. Plastow, B. W. Woodward, S. Bauck, H. Okut, X.-L. Wu, C. Sun, J. L. Aalhus, S. S. Moore, S. P. Miller, Z. Wang, and J. a. Basarab. 2014. Reliability of molecular breeding values for Warner-Bratzler shear force and carcass traits of beef cattle - An independent validation study. *J. Anim. Sci.* 92:2896–2904.
- Akanno, E. C., F. S. Schenkel, V. M. Quinton, R. M. Friendship, and J. A. B. Robinson. 2013. Meta-analysis of genetic parameter estimates for reproduction, growth and carcass traits of pigs in the tropics. *Livest. Sci.* 152:101–113.
- Allison, D. B., and M. Heo. 1998. Meta-analysis of linkage data under worst-case conditions: a demonstration using the human OB region. *Genetics* 148:859–65.
- Anderson, C. A., F. H. Pettersson, J. C. Barrett, J. J. Zhuang, J. Ragoussis, L. R. Cardon, and A. P. Morris. 2008. Evaluating the Effects of Imputation on the Power, Coverage, and Cost Efficiency of Genome-wide SNP Platforms. *Am. J. Hum. Genet.* 83:112–119.

- Aulchenko, Y. S., S. Ripke, A. Isaacs, and C. M. van Duijn. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23:1294–6.
- Badke, Y. M., R. O. Bates, C. W. Ernst, J. Fix, and J. P. Steibel. 2014a. Accuracy of Estimation of Genomic Breeding Values in Pigs Using Low Density Genotypes and Imputation. *G3 (Bethesda)*:1–24.
- Badke, Y. M., R. O. Bates, C. W. Ernst, J. Fix, and J. P. Steibel. 2014b. Accuracy of estimation of genomic breeding values in pigs using low-density genotypes and imputation. *G3 Genes|Genomes|Genetics* 4:623–31.
- Badke, Y. M., R. O. Bates, C. W. Ernst, C. Schwab, J. Fix, C. P. Van Tassell, and J. P. Steibel. 2013. Methods of tagSNP selection and other variables affecting imputation accuracy in swine. *BMC Genet.* 14:8.
- Badke, Y. M., R. O. Bates, C. W. Ernst, C. Schwab, and J. P. Steibel. 2012. Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics* 13:24.
- Begum, F., D. Ghosh, G. C. Tseng, and E. Feingold. 2012. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res.* 40:3777–84.
- Bendall, J. R. 1973. Post mortem changes in muscle. In: G. H. Bourne, editor. *Structure and Function of Muscle*. Academic Press, New York, NY.
- Bendixen, E. 2005. The use of proteomics in meat science. *Meat Sci.* 71:138–149.

- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57:289–300.
- Bermingham, M. L., S. C. Bishop, J. a Woolliams, R. Pong-Wong, a R. Allen, S. H. McBride, J. J. Ryder, D. M. Wright, R. a Skuce, S. W. McDowell, and E. J. Glass. 2014. Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. *Heredity (Edinb)*. 112:543–51.
- Bernal Rubio, Y. L., J. L. Gualdrón Duarte, R. O. Bates, C. W. Ernst, D. Nonneman, G. A. Rohrer, D. A. King, S. D. Shackelford, T. L. Wheeler, R. J. C. Cantet, and J. P. Steibel. 2014. Meta-analysis of genome wide association studies for pork quality traits. In: *Proceedings of 10th World Congress of Genetics Applied to Livestock Production – WCGALP*. Vancouver, Canada. https://asas.org/docs/default-source/wcgalp-proceedings-oral/185_paper_8800_manuscript_1645_0.pdf?sfvrsn=2.
- Boddicker, N. J., A. Bjorkquist, R. R. R. Rowland, J. K. Lunney, J. M. Reecy, and J. C. M. Dekkers. 2014. Genome-wide association and genomic prediction for host response to porcine reproductive and respiratory syndrome virus infection. *Genet. Sel. Evol.* 46:18.
- Boler, D. D., A. C. Dilger, B. S. Bidner, S. N. Carr, J. M. Eggert, J. W. Day, M. Ellis, F. K. McKeith, and J. Killefer. 2010. Ultimate pH explains variation in pork quality traits. *J. Muscle Foods* 21:119–130.

- Bolormaa, S., J. E. Pryce, A. Reverter, Y. Zhang, W. Barendse, K. Kemper, B. Tier, K. Savin, B. J. Hayes, and M. E. Goddard. 2014. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genet.* 10:e1004198.
- Bovo, S., F. Bertolini, G. Schiavo, G. Mazzoni, S. Dall'Olio, and L. Fontanesi. 2015. Reduced representation libraries from DNA pools analysed with next generation semiconductor based-sequencing to identify SNPs in extreme and divergent pigs for back fat thickness. *Int. J. Genomics* Volume 201:Article ID 950737.
- Bravata, D. M., and I. Olkin. 2001. Simple Pooling versus Combining in Meta-Analysis. *Eval. Health Prof.* 24:218–230.
- Briscoe, D., J. C. Stephens, and S. J. O'Brien. 1994. Linkage disequilibrium in admixed populations: applications in gene mapping. *J. Hered.*:59–63.
- Browning, S. R., and B. L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81:1084–97.
- Bush, W. S., and J. H. Moore. 2012. Chapter 11: Genome-wide association studies. *PLoS Comput. Biol.* 8:e1002822.
- Cameron, N. D., P. D. Warriss, S. J. Porter, and M. B. Enser. 1990. Comparison of Duroc and British Landrace pigs for meat and eating quality. *Meat Sci.* 27:227.

- Cameron, T. L., K. M. Bell, L. Tatarczuch, E. J. Mackie, M. H. Rajpar, B. T. McDermott, R. P. Boot-Handford, and J. F. Bateman. 2011. Transcriptional Profiling of chondrodysplasia growth plate cartilage reveals adaptive ER-stress networks that Allow Survival but Disrupt Hypertrophy. *PLoS One* 6:e24600.
- Cannon, J. E., J. B. Morgan, F. K. Mckeith, and G. C. Smith. 1996. Pork chain audit survey: Quantification of pork quality characteristics. *J. Muscle Foods* 7:29–44.
- Casas, E., R. J. Leach, T. a Reinhardt, R. M. Thallman, J. D. Lippolis, G. L. Bennett, and L. a Kuehn. 2013. A genomewide association study identified CYP2J2 as a gene controlling serum vitamin D status in beef cattle. *J. Anim. Sci.* 91:3549–56.
- Cassens, R. G. 2000. Historical perspectives and current aspects of pork meat quality in the USA. *Food Chem.* 69:357–363.
- Chakraborty, A. ., and M. Chaterjee. 2013. On multivariate folded normal distribution. *Indian J. Stat.* 75-B:1–15.
- Chen, C. Y., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2011. Effect of different genomic relationship matrices on accuracy and scale. *Journal of Animal Science*, in press. *J. Anim. Sci.* 89:2673–2679.
- Chen, L., L. Changxi, M. Sargolzaei, and F. S. Schenkel. 2014. Impact of genotype imputation on the performance of GBLUP and Bayesian Methods for genomic prediction. *PLoS One* 9:e101544.

- Cheng, T., K. Hitomi, I. M. van Vlijmen-Willems, G. J. de Jongh, K. Yamamoto, K. Nishi, C. Watts, T. Reinheckel, J. Schalkwijk, and P. L. Zeeuwen. 2006. Cystatin M/E is a high affinity inhibitor of cathepsin V and cathepsin L by a reactive site that is distinct from the legumain-binding site. A novel clue for the role of cystatin M/E in epidermal cornification. *J. Biol. Chem.* 281:15893–15899.
- Cherel, P., J. Pires, J. Glénisson, D. Milan, N. Iannuccelli, F. Hérault, M. Damon, and P. Le Roy. 2011. Joint analysis of quantitative trait loci and major-effect causative mutations affecting meat quality and carcass composition traits in pigs. *BMC Genet.* 12:76.
- Choi, I., J. P. Steibel, R. O. Bates, N. E. Raney, J. M. Rumph, and C. W. Ernst. 2011. Identification of Carcass and Meat Quality QTL in an F(2) Duroc × Pietrain Pig Resource Population Using Different Least-Squares Analysis Models. *Front. Genet.* 2:18.
- CIE International. 1976. International Commission on Illumination, Colorimetry: Official Recommendations of the International Commission on Illumination.
- Ciobanu, D., J. Bastiaansen, M. Malek, J. Helm, J. Woollard, G. Plastow, and M. Rothschild. 2001. Evidence for new alleles in the protein kinase adenosine monophosphate-activated gamma(3)-subunit gene associated with low glycogen content in pig skeletal muscle and improved meat quality. *Genetics* 159:1151–62.

- Ciobanu, D. C., J. W. Bastiaansen, S. M. Lonergan, H. Thomsen, J. C. Dekkers, G. S. Plastow, and M. F. Rothschild. 2004. New alleles in calpastatin gene are associated with meat quality traits in pigs. *J. Anim. Sci.* 82:2829–2839.
- Cleveland, W. S., E. Grosse, and W. M. Shyu. 1992. Local regression models. In: J. M. Chambers and T. J. Hastie, editors. *Statistical Models in S*. Pacific Grove: Wadsworth & Brooks/Cole. p. 309–376.
- Cochran, W. G. 1954. The combination of estimated from different experiments. *Biometrics* 10:101–129.
- Cordell, H. J., and D. G. Clayton. 2005. Genetic association studies. *Lancet* 366:1121–1131.
- Crow, J. F., and M. Kimura. 1970. *An Introduction to Population Genetics Theory*. (H. and Row, editor.). New York, USA.
- Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185:1021–1031.
- Dassonneville, R., R. F. Brondum, T. Druet, S. Fritz, F. Guillaume, B. Guldbbrandtsen, M. S. Lund, V. Ducrocq, and G. Su. 2011. Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. *J. Dairy Sci.* 94:3679–3686.

- Daszkiewicz, T., S. Wajda, and T. Bąk. 2003. Intermuscular fat and eating quality of meat. *Gospod. Mięsna* 2:26–29.
- Dekkers, J. C. . 2004. Commercial application of marker-and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82:E313–E328.
- Do, D. N., T. Ostensen, A. B. Strathe, T. Mark, J. Jensen, and H. N. Kadarmideen. 2014. Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. *BMC Genet.* 15:27.
- Do, D. N., A. B. Strathe, T. Ostensen, J. Jensen, T. Mark, and H. N. Kadarmideen. 2013. Genome-wide association study reveals genetic architecture of eating behavior in pigs and its implications for humans obesity by comparative mapping. *PLoS One* 8:e71509.
- Dransfield, E., T. M. Ngapo, N. A. Nielsen, L. Bredahl, P. O. Sjöden, M. Magnusson, M. M. Campo, and G. R. Nutee. 2005. Consumer choice and suggested price for pork as influenced by its appearance, taste and information concerning country of origin and organic pig production. *Meat Sci.* 69:61–70.
- Du, F.-X., A. C. Clutter, and M. M. Lohuis. 2007. Characterizing linkage disequilibrium in pig populations. *Int. J. Biol. Sci.* 3:166–78.
- Dube, B., S. D. Mulugeta, and K. Dzama. 2013. Integrating economic parameters into genetic selection for Large White pigs. *Animal* 7:1231–1238.

- Edwards, D. B., R. O. Bates, and W. N. Osburn. 2003. Evaluation of Duroc- vs. Pietrain-sired pigs for carcass and meat quality measures. *J. Anim. Sci.* 81:1895–1899.
- Edwards, D. B., C. W. Ernst, N. E. Raney, M. E. Doumit, M. D. Hoge, and R. O. Bates. 2008. Quantitative trait locus mapping in an F2 Duroc x Pietrain resource population: II. Carcass and meat quality traits. *J. Anim. Sci.* 86:254–66.
- Ertbjerg, P., P. Henckel, A. Karlsson, L. M. Larsen, and A. J. Møller. 1999. Combined effect of epinephrine and exercise on calpain/calpastatin and cathepsin B and L activity in porcine Longissimus Muscle. *J. Anim. Sci.* 77:2428–2436.
- Evangelou, E., and J. P. A. Ioannidis. 2013. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* 14:379–89.
- Fan, B., S. K. Onteru, Z. Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *PLoS One* 6:e14726.
- Fernandez, X., G. Monin, A. Talmant, J. Mourot, and B. Lebret. 1999. Influence of intramuscular fat content on the quality of pig meat – 1. Composition of the lipid fraction and sensory characteristics of m. longissimus lumborum. *Meat Sci.* 53:59–65.
- Fernando, R. L., D. Nettleton, B. R. Southey, J. C. M. Dekkers, M. F. Rothschild, and M. Soller. 2004. Controlling the proportion of false positives in multiple dependent tests. *Genetics* 166:611–619.

- Finno, C. J., M. Aleman, R. J. Higgins, J. E. Madigan, and D. L. Bannasch. 2014. Risk of false positive genetic associations in complex traits with underlying population structure: A case study. *Vet. J.* 202:543–549.
- Fleiss, J. 1993. The statistical basis of meta-analysis. *Stat. Methods Med. Res.* 2:121–145.
- Fontanesi, L., E. Scotti, L. Buttazzoni, S. Dall’Olio, A. Bagnato, D. P. Lo Fiego, R. Davoli, and V. Russo. 2010. Confirmed association between a single nucleotide polymorphism in the FTO gene and obesity-related traits in heavy pigs. *Mol. Biol. Rep.* 37:461–466.
- Fu, G., G. Saunders, and J. Stevens. 2014. Holm multiple correction for large-scale gene-shape association mapping. *BMC Genet.* 15:S5.
- Fu, W., J. C. Dekkers, W. R. Lee, and B. Abasht. 2015. Linkage disequilibrium in crossbred and pure line chickens. *Genet. Sel. Evol.* 47:1–12.
- Gao, X., L. C. Becker, D. M. Becker, J. D. Starmer, and M. a Province. 2010. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet. Epidemiol.* 34:100–5.
- García-Gámez, E., B. Gutiérrez-Gil, G. Sahana, J.-P. Sánchez, Y. Bayón, and J.-J. Arranz. 2012. GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the LALBA gene. *PLoS One* 7:e47782.

- Gianola, D., G. de los Campos, O. González-Recio, N. Long, H. Okut, G. J. M. Rosa, K. . Weigel, and X.-L. Wu. 2010. Statistical learning methods for genome-based analysis of quantitative traits. In: World Genetic Congress Applied to Livestock Production. Leipzig, Germany.
- Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. *Genetics* 183:347–63.
- Gjerlaug-Enger, E., L. Aass, J. Ødegard, and O. Vangen. 2010. Genetic parameters of meat quality traits in two pig breeds measured by rapid methods. *Animal* 4:1832–1843.
- Glass, G. V. 1976. Primary, Secondary, and Meta-Analysis of Research. *Educ. Res.* 5:3–8.
- Goddard, M. E., and B. J. Hayes. 2007. Genomic selection. *J. Anim. Breed. Genet.* 124:323–30.
- Goddard, M., B. Hayes, H. McPartlan, and A. J. Chamberlain. 2006. Can the same genetic markers be used in multiple breeds? In: Proceedings of the 8th World Congress on Genetics Applied to Livestock Production. Belo Horizonte, Brazil. p. 16–22.
- Golden, B., D. Garrick, S. Newman, and R. Enns. 2000. Economically relevant traits: A framework for the next generation of EPDs. In: Proceedings of the 32nd Research Symposium and Annual Meeting of the Beef Improvement Federation.

- Gondret, F., P. Hernandez, H. Remignon, and S. Combes. 2009. Skeletal muscle adaptations and biomechanical properties of tendons in response to jump exercise in rabbits. *J. Anim. Sci.* 87:544–553.
- González-Recio, O., D. Gianola, N. Long, K. . Weigel, G. J. M. Rosa, and S. Avendaño. 2008. Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in Broilers. *Genetics* 178:2305–2313.
- Gualdrón Duarte, J. L., R. O. Bates, C. W. Ernst, N. E. Raney, R. J. C. Cantet, and J. P. Steibel. 2013. Genotype imputation accuracy in a F2 pig population using high density and low density SNP panels. *BMC Genet.* 14:38.
- Gualdrón Duarte, J. L., R. J. C. Cantet, R. O. Bates, C. W. Ernst, N. E. Raney, and J. P. Steibel. 2014. Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. *BMC Bioinformatics* 15:246.
- Guo, G., F. Zhao, Y. Wang, Y. Zhang, L. Du, and G. Su. 2014. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* 15:30.
- Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2009. Genomic selection using low-density marker panels. *Genetics* 182:343–53.
- Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12.

- Habier, D., J. Tetens, F.-R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet. Sel. Evol.* 42:5.
- Hamm, R. 1960. *Advances in Food Research*. Academic Press, New York and London.
- Han, B., and E. Eskin. 2012. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* 8:e1002555.
- Hang, B., H. M. Kang, and E. Eskin. 2009. Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. *PLoS One* 7(2):e3054.
- Hayes, B., and M. Goddard. 2010. Genome-wide association and genomic selection in animal breeding. *Genome* 53:876–83.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, K. Savin, C. P. van Tassell, T. S. Sonstegard, and M. E. Goddard. 2009. A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PLoS One* 4:e6676.
- Hayes, B. J., P. M. Visscher, H. C. McPartlan, and M. E. Goddard. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.* 13:635–643.
- Hedges, L. V., and I. Olkin. 1985. *Statistical methods for meta-analysis*. New York: Academic Press.

- Hedges, L. V., and J. L. Vevea. 1998. Fixed and random effects models in meta-analysis. *Psychol. Methods* 3:486–504.
- Hedges, L. V. 1983. A random effects model for effect sizes. *Psychol. Bull.* 93:388–395.
- Herault, F., A. Vincent, O. Dameron, P. Le Roy, P. Cherel, and M. Damon. 2014. The Longissimus and Semimembranosus muscles display marked differences in their gene expression profiles in pig. *PLoS One* 9:e100113.
- Hermesch, S., and R. M. Jones. 2012. Genetic parameters for haemoglobin levels in pigs and iron content in pork. *Animal* 6:1904–1912.
- Hermesch, S., B. G. Luxford, and H. U. Graser. 2000. Genetic parameters for lean meat yield, meat quality, reproduction and feed efficiency traits for Australian pigs. 2. Genetic relationships between production, carcass and meat quality traits. *Livest. Prod. Sci.* 65:249–259.
- Hernandez, P., and A. Dalle Zotte. 2010. Influence of diet on rabbit meat quality. In: C. de Blas and J. Wiseman, editors. *Nutrition of the rabbit*. p. 163–178.
- Hickey, J. M., J. Crossa, R. Babu, and G. de los Campos. 2012. Factors Affecting the Accuracy of Genotype Imputation in Populations from Several Maize Breeding Programs. *Crop Sci.* 52:654.
- Hirschhorn, J. N., and M. J. Daly. 2005. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* 6:95–108.

Houlston, R. S., J. Cheadle, S. E. Dobbins, A. Tenesa, A. M. Jones, K. Howarth, S. L. Spain, P. Broderick, E. Domingo, S. Farrington, J. G. D. Prendergast, A. M. Pittman, E. Theodoratou, C. G. Smith, B. Olver, A. Walther, R. a Barnetson, M. Churchman, E. E. M. Jaeger, S. Penegar, E. Barclay, L. Martin, M. Gorman, R. Mager, E. Johnstone, R. Midgley, I. Niittymäki, S. Tuupanen, J. Colley, S. Idziaszczyk, H. J. W. Thomas, A. M. Lucassen, D. G. R. Evans, E. R. Maher, T. Maughan, A. Dimas, E. Dermitzakis, J.-B. Cazier, L. a Aaltonen, P. Pharoah, D. J. Kerr, L. G. Carvajal-Carmona, H. Campbell, M. G. Dunlop, and I. P. M. Tomlinson. 2010. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.* 42:973–7.

Huff-Lonergan, E., and S. M. Lonergan. 2005. Mechanisms of water-holding capacity of meat: The role of postmortem biochemical and structural changes. *Meat Sci.* 71:194–204.

Huff-Lonergan, E., T. Mitsuhashi, D. D. Beekman, F. C. Parrish, D. G. Olson, and R. M. Robson. 1996. Proteolysis of specific muscle structural proteins by μ -calpain at low pH and temperature is similar to degradation in postmortem bovine muscle. *J. Anim. Sci.* 74:993–1008.

Hunter, L. V., and F. L. Schmidt. 1990. *Methods of meta-analysis: Correcting error and bias in research findings.*

- Jiao, S., L. Hsu, C. M. Hutter, and U. Peters. 2011. The use of imputed values in the meta-analysis of genome-wide association studies. *Genet. Epidemiol.* 35:597–605.
- Jung, E. J., H. B. Park, J. B. Lee, C. K. Yoo, B. M. Kim, H. I. Kim, B. W. Kim, and H. T. Lim. 2014. Genome-wide association analysis identifies quantitative trait loci for growth in a Landrace purebred population. *Anim. Genet.* 45:442–4.
- Kang, H. M., J. H. Sul, S. K. Service, N. a Zaitlen, S.-Y. Kong, N. B. Freimer, C. Sabatti, and E. Eskin. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42:348–54.
- Karisa, B. K., J. Thomson, Z. Wang, P. Stothard, S. S. Moore, and G. S. Plastow. 2013. Candidate genes and single nucleotide polymorphisms associated with variation in residual feed intake in beef cattle. *J. Anim. Sci.* 91:3502–3513.
- Kauffman, R. G., W. Sybesma, and G. Eikelenboom. 1990. In Search of Quality. *J. Can. Inst. Food Sci. Technol.* 23:160–164.
- Kaufman, R. J. 1999. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. *Genes Dev.* 13:1211–1233.
- Khatkar, M. S., P. C. Thomson, I. Tammen, and H. W. Raadsma. 2004. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.* 36:163–190.

- Kim, J., M. F. Rothschild, J. Beever, S. Rodriguez-Zas, and J. C. M. Dekkers. 2005. Joint analysis of two breed cross populations in pigs to improve detection and characterization of quantitative trait loci. *J. Anim. Sci.* 83:1229–1240.
- King, D. A., S. D. Shackelford, and T. L. Wheeler. 2011. Use of visible and near-infrared spectroscopy to predict pork longissimus lean color stability. *J. Anim. Sci.* 89:4195–4206.
- Koohmaraie, M. 1992. Effect of pH, temperature, and inhibitors on autolysis and catalytic activity of bovine skeletal muscle mu-calpain. *J. Anim. Sci.* 70:3071–3080.
- Laird, N. M., and C. Lange. 2011. Population Substructure in Association Studies. In: *The Fundamentals of Modern Statistical Genetics. Statistics for Biology and Health* . p. 125–137.
- Lam, A. C., J. Powell, W.-H. Wei, D. J. de Koning, and C. S. Haley. 2009. A combined strategy for quantitative trait loci detection by genome-wide association. *BMC Proc.* 3:S6.
- Lander, E., and L. Kruglyak. 1995. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat. Genet.* 11:241–247.
- Lans, L., G. de Los Campos, N. Sheehan, and D. Sorensen. 2012. Inferences from genomic models in stratified populations. *Genetics* 192:693–704.

- Leone, F. C., L. S. Nelson, and R. B. Nottingham. 1961. The Folded Normal Distribution. *Technometrics* 3:543–550.
- Li, H. D., M. S. Lund, O. F. Christensen, V. R. Gregersen, P. Henckel, and C. Bendixen. 2010. Quantitative trait loci analysis of swine meat quality traits. *J. Anim. Sci.* 88:2904–12.
- Lindholm-Perry, A. K., G. A. Rohrer, J. W. Holl, S. D. Shackelford, T. L. Wheeler, M. Koohmaraie, and D. J. Nissemann. 2009. Relationships among calpastatin single nucleotide polymorphisms, calpastatin expression and tenderness in pork longissimus. *Anim. Genet.* 40:713–21.
- Lipsey, M. W., and D. B. Wilson. 2001. *Practical Meta-Analysis (Applied Social Research Methods Series, Vol 49)*.
- Liu, G., D. G. J. Jennen, E. Tholen, H. Juengst, T. Kleinwächter, M. Hölker, D. Tesfaye, G. Un, H.-J. Schreinemachers, E. Murani, S. Ponsuksili, J.-J. Kim, K. Schellander, and K. Wimmers. 2007. A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. *Anim. Genet.* 38:241–52.
- Liu, R., Y. Sun, G. Zhao, F. Wang, D. Wu, M. Zheng, J. Chen, L. Zhang, Y. Hu, and J. Wen. 2013. Genome-wide association study identifies loci and candidate genes for body composition and meat quality traits in Beijing-You chickens. *PLoS One* 8:e61172.

- Lo, L. L., D. G. McLaren, F. K. McKeith, F. L. Fernando, and J. Novakofski. 1992. Genetic analysis of growth, real time ultrasound, carcass and pork quality traits in Duroc and Landrace pigs: II. Heritabilities and correlations. *J. Anim. Sci.* 70:2357–2396.
- Lonergan, S. M., E. Huff-Lonergan, L. J. Rowe, D. L. Kuhlbers, and S. B. Jungst. 2001. Selection for lean growth efficiency in Duroc pigs: Influence on pork quality. *J. Anim. Sci.* 79:2075–2085.
- De Los Campos, G., D. Gianola, and D. B. Allison. 2010. Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nat. Rev. Genet.* 11:880–886.
- De Los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193:327–45.
- De los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel, and J. M. Cotes. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182:375–385.
- Ma, J., J. Yang, L. Zhou, Z. Zhang, H. Ma, X. Xie, F. Zhang, X. Xiong, L. Cui, H. Yang, X. Liu, Y. Duan, S. Xiao, H. Ai, J. Ren, and L. Huang. 2013. Genome-wide association study of meat quality traits in a White Duroc×Erhualian F2 intercross and Chinese Sutai pigs. *PLoS One* 8:e64047.

- MacLeod, I. M., B. J. Hayes, K. W. Savin, A. J. Chamberlain, H. C. McPartlan, and M. E. Goddard. 2010. Power of a genome scan to detect and locate quantitative trait loci in cattle using dense single nucleotide polymorphisms. *J. Anim. Breed. Genet.* 127:133–142.
- Mägi, R., and A. P. Morris. 2010. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* 11:288.
- Malhotra, J. D., and R. J. Kaufman. 2007. The endoplasmic reticulum and the unfolded protein response. *Semin. Cell Dev. Biol.* 18:716–731.
- Marchini, J., and B. Howie. 2010. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 11:499–511.
- Marin-Martinez, F., and J. Sanchez-Meca. 2010. Weighting by Inverse Variance or by Sample Size in Random-Effects Meta-Analysis. *Educ. Psychol. Meas.* 70:56–73.
- Maxa, J., M. Neuditschko, I. Russ, M. Förster, and I. Medugorac. 2012. Genome-wide association mapping of milk production traits in Braunvieh cattle. *J. Dairy Sci.* 95:5357–64.
- McClure, M. C., H. Ramey, M. M. Rolf, S. D. McKay, J. E. Decker, R. H. Chapple, J. W. Kim, T. M. Taxis, R. L. Weaver, R. D. Schnabel, and J. F. Taylor. 2012. Genome-wide association analysis for quantitative trait loci influencing Warner–Bratzler shear force in five taurine cattle breeds. *Anim. Genet.* 43:662–673.
- Meisinger, D. 2002. A system for assuring pork quality. Des Moines, IA.

- Meuwissen, T. H. E., and M. E. Goddard. 1996. The use of marker haplotypes in animal breeding schemes. *Genet. Sel. Evol.* 28:161–176.
- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–29.
- Meuwissen, T., B. Hayes, and M. Goddard. 2013. Accelerating Improvement of Livestock with Genomic Selection. *Annu. Rev. Anim. Biosci.* 1:221–237.
- Meyers, S. N., and J. E. Beever. 2008. Investigating the genetic basis of pork tenderness: genomic analysis of porcine CAST. *Anim. Genet.* 39:531–543.
- Miar, Y., G. S. Plastow, S. S. Moore, G. Manafiazar, P. Charagu, R. A. Kemp, B. Van Haandel, A. E. Huisman, C. Y. . Zhang, R. M. McKay, H. L. . Bruce, and Z. Wang. 2014. Genetic and phenotypic parameters for carcass and meat quality traits in commercial crossbred pigs. *J. Anim. Sci.* 92:2869–2884.
- Milan, D., J.-T. Jeon, C. Looft, V. Amarger, A. Robic, M. Thelander, C. Rogel-Gaillard, S. Paul, N. Iannuccelli, L. Rask, H. Ronne, K. Lundström, N. Reinsch, J. Gellin, E. Kalm, P. Le Roy, P. Chardon, and L. Andersson. 2000. A Mutation in PRKAG3 Associated with Excess Glycogen Content in Pig Skeletal Muscle. *Science* (80-.). 288:1248–1251.
- Minozzi, G., J. L. Williams, A. Stella, F. Strozzi, M. Luini, M. L. Settles, J. F. Taylor, R. H. Whitlock, R. Zanella, and H. L. Neiberghs. 2012. Meta-analysis of two genome-wide association studies of bovine paratuberculosis. *PLoS One* 7:e32578.

- Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92:4648–55.
- Möhring, J., and H.-P. Piepho. 2009. Comparison of Weighting in Two-Stage Analysis of Plant Breeding Trials. *Crop Sci.* 49:1977.
- Muñoz, G., C. Ovilo, L. Silió, A. Tomás, J. L. Noguera, and M. C. Rodríguez. 2009. Single- and joint-population analyses of two experimental pig crosses to confirm quantitative trait loci on *Sus scrofa* chromosome 6 and leptin receptor effects on fatness and growth traits. *J. Anim. Sci.* 87:459–68.
- Neves, J. V., J. M. Wilson, H. Kuhl, R. Reinhardt, F. C. Castro, and P. N. S. Rodrigues. 2011. Natural history of SLC11 genes in vertebrates: tales from the fish world. *BMC Evol. Biol.* 11:106.
- Nollet, L. M. L. 2012. *Handbook of Meat, Poultry and Seafood Quality*. Second. John Wiley & Sons, Inc.
- Nonneman, D. J., S. D. Shackelford, D. A. King, T. L. Wheeler, R. T. Wiedmann, W. M. Snelling, and G. A. Rohrer. 2013. Genome-wide association of meat quality traits and tenderness in swine. *J. Anim. Sci.* 91:4043–4050.
- Nonneman, D., a K. Lindholm-Perry, S. D. Shackelford, D. a King, T. L. Wheeler, G. a Rohrer, C. D. Bierman, J. F. Schneider, R. K. Miller, H. Zerby, and S. J. Moeller. 2011. Predictive markers in calpastatin for tenderness in commercial pig populations. *J. Anim. Sci.* 89:2663–72.

- Nordmann, A. J., B. Kasenda, and M. Briel. 2012. Meta-analyses: what they can and cannot do. *Swiss Med. Wkly.* 142:w13518.
- Otto, G., R. Roehe, H. Looft, L. Thoelking, P. W. Knap, M. F. Rothschild, G. S. Plastow, and E. Kalm. 2007. Associations of DNA markers with meat quality traits in pigs with emphasis on drip loss. *Meat Sci.* 75:185–95.
- Ouhayoun, J., and A. Dalle Zotte. 1993. Muscular energy metabolism and related traits in rabbit. In: *Vth Hungarian Rabbit Science Symposium, Kaposvar.*
- Ovilo, C., A. Clopt, J. L. Noguera, M. A. Oliver, C. Barragán, C. Rodríguez, L. Silio, M. A. Toro, A. Coll, J. M. Folch, A. Sánchez, D. Babot, L. Varona, and M. Pérez-Enciso. 2002. Quantitative trait locus mapping for meat quality traits in an Iberian × Landrace F2 pig population. *J. Anim. Sci.* 80:2801–2808.
- Page, B. T., E. Casas, R. L. Quaas, R. M. Thallman, T. L. Wheeler, S. D. Shackelford, M. Koohmaraie, S. N. White, G. L. Bennett, J. W. Keele, M. E. Dikeman, and T. P. Smith. 2004. Association of markers in the bovine CAPN1 gene with meat tenderness in large crossbred populations that sample influential industry sires. *J. Anim. Sci.* 82:3474–3481.
- Paszek, A. A., P. J. Wilkie, G. H. Flickinger, L. M. Miller, C. F. Louis, G. A. Rohrer, L. J. Alexander, C. W. Beattie, and L. B. Schook. 2001. Interval mapping of carcass and meat quality traits in a divergent swine cross. *Anim. Biotechnol.* 12:155–165.
- Pérez-Enciso, M., A. Mercadé, J. P. Bidanel, H. Geldermann, S. Cepica, H. Bartenschlager, L. Varona, D. Milan, and J. M. Folch. 2005. Large-scale,

multibreed, multitrait analyses of quantitative trait loci experiments: The case of porcine X chromosome. *J. Anim. Sci.* 83:2289–2296.

Porto Neto, L. R., E. K. Piper, N. N. Jonsson, W. Barendse, and C. Gondro. 2010. Meta-analysis of Genome Wide Association and Gene Expression Studies to Identify Candidate Genes for Tick Burden in Cattle. In: 9th World Congress of Genetics Applied to Livestock Production. Leipzig, Germany. p. p. 664.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38:904–909.

Price, A. L., N. a Zaitlen, D. Reich, and N. Patterson. 2010. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* 11:459–63.

Pritchard, J. K., and P. Donnelly. 2001. Case-control studies of association in structured or admixed populations. *Theor. Popul. Biol.* 60:227–37.

Pritchard, J. K., and M. Przeworski. 2001. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* 69:1–14.

Pritchard, J. K., and N. A. Rosenberg. 1999. Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.* 65:220–228.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–59.

- Pritchard, J. K., M. Stephens, N. A. Rosenberg, and P. Donnelly. 2000. Association mapping in structured populations. *Am. J. Hum. Genet.* 67:170–81.
- Rabinowitz, D. 1997. A Transmission Disequilibrium Test for Quantitative Trait Loci. *Hum. Hered.* 47:342–350.
- Ramos, A. M., R. P. M. Crooijmans, N. A. Affara, A. J. Amaral, A. L. Archibald, J. E. Beever, C. Bendixen, C. Churcher, R. Clark, P. Dehais, M. S. Hansen, J. Hedegaard, Z.-L. Hu, H. H. Kerstens, A. S. Law, H.-J. Megens, D. Milan, D. J. Nonneman, G. A. Rohrer, M. F. Rothschild, T. P. L. Smith, R. D. Schnabel, C. P. Van Tassell, J. F. Taylor, R. T. Wiedmann, L. B. Schook, and M. A. M. Groenen. 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* 4:e6524.
- RDC Team. 2013. R: A language and environment for statistical computing.
- Reverter, A., Y. H. Wang, K. A. Byrne, S. H. Tan, G. S. Harper, and S. A. Lehnert. 2004. Joint analysis of multiple cDNA microarray studies via multivariate mixed models applied to genetic improvement of beef cattle 1. *J. Anim. Sci.* 82:3430–3439.
- Riggio, V., R. Pong-Wong, G. Sallé, M. G. Usai, S. Casu, C. R. Moreno, O. Matika, and S. C. Bishop. 2014. A joint analysis to identify loci underlying variation in nematode resistance in three European sheep populations. *J. Anim. Breed. Genet.* 131:426–436.

- Roberts, R., G. a Wells, A. F. R. Stewart, S. Dandona, and L. Chen. 2010. The genome-wide association study--a new era for common polygenic disorders. *J. Cardiovasc. Transl. Res.* 3:173–182.
- Rohrer, G. A., D. J. Nonneman, R. K. Miller, H. Zerby, and S. J. Moeller. 2012. Association of single nucleotide polymorphism (SNP) markers in candidate genes and QTL regions with pork quality traits in commercial pigs. *Meat Sci.* 92:511–8.
- Rolf, M. M., J. F. Taylor, R. D. Schnabel, S. D. McKay, M. C. McClure, S. L. Northcutt, M. S. Kerley, and R. L. Weaber. 2012. Genome-wide association analysis for feed efficiency in Angus cattle. *Anim. Genet.* 43:367–74.
- Ron, D., and P. Walter. 2007. Signal integration in the endoplasmic reticulum unfolded protein response. *Nat. Rev. Mol. cell Biol.* 8:519–529.
- De Roos, a P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179:1503–12.
- Rosenvold, K., and H. J. Andersen. 2003. Factors of significance for pork quality - a review. *Meat Sci.* 64:219–237.
- Rothschild, M. F., and A. Ruvinsky. 1998. *The genetics of the pig.* (C. International, editor.). Wallingford.
- Rousseau, S., N. Iannuccelli, M. J. Mercat, C. Naylies, J.-C. Thouly, B. Servin, D. Milan, E. Pailhoux, and J. Riquet. 2013. A genome-wide association study points

out the causal implication of SOX9 in the sex-reversal phenotype in XX pigs. *PLoS One* 8:e79882.

Ryan, M. T., R. M. Hamill, A. M. O'Halloran, G. C. Davey, J. McBryan, A. M. Mullen, C. McGee, M. Gispert, O. I. Southwood, and T. Sweeney. 2012. SNP variation in the promoter of the PRKAG3 gene and association with meat quality traits in pig. *BMC Genet.* 13:66.

Saatchi, M., M. C. McClure, S. D. McKay, M. M. Rolf, J. Kim, J. E. Decker, T. M. Taxis, R. H. Chapple, H. R. Ramey, S. L. Northcutt, S. Bauck, B. Woodward, J. C. M. Dekkers, R. L. Fernando, R. D. Schnabel, D. J. Garrick, and J. F. Taylor. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genet. Sel. Evol.* 43:40.

Sahana, G., B. Guldbrandtsen, C. Bendixen, and M. S. Lund. 2010. Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. *Anim. Genet.* 41:579–588.

Sahana, G., V. Kadlecová, H. Hornshøj, B. Nielsen, and O. F. Christensen. 2013. A genomewide association scan in pig identifies novel regions associated with feed efficiency trait. *J. Anim. Sci.* 91:1041–1050.

Salmi, B., L. Trefan, J. Bloom-Hansen, J. P. Bidanel, A. B. Doeschl-Wilson, and C. Larzul. 2010. Meta-analysis of the effect of the halothane gene on 6 variables of pig meat quality and on carcass leanness. *J. Anim. Sci.* 88:2841–55.

- Sanchez, M. P., T. Tribout, N. Iannuccelli, M. Bouffaud, B. Servin, A. Tenghe, P. Dehais, N. Muller, M. P. Del Schneider, M. J. Mercat, C. Rogel-Gaillard, D. Milan, J. P. Bidanel, and H. Gilbert. 2014. A genome-wide association study of production traits in a commercial population of Large White pigs: evidence of haplotypes affecting meat quality. *Genet. Sel. Evol.* 46:12.
- Scheffler, T. L., S. Park, and D. E. Gerrard. 2011. Lessons to learn about postmortem metabolism using the AMPK γ 3(R200Q) mutation in the pig. *Meat Sci.* 89:244–250.
- Schenkel, F. S., S. P. Miller, Z. Jiang, I. B. Mandell, X. Ye, H. Li, and J. W. Wilton. 2006. Association of a single nucleotide polymorphism in the calpastatin gene with carcass and meat quality traits of beef cattle. *J. Anim. Sci.* 84:291–299.
- Shackelford, S. D., D. A. King, and T. L. Wheeler. 2012. Chilling rate effects on pork loin tenderness in commercial processing plants. *J. Anim. Sci.* 90:2842–2849.
- Shackelford, S. D., T. L. Wheeler, and M. Koohmaraie. 1999. Evaluation of Slice Shear Force as an Objective Method of Assessing Beef Longissimus Tenderness. *J. Anim. Sci.* 77:2693–2699.
- Shackelford, S. D., and T. L. Wheeler. 2009. Slice Shear Force. *Cattlemen's Beef Board Natl. Cattlemen's Beef Assoc.*
- Shepherd, R. K., M. J. Drumm, and J. Yang. 2013. Using cross-validation in a fast EM algorithm for genomic selection and complex trait prediction. In: *Assoc. Advmt. Anim. Breed. Genet.* p. 20:37–40.

- Silva, K. M., J. W. M. Bastiaansen, E. F. Knol, J. W. M. Merks, P. S. Lopes, S. E. F. Guimarães, and J. a M. van Arendonk. 2010. Meta-analysis of results from quantitative trait loci mapping studies on pig chromosome 4. *Anim. Genet.* 42:280–92.
- Skol, A. D., L. J. Scott, G. R. Abecasis, and M. Boehnke. 2006. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* 38:209–13.
- Spelman, R. J., A. E. Huisman, S. R. Singireddy, W. Coppieters, J. Arranz, M. Georges, and D. J. Garrick. 1999. Short communication: quantitative trait loci analysis on 17 nonproduction traits in the New Zealand dairy population. *J. Dairy Sci.* 82:2514–2516.
- Spencer, C. C. A., Z. Su, P. Donnelly, and J. Marchini. 2009. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet.* 5:e1000477.
- Stankowich, T., and D. T. Blumstein. 2005. Fear in animals: a meta-analysis and review of risk assessment. *Proc. Biol. Sci.* 272:2627–34.
- Strandén, I., and D. J. Garrick. 2009. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.* 92:2971–5.

- Stratz, P., R. Wellmann, S. Preuss, K. Wimmers, and J. Bennewitz. 2014. Genome-wide association analysis for growth, muscularity and meat quality in Piétrain pigs. *Anim. Genet.* 45:350–6.
- Tait, R. G., S. D. Shackelford, T. L. Wheeler, D. A. King, E. Casas, R. M. Thallman, T. P. Smith, and G. L. Bennett. 2014. μ -Calpain, calpastatin, and growth hormone receptor genetic effects on preweaning performance, carcass quality traits, and residual variance of tenderness in Angus cattle selected to increase minor haplotype and allele frequencies. *J. Anim. Sci.* 92:456–466.
- Topel, D. G. 1986. Future meat-animal composition: Industry adaptation of new technologies. *J. Anim. Sci.* 63:633.
- Trefan, L., A. Doeschl-Wilson, J. A. Rooke, C. Terlouw, and L. Bünger. 2013. Meta-analysis of effects of gender in combination with carcass weight and breed on pork quality. *J. Anim. Sci.* 91:1480–1492.
- Turk, V., and W. Bode. 1991. The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett.* 285:213–219.
- Uimari, P., A. Sironen, and M. L. Sevón-Aimonen. 2013. Evidence for three highly significant QTL for meat quality traits in the Finnish Yorkshire pig breed. *J. Anim. Sci.* 91:2001–2011.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16–24.

- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–23.
- Veyrieras, J.-B., B. Goffinet, and A. Charcosset. 2007. MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments. *BMC Bioinformatics* 8:49.
- Visscher, P. M., S. Macgregor, B. Benyamin, G. Zhu, S. Gordon, S. Medland, W. G. Hill, J.-J. Hottenga, G. Willemsen, D. I. Boomsma, Y.-Z. Liu, H.-W. Deng, G. W. Montgomery, and N. G. Martin. 2007. Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am. J. Hum. Genet.* 81:1104–10.
- De Vries, A. G., P. G. van der Wal, T. Long, G. Eikelenboom, and J. W. M. Merks. 1994. Genetic parameters of pork quality and production traits in Yorkshire populations. *Livest. Prod. Sci.* 40:277–289.
- Walling, G. A., P. M. Visscher, L. Andersson, M. F. Rothschild, L. Wang, G. Moser, M. A. Groenen, J. P. Bidanel, S. Cepica, A. L. Archibald, H. Geldermann, D. J. de Koning, D. Milan, and C. S. Haley. 2000. Combined analyses of data from quantitative trait loci mapping studies. Chromosome 4 effects on porcine growth and fatness. *Genetics* 155:1369–78.
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, R. L. Fernando, Z. Vitezica, R. Okimoto, R. Hawken, and W. M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front. Genet.* 5:134.

- Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb)*. 94:73–83.
- Wang, L., P. Sørensen, L. Janss, T. Ostersen, and D. Edwards. 2013. Genome-wide and local pattern of linkage disequilibrium and persistence of phase for 3 Danish pig breeds. *BMC Genet*. 14:115.
- Wang, L., Y. Xiong, B. Zuo, M. Lei, Z. Ren, and D. Xu. 2012. Molecular and functional characterization of glycogen synthase in the porcine satellite cells under insulin treatment. *Mol. Cell. Biochem*. 360:169–180.
- Wang, W. Y. S., B. J. Barratt, D. G. Clayton, and J. a Todd. 2005. Genome-wide association studies: theoretical and practical concerns. *Nat. Rev. Genet*. 6:109–18.
- Weigel, K. ., G. de los Campos, A. I. Vasquez, G. J. M. Rosa, D. Gianola, and C. P. Van Tassell. 2010. Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. *J. Dairy Sci*. 93:5423–5435.
- White, S. N., E. Casas, T. L. Wheeler, S. D. Shackelford, M. Koohmaraie, D. G. Riley, C. C. Chase, D. D. Johnson, K. J.W., and T. P. Smith. 2005. A new single nucleotide polymorphism in CAPN1 extends the current tenderness marker test to include cattle of *Bos indicus*, *Bos taurus*, and crossbred descent. *J. Anim. Sci*. 83:2001–2008.

- Willer, C. J., Y. Li, and G. R. Abecasis. 2010. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190–1.
- Woelfel, R. L., C. M. Owens, E. M. Hirschler, R. Martinez-Dawson, and A. R. Sams. 2002. The Characterization and Incidence of Pale, Soft, and Exudative Broiler Meat in a Commercial Processing Plant. *Poult. Sci.* 81:579–584.
- Wood, I., G. Moser, D. L. Burrell, K. L. Mengersen, and D. J. S. Hetzel. 2006. A meta-analytic assessment of a thyroglobulin marker for marbling in beef cattle. *Genet. Sel. Evol.* 38:479–94.
- Wood, J. D., J. Wiseman, and D. J. A. Cole. 1994. Control and manipulation of meat quality. In: M. A. V. D.J.A. Cole, J. Wiseman, editor. *Principles of Pig Science*. Nottingham University Press, Nottingham, UK. p. 433–456.
- Wu, C., A. DeWan, J. Hoh, and Z. Wang. 2011. A Comparison of Association Methods Correcting for Population Stratification in Case–Control Studies. *Ann. Hum. Genet.* 75:418–427.
- Ying, W., M. D. Tokach, J. M. DeRouche, T. E. Houser, S. S. Dritz, R. D. Goodband, and J. L. Nelssen. 2014. Effects of dietary L-carnitine and dried distillers grains with solubles on growth, carcass characteristics, and loin and fat quality of growing-finishing pigs. *J. Anim. Sci.* 91:3211–3219.
- Zeeuwen, P. L., T. Cheng, and J. Schalkwijk. 2009. The biology of cystatin M/E and its cognate target proteases. *J. Invest. Dermatol.* 129:1327–1338.

- Zhang, L., J. Liu, F. Zhao, H. Ren, L. Xu, J. Lu, S. Zhang, X. Zhang, C. Wei, G. Lu, Y. Zheng, and L. Du. 2013. Genome-wide association studies for growth and meat production traits in sheep. *PLoS One* 8:e66569.
- Zhang, Z., E. Ersoz, C.-Q. Lai, R. J. Todhunter, H. K. Tiwari, M. a Gore, P. J. Bradbury, J. Yu, D. K. Arnett, J. M. Ordovas, and E. S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42:355–60.
- Zhang, Z., U. Ober, M. Erbe, H. Zhang, N. Gao, J. He, J. Li, and H. Simianer. 2014. Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association studies. *PLoS One* 9:e93017.
- Zhou, B., J. Shi, and A. S. Whittemore. 2011. Optimal methods for meta-analysis of genome-wide association studies. *Genet. Epidemiol.* 35:581–91.
- Zhu, M., and S. Zhao. 2007. Candidate gene identification approach: progress and challenges. *Int. J. Biol. Sci.* 3:420–427.