GENOME-WIDE ASSOCIATION STUDY IN AN F2 DUROC X PIETRAIN RESOURCE POPULATION FOR MEAT QUALITY AND CARCASS TRAITS

By

Sebastián Casiró

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Animal Science-Master of Science

2016

ABSTRACT

GENOME-WIDE ASSOCIATION STUDY IN AN F2 DUROC X PIETRAIN RESOURCE POPULATION FOR MEAT QUALITY AND CARCASS TRAITS

By

Sebastián Casiró

Accurate association mapping in livestock populations is challenging, thus I tested the properties of three methods to derive 95% confidence intervals (CI) for QTL positions [Parametric Method (PM), non-parametric centered (NPC) and non-parametric non-centered (NPNC)]. The NPC failed to provide adequate coverage for the 95% CI for the true QTL position. The 95% CI obtained with NPNC and PM had similar coverage, however the PM had shorter intervals, therefore, I recommend to use PM. Furthermore, to map regions comprising SNP associated with meat quality and carcass traits I performed Genome-wide Association analysis for 948 F2 Duroc x Pietrain resource population pigs for 38 meat quality and carcass traits using 44,911 SNP. Type I error rate was controlled at a False Discovery Rate of 5%. I found nine QTL associated with 15 traits. Three of those nine QTL [one on SSC1 (tenth rib backfat thickness), one on SSC7 (dressing percentage and loin muscle area) and one on SSC11 (belly weight)] were mapped to a specific genomic segment in this study. Moreover, two novel QTL associated with tenderness were located on SSC3 and SSC5. Also, I propose the candidate genes A Kinase (PRKA) Anchor Protein 3 for the QTL on SSC5 and Carnitine O-Acetyltransferase for the QTL on SSC1. Finally, this study shows that the variants of Protein Kinase AMP-activated γ 3-subunit, I199V and T30N are not associated with pH 24 post mortem and related traits in this population.

This thesis is gratefully dedicated to my wife Tatiana, Mom, Dad, my brothers Dami and Cheche and my grandparents Cacho, Cocho, Perla and Tamara, their love and constant support helped me to accomplish everything I proposed in my life.

AKNOWLEDGMENTS

To accomplish such an endeavor as this two years' journey, it requires a los of support encouragement and guidance. First, I would love to show my gratitude to Drs. Alvarez, Morado, Perez Aguirreburualde, Rebuelto and Rodriguez. They always believed in me and encouraged me to pursue and advance degree. I am deeply grateful to Dr. Alvarez, who showed me the research side of veterinary medicine and has been a great counselor since I met him.

I owe my deepest gratitude to my advisor Dr. Juan Pedro Steibel, his constant guidance, encouragement and support helped me to accomplish this thesis and several projects in his lab. Furthermore, I will always treasure his advice and guidance with regards of my future goals.

Also, I would love to thank my guidance committee Drs. Bates, Ernst, Lu and Steibel, who listened to my goals and suggested courses to tailor my graduate studies to achieve those goals. Moreover, I am grateful to Dr. Bates who helped me having a better understanding of the pig production in US. Also, I am grateful to Dr. Ernst who advised me throughout the candidate gene and several issues regarding the graduate student association.

I would like to thank Deborah, Kaitlin, Kaitlyn and Scott for donating their time and assisted me when I needed. Also I am grateful to Carly, Kaitlin and Kaitlyn, the sibling's relationship that we have was crucial making this journey easier. Additionally, I want to thank Pablo and Maria, they opened the door of their house and treated me like family since the first day I arrived. Since then, they also became my extended family in East Lansing.

Also I am deeply thankful to my friends in Argentina whose daily contact made me feel I was at home. Thanks, Andy, Bari, Fede, Gri, Huevo, Juli, Leo, Lio, Pelu, Niky and Motor. Furthermore, I am grateful to my extended family in Argentina, Bety, Dani, Mica and Nati for their love and

support since I was born. Moreover, I would love to thank Susy and Jorge, for rising such a strong, kind, funny, intelligent and lovely daughter.

I am very thankful to my entire family who always supported and encouraged me to peruse whatever I wanted to do in this life. A special thanks to my aunt Jessi who encouraged me to study abroad and advised me in every step of this journey. Thanks to Sol and Minu for the unconditional love they gave me. Also, to my grandmothers and brothers who always been there for me. Furthermore, I want to thank my parents who taught me to treasure the family and that with hard work you can accomplish every goal that you set, they are an inspiration for me. Finally, I want to thank my amazing wife Tatiana, she done the greatest sacrifice for my education. I would not be able to recover all the time that I did not spend with her these past years, and I hope I can find a way to thank her for her constant support and understanding. I think she deserves an honorary degree for everything that she done for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Introduction

Traditional swine genetic improvement programs were tailored to enhance efficiency traits, such as maximizing lean growth, reducing backfat thickness and improving feed conversion to produce the highest quantity of meat at the lowest cost. Providing high value protein at a lower price is crucial for the industry; however, consumers are not only looking for quantity but also for quality, depending on their preferences and perception of the meat. Thus, the breeding goals shifted and nowadays, they are also directed towards improving meat quality. Meat quality is subject to genetic control, as revealed by between breed differences and by within breed heritabilities of relevant traits. For instance, meat quality traits heritabilities range from 0.08 to 0.30, while carcass traits heritabilities range from 0.30 to 0.60 (Sellier, 1998; van Wijk et al., 2005).

Traditional breeding methods have been used to improve efficiency, carcass and meat quality traits. In particular, crossbreeding has been well exploited in swine growth and carcass traits (Schneider et al., 1982) to produce individuals with better performance than the average of their parents. Also, purebred lines have been selected to exploit within-breed variation using selection indexes (Hazel, 1943). However, selection for one trait may affect the response of correlated traits. For instance, selecting for lean growth efficiency has resulted in meat that had normal color, but was softer and more exudative (Lonergan et al., 2001). This happened because meat quality traits are negatively correlated with lean growth. For instance van Wijk et al. (2005) reported that backfat and pH 24 hours post mortem have a genetic correlation r=-0.24. Furthermore, those authors also showed that percentage of lean meat is negatively correlated with meat color traits such as Minolta L*, a* and b* in ham and loin (-0.16<r<-0.62) and with meat firmness (r= -0.21) among other traits. Likewise, van Wijk et al. (2005) estimated that

average daily gain is negatively correlated with meat firmness (r= -0.67). Additionally, Sellier, (1998) reported negative genetic correlations between growth traits and meat quality traits ranging from r= -0.2 to r= -0.82. Consequently, selecting for growth traits alone would lead to faster growing pigs with lower meat quality. To circumvent this problem, selection indexes should include both growth and meat quality traits. However, meat quality traits, carcass traits and some growth traits are expensive to measure because most are expressed later in life and require access to an abattoir for post-mortem data collection. This imposes a challenge to traditional breeding methods that rely on accurately estimating multi-trait breeding values from an animal's own records and from progeny records to construct selection indexes.

To overcome the problem of increased cost of measuring traits later in life, having an early selection criterion would allow the swine industry to make rapid genetic progress at a reasonable cost. Currently genomic selection (Meuwissen et al., 2001; Goddard & Hayes, 2007) is widely used in livestock species (Hayes et al., 2009; Chen et al., 2011; Wang et al., 2012; Akanno et al., 2014) to estimate Genomic Breeding Values (GEBV). Genomic selection is a form of marker assisted selection, which relies on exploiting Linkage Disequilibrium (LD) between Quantitative Trait Loci (QTL) of genetic markers densely distributed along the genome (Goddard and Hayes, 2007). Thus, genomic selection enables prediction of breeding values for several traits at birth using the animal's genotype. However, because genomic selection relies on LD between markers and the actual causal alleles, it may have some shortcomings. For instance, the marker effects need to be re-estimated periodically (Habier et al., 2007) or marker effects estimated in one population are not informative to predict breeding values in a different population. This problem could be circumvented if the actual causative alleles, for instance, Quantitative Trait Nucleotides (QTN) are used in selection indexes (Weller and Ron, 2011).

Furthermore, finding QTN can facilitate the selection for negatively correlated traits, because the selection index can incorporate information about QTN that are neutral for one trait and significantly improve the other trait. For instance, several studies conducted in swine found causative variants on the Protein Kinase AMP-activated γ 3-subunit gene (PRKAG3) affecting meat quality (Milan et al., 2000; Ciobanu et al., 2001; Ryan et al., 2012; Uimari & Sironen, 2014), but these alleles do not seem to be associated with growth.

Finally, knowing the actual QTN behind a QTL has biotechnological value. For instance, genome editing technologies could be used to "fix" all desirable alleles in one generation. This technology is already being used in pigs to improve disease resistance traits (Whitworth et al., 2016) and, in combination with genomic selection, promises to significantly increase selection response (Jenko et al., 2015).

Discovery of QTN of complex traits is a tedious and difficult task due to the polygenic nature of the genetic architecture of most traits (Manolio et al., 2009). A very common first step used to identify genomic regions containing QTN is to conduct a Genome-Wide Association (GWA) analysis (Weller and Ron, 2011).

In swine, the development of the Illumina PorcineSNP60 BeadChip (Ramos et al., 2009) allowed the implementation of genomic selection and GWA for meat quality and carcass traits (Becker et al.,2013; Ma et al., 2013; Nonneman et al., 2013; Uimari et al., 2013; Badke et al., 2014;Sanchez et al., 20 14; Bernal Rubio et al., 2015a;). A very common method used to implement GWA in livestock species in general and in swine in particular consists in the transformation of animal specific breeding values into SNP effects (Wang et al., 2012; Gualdrón Duarte et al., 2014; Bernal Rubio et al., 2015b). The output of these GWA analyses consists of a list of SNP-specific effects and/or p-values which need to be filtered for statistical significance

and spatially clustered to define putative QTL regions. It is expected that the QTL regions will contain the causative variants or genes behind the association signal. However accurately defining genomic regions from SNP association tests results is still challenging due to long range persistence of LD among other reasons.

In GWA different methods have been used to bound a genomic region around a QTL peak. One alternative is to select the region defined by all significant and contiguous SNP, another alternative is to consider a fixed number of SNP downstream and upstream of the peak (i.e. Hayes et al., 2010) or to use a fixed physical distance around the peak (i.e. Gualdrón Duarte et al., 2014). However, defining those fixed distances is challenging and arbitrary. A better idea is to define the region based on the LD blocks surrounding the QTL peak (Bernal Rubio et al., 2015a). But LD blocks are difficult to define objectively. Ideally, putative QTL regions should be defined using a statistical support interval. The derivation of confidence intervals (CI) has been widely studied in linkage analysis QTL mapping (Lander & Botstein, 1989; Visscher et al., 1996). For GWA analyses, Hayes (2013) proposed to calculate CI using a data partition algorithm. However, this method has never been applied, and its properties remain unknown. Some important questions that should be answered before using the proposed CI computation methods are: Does the nominal 95% confidence intervals obtained with Hayes (2013) method provides actual 95% coverage? Is there a more efficient method that results in shorter intervals of similar coverage? These questions are important because a shorter CI will contain less genes to study as potential candidates, but at the same time, a shorter CI will have less chances of containing the actual causative gene. Consequently, implementing GWA of meat quality traits in swine using sound statistical methods that produce putative QTL regions could facilitate discovery of causative genes.

The overall goal of this study was to find novel QTL, to refine known QTL for meat quality and carcass traits and to propose candidate genes for further studies. To attain this overall goal, two specific aims were proposed:

1. Implement and test properties of methods for computing the confidence interval of a QTL position in the context of GWA from mixed effects GBLUP models.

2. Perform GWA of meat quality and carcass traits in an F2 Duroc x Pietrain resource population using the methods tested under aim 1 and propose candidate genes for further study.

LITERATURE CITED

LITERATURE CITED

Akanno, E. C., Plastow, G., B. W. Woodwards, S. Bauck, H. Okut, X.-L. Wu, C. Sun, J. L. Aalhus, S. S. Moore, S. P. Miller, W. Z., 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. http://doi.org/10.2527/jas.2013-7374

Badke, Y. M., R. O. Bates, C. W. Ernst, J. Fix, and J. P. Steibel. 2014. Accuracy of estimation of genomic breeding values in pigs using low-density genotypes and imputation. G3 (Bethesda). 4:623–31. http://doi.org/10.1534/g3.114.010504

Becker, D., K. Wimmers, H. Luther, A. Hofer, and T. Leeb. 2013. A Genome-Wide Association Study to Detect QTL for Commercially Important Traits in Swiss Large White Boars. PLoS One 8:1–6. http://doi.org/10.1371/journal.pone.0055951

Bernal Rubio, Y. L., J. L. Gualdrón Duarte, R. O. Bates, C. W. Ernst, D. Nonneman, G. A. Rohrer, A. King, S. D. Shackelford, T. L. Wheeler, R. J. C. Cantet, and J. P. Steibel. 2015a. Implementing meta-analysis from genome-wide association studies for pork quality traits 1. J. Anim. Sci. 93:5607–5617. http://doi.org/10.2527/jas2015-9502

Bernal Rubio, Y. L., J. L. Gualdrón Duarte, R. O. Bates, C. W. Ernst, D. Nonneman, G. A. Rohrer, A. King, S. D. Shackelford, T. L. Wheeler, R. J. C. Cantet, and J. P. Steibel. 2015b. Meta-analysis of genome-wide association from genomic prediction models. Anim. Genet.:36–48. http://doi.org/10.1111/age.12378

Chen, C. Y., I. Misztal, I. Aguilar, S. Tsuruta, the Meuwissen, S. E. Aggrey, T. Wing, and W. M. Muir. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotyping. J. Anim. Sci. 89:23. http://doi.org/10.2527/jas.2010-3071

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 3 -Subunit Gene Associated With Low Glycogen Content in Pig Skeletal Muscle and Improved Meat Quality. Genetics 159:1151–1162.

Goddard, M. E., and B. J. Hayes. 2007. Genomic selection. J.Anim.Breed.Genet 124:323–330. http://doi.org/10.1080/09064700801959395

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. http://doi.org/10.1186/1471-2105-15-246

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397. http://doi.org/10.1534/genetics.107.081190

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433–443.

http://doi.org/10.3168/jds.2008-1646

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic Prediction: Coat colour, Milk-fat percentage, and type in holstein cattle as contrasting model traits. PLoS Genet. 6. http://doi.org/10.1371/journal.pgen.1001139

Hayes, B. J. 2013. Overview of statistical methods for genome-wide association studies. C. Gondor, B. van der Werf, & B. J. Hayes, editors, Genome-wide association studies and genomic prediction. Human Press, New York, NY. p. 156-157.

Hazel, L. N. 1943. The Genetic Basis for Constructing Selection Indexes. Genetics 28:476–490.

Jenko, J., G. Gorjanc, M. A. Cleveland, R. K. Varshney, C. B. A. Whitelaw, J. A. Woolliams, and J. M. Hickey. 2015. Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. Genet. Sel. Evol. 47:55. http://doi.org/10.1186/s12711-015-0135-3

Lander, E. S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP limkage maps. Genetics 121:185–199.

Lonergan, S. M., E. Huff-Lonergan, L. J. Rowe, D. L. Kuhlers, and S. B. Jungst. 2001. Selection for lean growth efficiency in Duroc pigs influences pork quality. J. Anim. Sci. 79:2075–2085. http://doi.org//2001.7982075x

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. http://doi.org/10.1371/journal.pone.0064047

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. C. Mackay, S. A. McCarroll, and P. M. Visscher. 2009. Finding the missing heritability of complex diseases. Nature 461:747–753. http://doi.org/10.1038/nature08494

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. http://doi.org/11290733

Milan, D., J. T. Jeon, C. Looft, V. Amarger, A. Robic, M. Thelander, C. Rogel-Gillard, 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. http://doi.org/10.1126/science.288.5469.1248

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 D. J. Nonneman, S. D. Shackelford, D. A. King, T. L. Wheeler, R. T.

Wiedmann, W. M. Snelling and G. A. Rohrer. J. Anim. Sci.:4043–4050. http://doi.org/10.2527/jas2013-6255

Ramos, A. M., R. P. M. A. 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. http://doi.org/10.1371/journal.pone.0006524

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. http://doi.org/10.1186/1471-2156-13-66

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. http://doi.org/10.1186/1297-9686-46-12

Schneider, J., L. . Christian, and D. L. Kuhlers. 1982. Crossbreeding in swine: Genetic effects on litter performance. J. Anim. … 54:739–746.

Sellier, P. 1998. Genetics of meat and carcass traits. Rothschild M.F. and Ruvinsky A. editors, The genetics of the pigs. CAB International, Wallingford, UK, pp. 463-510.

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. http://doi.org/10.2527/jas.2012-5811

Uimari, P., and A. Sironen. 2014. A combination of two variants in PRKAG3 is needed for a positive effect on meat quality in pigs. BMC Genet. 15:29. http://doi.org/10.1186/1471-2156-15-29

Visscher, P. M., R. Thompson, and C. S. Haley. 1996. Confidence Intervals in QTL Mapping by Bootstrapping. Genetics 143:1013–1020.

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. http://doi.org/10.1017/s0016672312000274

Weller, J. I., and M. Ron. 2011. Invited review: quantitative trait nucleotide determination in the era of genomic selection. J. Dairy Sci. 94:1082–1090. http://doi.org/68/jds.2010-3793

Whitworth, K. M., R. R. R. Rowland, C. L. Ewen, B. R. Trible, M. A. Kerrigan, A. G. Cino-Ozuna, M. S. Samuel, J. E. Lightner, D. G. McLaren, A. J. Mileham, K. D. Wells, and R. S. Prather. 2016. Gene-edited pigs are protected from porcine reproductive and respiratory

syndrome virus. Nat. Biotechnol. 34:20–22. http://doi.org/10.1038/nbt.3434

van Wijk, H. J., D. J. G. Arts, J. O. Matthews, M. Webster, B. J. Ducro, and E. F. Knol. 2005. Genetic parameters for carcass composition and pork quality estimated in a commercial production chain. J. Anim. Sci 83:324–333.

CHAPTER TWO

Confidence intervals for Quantitative Trait Loci position in Genome-wide Association analysis obtained with Genomic Best Linear Unbiased Predictor models

ABSTRACT

Deriving confidence intervals (CI) for the position of quantitative trait loci has been widely studied in Linkage Analysis. However, this problem has not been fully studied for Genome-Wide Association (GWA) analyses based on mixed models. The objective of this study was to propose and test the properties of two non-parametric methods to compute CI; the non-parametric centered CI (NPC), and the non-parametric non-centered CI (NPNC). Also, we tested the properties of a previously published parametric method (PM) that, so far, had remained untested. The 10^{th} rib backfat thickness (BF10) measurements for 947 F2 animals from the Michigan State University Duroc x Pietrain resource population were used as a base data set for simulations. We simulated 200 plasmode datasets based on the BF10 phenotypes and associated genotypes with a QTL at position 133.88Mb on SSC6 accounting for 12.6% of the genetic variation. We fitted a Gaussian linear mixed model to estimate the breeding values, and we divided the data in two halves and performed the GWA, repeating the process 200 times saving the physical position of the most significant SNP in each half-data set. Finally, we calculated the 95% CI using the empirical distribution for NPC and NPNC and assuming asymptotic normality for the PM. The 95% CI derived from the real data set showed that the NPC had the shortest interval (7.79 Mb.), followed by PM (8.84 Mb.) and NPNC (12.22 Mb.). However, the nominal 95% NPC CI only provided 89.5% coverage of the true QTL position. On the other hand, the nominal 95% PM CI and NPNC CI were slightly conservative covering 96.5% and 96% of the true QTL position. However, NPNC resulted in wider CI than the PM (9.4 vs 8 Mb.). Therefore, the PM provides

the shortest and most conservative CI for the QTL position. The calculation of CI with good properties (coverage and minimum length) is crucial to refine QTL regions. For instance, for BF10 the genomic region containing significantly associated SNP spanned 65.42 Mb, but it was reduced to an 8.84 Mb region using the PM CI. Having a narrower region with high probability of containing the QTL helps with candidate gene search and in the design of resequencing experiments.

Key words: optimal confidence interval, GBLUP based GWA, simulation.

INTRODUCTION

Genomic selection is widely applied in livestock species to predict genomic breeding values or GEBV (Meuwissen et al., 2001; Goddard & Hayes, 2007,Hayes et al., 2009; Chen et al., 2011; Wang et al., 2012; Akanno et al., 2014). Following the estimation of the GEBV, it is common to perform Genome-Wide Association (GWA) analyses to identify SNP associated with phenotypes. A computationally efficient way of performing GWA is to estimate marker effects through linear transformation of GEBV and to estimate their variance to implement hypothesis testing (Gualdrón Duarte et al., 2014; Bernal Rubio et al., 2015). Thus, genomic regions comprising SNP associated with the traits can be mapped.

It is common that due to persistence of LD, sizable genomic regions are defined by GWA. Hence, there is a need to define a confidence interval (CI) for the true Quantitative Trait Loci (QTL) position. Refining the genomic regions by deriving the CI for a QTL narrows the list of candidate genes in a QTL region, while maximizing the chances that the actual causal genes are retained in the list. Moreover, by focusing on a narrower region, less genes can be further studied

to validate them as causative genes. Additionally, if further genotyping is needed, sequencing a narrower region with high confidence of containing causative genes optimizes the use of resources to characterize key genomic regions. Therefore, accurately calculating the CI around a QTL peak is beneficial for post-GWA studies in vitro and in silico.

The problem of computing the confidence interval for the position of a QTL was addressed in least squares based linkage analysis using bootstrapping (Visscher et al., 1996). However, deriving CI for the position of a QTL in a mixed model based GWA imposes a challenge: there is a need to perform randomization while accounting for the sample structure. Hayes (2013) proposed a data partition algorithm to compute the CI for the position of a QTL assuming asymptotic normality for the length of the QTL interval. However, this method has never been used and the properties of the CI using such parametric method (PM) remain unknown.

The main objective of this study is to use plasmode simulations to test a previously proposed CI calculation method for GWA and to compare to non-parametric alternatives also based on randomized data-partition methods.

MATERIALS AND METHODS

Ethical Statement

Animal protocols were approved by the Michigan State University All University Committee on Animal Use and Care (AUF# 09/03-114-00)

Real Dataset

The experimental population consisted of an F2 Duroc X Pietrain resource population created at Michigan State University (Edwards et al., 2008). Briefly, the F0 generation consisted of four

Duroc sires and 15 Pietrain dams. From those matings six F1 boars (representing 3 of the Duroc sires) and 50 F1 dams were selected. The F2 generation comprised a total of 1259 non inbred pigs of both sexes from 142 litters. Phenotypic data for 38 meat quality and carcass traits were recorded for approximately 948 F2 animals (Edwards et al., 2008). For illustration purposes, in this study we used backfat thickness at the 10^{th} rib (BF10) obtained at slaughter, which is an economically important trait with a heritability of 0.449. We selected this trait because there is evidence of a QTL on SSC6 that has been reported previously by our group for this population (Edwards et al., 2008; Choi et al., 2011), and also in a Berkshire x Yorkshire population (Malek et al., 2001, Kim et al., 2005). More interestingly, the QTL region exhibits long range persistence of linkage disequilibrium, which makes it difficult to bound the QTL position. Therefore, this trait and the QTL on SSC6 is a very good example to illustrate the application of confidence interval computation methods to narrow down the position of important QTL.

Genotyping and genotype editing

The animals from this experimental population were genotyped using two different SNP chips (Gualdrón Duarte et al., 2013). All the Grandparents, parents and 336 F2 animals were genotyped with the Illumina PorcineSNP60 BeadChip (Ramos et al., 2009) which encompasses 62,163 Single Nucleotide Polymorphism (SNP). The remaining 612 F2 animals were genotyped using the GeneSeek Genomic Profiler for Porcine LD (GGP-Porcine LD, GeneSeek a Neogen Company, Lincoln, NE), which is a SNP chip with lower resolution with 8,836 Tag SNP (Badke et al., 2013). We removed 2,277 SNP whose genotypes were missing for all animals in the Illumina PorcineSNP60 BeadChip. Furthermore, we checked for Mendelian inconsistencies (Forneris et al., 2015) and removed 1,155 SNP that did not fit the Mendelian inheritance rules

($p<8.4E^{-8}$). To account for multiple testing, the threshold ($p<8.4E^{-8}$) was calculated performing a Bonferroni correction. Following these edits we imputed all the missing genotypes for both chips (Badke et al., 2013) using the FImpute software with default settings (Sargolzaei et al., 2014). During the imputation, SNP-specific accuracies where estimated having an average accuracy r^2=0.97. At this point we removed 712 SNP which had an imputation accuracy r^2<0.64. Moreover, 101 SNP which had more than 10% of missing genotypes were removed after the imputation, because we were unable to calculate a reliable imputation accuracy. Finally, The FImpute software detected 147 SNP and 9 animals having genotyping errors, therefore we removed those individuals and SNP from our data. Altogether the dataset used in this study has 947 F2 pigs with phenotypic records for BF10 and 44,911 SNP.

Statistical Analysis

We performed a Genomic Best Linear Unbiased Predictor (GBLUP) based on a GWA, for the association study (Gualdrón Duarte et al., 2014), fitting an animal-centric Gaussian linear mixed model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{a} + \boldsymbol{e}, \qquad (2.1)$$

where $\boldsymbol{y}$ is the vector containing the phenotypes for BF10 in *mm*, $\boldsymbol{X}$ is the incidence matrix which relates the individual records with the fixed effects of sex, slaughter group and carcass weight in $\boldsymbol{\beta}$ (Edwards et al., 2008), $\boldsymbol{a}$~N $(0, \mathbf{G}\sigma^2_A)$ is a vector of random breeding values where we used the marker based relationship matrix **G** as a covariance to account for population substructure (Janss et al., 2012). Where **G** was calculated by multiplying **Z** which is the standardized allelic dosages (counts of the "B" allele) matrix (VanRaden, 2008) by its transpose, and Z is equal to:

$$\boldsymbol{Z}_{ij} = \frac{\boldsymbol{M}_{ij} - 2\,p_j}{\sqrt{m\,(\,2\,p_j\,(1 - p_j))}}$$

$\boldsymbol{M}$ is the allelic dosage matrix, p_j is the allelic frequency at the j^{th} marker, i is the i^{th} animal and m is the number of markers. Lastly, $\boldsymbol{e}$~N (0, $\boldsymbol{I}\sigma^2_e$) is the vector of residuals. We proceeded to estimate the SNP effect ($\widehat{\boldsymbol{g}}$) performing a linear transformation of the estimated breeding values ($\widehat{\boldsymbol{a}}$) obtained from fitting the model in (2.1) and the estimated SNP effect variance ($Var(\widehat{\boldsymbol{g}})$) following (Gualdrón Duarte et al., 2014) :

$$\widehat{\boldsymbol{g}} = \boldsymbol{Z}'\boldsymbol{G}^{-1}\ \widehat{\boldsymbol{a}}\,, \tag{2.2}$$

$$Var(\widehat{\boldsymbol{g}}) = \boldsymbol{Z}'\boldsymbol{G}^{-1}\boldsymbol{Z}\ \sigma_A^2 - \boldsymbol{Z}'\boldsymbol{G}^{-1}\boldsymbol{C}^{\boldsymbol{aa}}\ \boldsymbol{G}^{-1}\ \boldsymbol{Z}, \tag{2.3}$$

where $\boldsymbol{C}^{\boldsymbol{aa}}$ is the portion of the inverse of the mixed model equations that correspond to animal effects, $\boldsymbol{G}^{-1}$ is the inverse of the relationship matrix and the other terms have been previously described in equation (2.1). After estimating the SNP effect and its variance we calculated the test statistics by standardizing $\widehat{\boldsymbol{g}}$ and then p-values were obtained from the Gaussian distribution:

$$t_j = \frac{\hat{g}_j}{\sqrt{var(\widehat{g_j})}}\,, \tag{2.4}$$

$$p - value_j = 2(1\text{- }\phi\ (|t_j|)), \tag{2.5}$$

where the subscript *j is* the j^{th} SNP, and $\emptyset(x)$ is the cumulative density function of the normal distribution. Gualdrón Duarte et al. (2014) and Bernal Rubio et al. (2015), showed that this procedure is equivalent to testing one fixed SNP at a time conditional on estimated variance ratio from model (2.1). All the computations performed in this study were done with the gwaR package (https://github.com/steibelj/gwaR.git) in the R environment (https://cran.r-project.org). To control for multiple testing a genome-wide significant threshold was determined using an FDR=5% (Benjamini & Hochberg, 1995; Storey, 2002; Storey & Tibshirani, 2003).

Stochastic simulations

A stochastic simulation was performed to evaluate the properties of the confidence intervals. Genotypes from the dataset were assumed fixed (observed genotypes) and the phenotype were simulated following:

$$\boldsymbol{y_{sim}} = \boldsymbol{X}\widehat{\beta_1} + \boldsymbol{a_{sim}} + \boldsymbol{Z_{peak}}\widehat{\beta_{peak}} + \boldsymbol{e_{sim}}\,, \tag{2.6}$$

where $\boldsymbol{y_{sim}}$ is the simulated phenotype, $\boldsymbol{X}$ is the incidence matrix which relates the records with the fixed effects in $\widehat{\beta_1}$. The only fixed effect used in this simulation was the general mean of the phenotypic records obtained from the real dataset, therefore $\widehat{\beta_1} = 24.16$ mm of BF10. $\boldsymbol{a_{sim}}$ is a vector containing the random breeding values which are simulated from $\boldsymbol{a_{sim}}$~N (0, $\mathbf{G}\,\sigma^2_A$), where **G** is the genomic relationship matrix and σ^2_A= 12.67 is the additive variance estimated from real data using model (1). $\boldsymbol{e_{sim}}$~N (0, $\boldsymbol{I}\sigma^2_e$) is the simulated vector of the residuals where σ^2_e= 15.52 is the error variance of real data estimated using model (2.1). Finally, $\boldsymbol{Z_{peak}}$, is the standardized genotype at marker M1GA0008917 and $\widehat{\beta_{peak}} = -401.5$ is the estimated fixed effect of genotype at marker M1GA0008917 estimated from real data by fitting the following model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z_{peak}}\beta_{peak} + \boldsymbol{a} + \boldsymbol{e}, \tag{2.7}$$

where $\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{a}$ and $\boldsymbol{e}$ were previously described in equation (2.1) and $\boldsymbol{Z_{peak}}$ was previously explained in equation (2.6) and β_{peak} is the fixed marker effect. This simulation resulted in a QTL at position 133.88 Mb, which explains 12.6% of the phenotypic variance. Furthermore, after simulating the QTL using the SNP M1GA0008917, the SNP was removed from the data, to emulate the most common situation where the causative variant itself is not genotyped. Among the remaining SNP in the data, the closest SNP (ASGA0029650) was highly correlated with the causative variant (r^2=0.91)

Confidence interval based on a cross-validation

Hayes (2013) proposed an algorithm based on data partition to compute a CI around a QTL peak in mixed model GWA. We followed his procedure, and we proposed two non-parametric alternatives.

In general terms, the method consists of the following steps:

Step 1: Perform a GWA with the whole dataset to estimate the QTL position: p (corresponding to the SNP with smallest p-value)

Step 2: Randomly divide the dataset into two halves (x_1, x_2).

Step 3: Perform a GWA for each half separately, and record the physical position of the QTL peak for each half into the vectors v_1 and v_2

Step 4: Repeat step 2 and 3 n times growing the vectors v_1 and v_2 in each iteration.

Calculate the CI using v_1 and v_2 following the proposed methods. For this study the nominal value of the CI was 95% CI.

Alternative 1: Parametric (PM) 95% Confidence Interval (Hayes, 2013)

Step 5 A1: Calculate the Standard Error of the difference between the positions in v_1 and v_2 following:

$$se(\bar{x}) = \sqrt{\frac{1}{4n}\sum_{k=1}^{n}(v_{1k} - v_{2k})^2}$$

where, *k* corresponds to the k^{th} iteration from Step 4

Step 6 A1: Calculate the 95% CI using p from step 1, z value which corresponds to 97.5 percentile in a normal Gaussian distribution (z=1.96) and the $se(\bar{x})$ from step 6

$$CI = p \pm z_{97.5} se(\bar{x})$$

Alternative 2: Non-parametric CI centered on the maximum likelihood QTL position (NPC).

Step 5 A2: Calculate $v_{diff} = |v_2 - v_1|$.

Step 6 A2: Compute the j percentile of v_{diff} (i.e. $j = 95\%$) $l = j^{th} percentile\ of\ v_{diff}$.

Step 7 A2: Calculate the confidence interval, using p from step 1 following: p $\pm\, l/2$

Alternative 3: Non-Parametric CI not centered on the estimated QTL position (NPNC)

Step 5 A3: Reorder the pair v_{1k} and v_{2k} so the smallest physical position is stored in vectors v_1 and the largest in v_2. Where k is the k^{th} iteration from Step 4.

Step 6 A3: Select a desired level of confidence c (i.e 95%).

Step 7 A3: Define the area under the left tail of the distribution of QTL positions outside the CI lb (e.g.: lb=0.01 for 1%).

Step 8 A3: Define the area under the right tail of the distribution of QTL positions outside the CI: $ub = 1 - c - lb$ (e.g: 0.04 if lb=0.01 and c=0.95)

Step 9 A3: Calculate lower boundary (LB) from lb: LB= lbth percentile of v_1.

Step 10 A3: Calculate the upper boundary (UB) from ub: UB= ub^{th} percentile of v_2.

Step 11 A3: Calculate the length the CI ($length = UB -$ LB)

Step 12 A3: repeat steps 7-11 A3 for several values of lb and keep the CI of the shortest length

CI Coverage computation

To calculate the coverage of the nominal 95% CI, we compare the lower and upper boundary of the CI_{mp} with the true QTL position, where m corresponds to the m^{th} method and p corresponds to the p^{th} plasmode dataset. If the true QTL position was larger than the lower boundary of the

CI_{mp} and smaller than the upper boundary of the CI_{mp}, we consider that the CI_{mp} covered the true QTL position. If any of those conditions were not fulfilled the CI_{mp} did not cover the true QTL position. Finally, the percentage of coverage was calculated as follows:

$$CI_m\ coverage = \frac{Number\ of\ CI_m\ covering\ the\ true\ QTL\ position}{Number\ of\ CI_m\ computed} * 100$$

where *m* corresponds to the m^{th} method, and the number of CI_m computed was 200 in this study.

Post -GWAS analysis

The genomic region used for the functionally annotated gene identification was defined by the 95% confidence interval constructed around the peak. Annotated genes within those genomic regions were identified with the ENSEMBL annotation of *Sus scrofa* 10.2.83 (December 2015) assembly (http://useast.ensembl.org/biomart/martview/).

RESULTS AND DISCUSSION

Genome-wide Association Study in BF10

Backfat thickness is used in packing plants to determine the price of the carcass, in this study we used the BF10. This trait has a phenotypic mean ($\bar{x} = 24.16$), phenotypic variance (σ^2_{pheno} = 53.98) and heritability ($h^2 = 0.449$) in this population. The genome-wide association study shows three QTL regions associated with the trait on SSC1, SSC6 and SSC15 (Fig. 2.1 A). For this study we concentrate on the region on SSC6 which spans from 75.14 Mb (blue arrow Fig. 2.1 B). to 140.56 Mb (red arrow Fig. 2.1 B). The genotypes of the SNP corresponding to the QTL peak (M1GA0008917 located at 133.88 Mb) explain 12.6 % of the variance. As it can be seen in Fig. 2.1 B, there appears to be at least 4 peaks on SSC6 (black arrows). Hence in order to determine how many QTL may actually explain the observed association pattern, we fitted the

equation (2.7), which is a Gaussian linear mixed model conditioning on the genotype of the most significant SNP on SSC6 (M1GA0008917) as a fixed effect. After fitting those genotypes as fixed effects, no other SNP in the region was associated with the trait, thus we conclude that only one QTL is present in this region. This association peak corresponds to a well known QTL previously mapped in a low resolution linkage mapping for this population (Edwards et al., 2008; Choi et al., 2011) and for other populations (Malek et al., 2001; Kim et al., 2005).

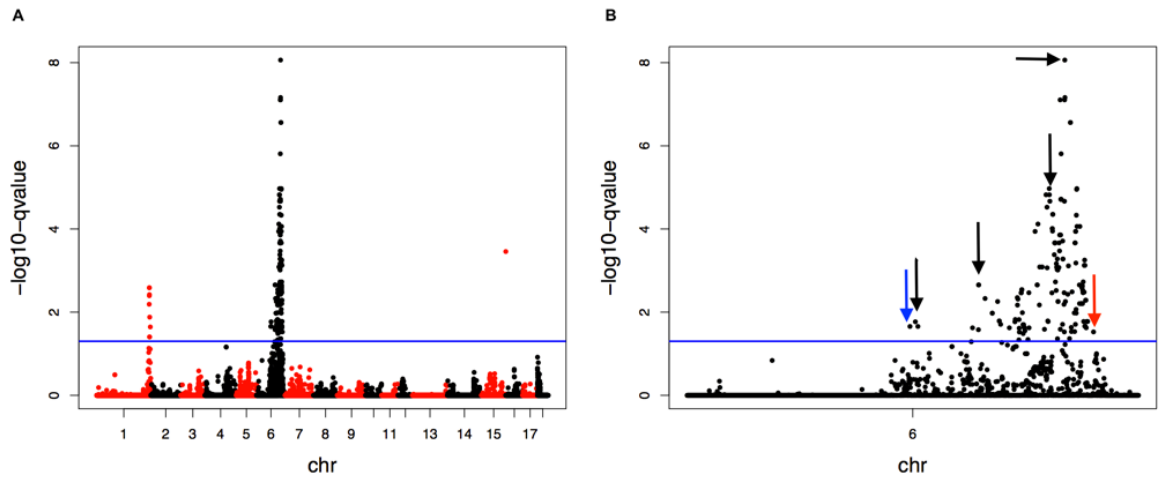

Figure 2.1 Manhattan plots for SNP associations with tenth-rib backfat thickness (BF10). Model (2.1) was fit with sex, slaughter group and carcass weight as fixed effects (FE). A, All Autosomes. B, Chromosome 6. Blue and red arrows point to the smallest and biggest physical position, respectively, of SNP associated with BF10. Black arrows point to four potential QTL peaks. Genome-wide significance threshold is shown with the blue line. False Discovery Rate (FDR<0.05)

Confidence Interval in the real dataset

The results for the parametric and non-parametric approaches are shown in Table 2.1. The non-parametric centered CI has the shortest length (Table 2.1), followed by the parametric method (Table 2.1). Furthermore, the non-parametric non-centered CI is the longest (Table 2.1). Next it is important to determine if the coverage of these CI are on par with the nominal 95% level, and we do so in the next sections.

Table 2.1 Comparison between the different regions defined on chromosome 6 for tenth rib backfat thickness

Method	Lower Boundary [a]	Upper Boundary [a]	Length [a]	Genes [b]
Significant region	75.14	140.56	65.42	502
Parametric CI	129.46	138.30	8.84	42
Non-Parametric Centered CI	129.98	137.78	7.79	38
Non-Parametric Non-Centered CI	126.04	138.26	12.22	64

[a] Results are expressed in Mb. [b] Number of annotated genes in the ENSEMBLE database within those boundaries. (CI) 95% Confidence Interval

Properties of the Confidence Intervals in the simulated dataset

We used simulated datasets to test the property of the CI obtained with of the three methods previously described (PM, NPC, NPNC). Figure 2.2 (A to C) shows the CI position calculated for each of the simulated datasets with each of the methods. These plots were ordered and color coded according the coverage of the CI. Located at the top of the graphic, are those datasets where the actual QTL was covered by CI produced by the three methods, and at the bottom of the graphics is represented the dataset whose QTL were not covered by any method. CI that cover the actual QTL position are presented in black and CI that do not cover the QTL position are presented in red. None of the methods applied to dataset 1 produced a CI that covered the true position of the QTL (Fig. 2.2 A to C red arrows). For the rest of the datasets, at least one method produced CI that covered the actual QTL position.

The QTL region associated with BF10 in this study has more significant SNP downstream from the peak than upstream (Fig. 2.1B). This means that the persistence of LD is larger downstream from the peak. As a consequence, in any particular simulated dataset, it is more likely to have an estimated QTL peak downstream from the actual QTL position that to have the estimated QTL peak upstream from the actual QTL. Consistent with this, all the CI that did not cover the true QTL position had an upper boundary which was below the actual QTL position (133.88 Mb), following the coverage criteria explained in the materials and methods section.

The anisotropic extent of LD seems to be better captured by CI obtained with the NPNC method (Fig. 2.2 C), which tends to produce CI that extend further downstream from the estimated QTL position. The asymmetry in the CI is an advantage for the NPNC method when the estimated QTL peak position is biased further downstream. For instance, two of the simulated datasets (Fig. 2.2 A to C orange arrow) have an estimated QTL peak at position 118 and 126 Mb, clearly biased downstream from the actual QTL (133.88Mb). Interestingly, the two centered CI (PM, and NPC) calculated for those two datasets did not cover the true QTL peak position, while with CI calculated using the NPNC did cover the actual QTL position. Therefore, the NPNC CI seem to adapt better to the density of significant SNP around the peak position. However, this adaptive characteristic comes at a price in length of the interval, as we explain next.

There are two properties of a CI that should be compared to determine which method is the best one; a) CI should cover the true QTL position at the nominal confidence level, b) CI should be as short as possible. The results of this comparison are shown in Table 2.2, where the average of the 200 CI calculated with the PM and NPC approaches had very similar lengths while the NPNC produced longer CI. Moreover, the 95% CI computed with PM has 96.5 % realized coverage (Table 2.2), but the 95% CI computed with the NPC had a realized coverage of only 89.5%

(Table 2.2). These results pose a question: how is it possible that two methods that produce CI of similar length have such a difference in the coverage? Looking in further detail at the NPC confidence intervals that do not cover the true QTL position, 16 of those had an interval length less than 1Mb centered around a peak at position 132.32Mb, while for those datasets the PM CI were on average 2.056 Mb and covered the true QTL position. On the other hand, for most datasets represented at the top of Fig. 2.2, the CI from the NPC tended to be longer than the PM CI. However, in all those datasets, the estimated QTL position was very close (<1.5Mb) to the actual QTL, so that virtually any interval at least 2 Mb in length centered at the peak contained the true position. The 95% CI obtained with NPNC had 96% realized coverage (Table 2.2), at the expense of, on average, longer intervals.

Table 2.2 Summary of the confidence intervals defined on chromosome 6 for tenth rib backfat thicknes in the simulated dataset

Method	**Lower Boundary** [a]	**Upper Boundary** [a]	**Length** [a]	**Coverage** [b]
Parametric	129.4	137.4	8	96.5
Non-parametric centered	129.3	137.4	8.1	89.5
Non-parametric Length Optimized	125.8	135.2	9.4	96

[a]Results are expressed as the average values for the 200 simulated datasets in Mb. [b] Results are expressed in percentage.

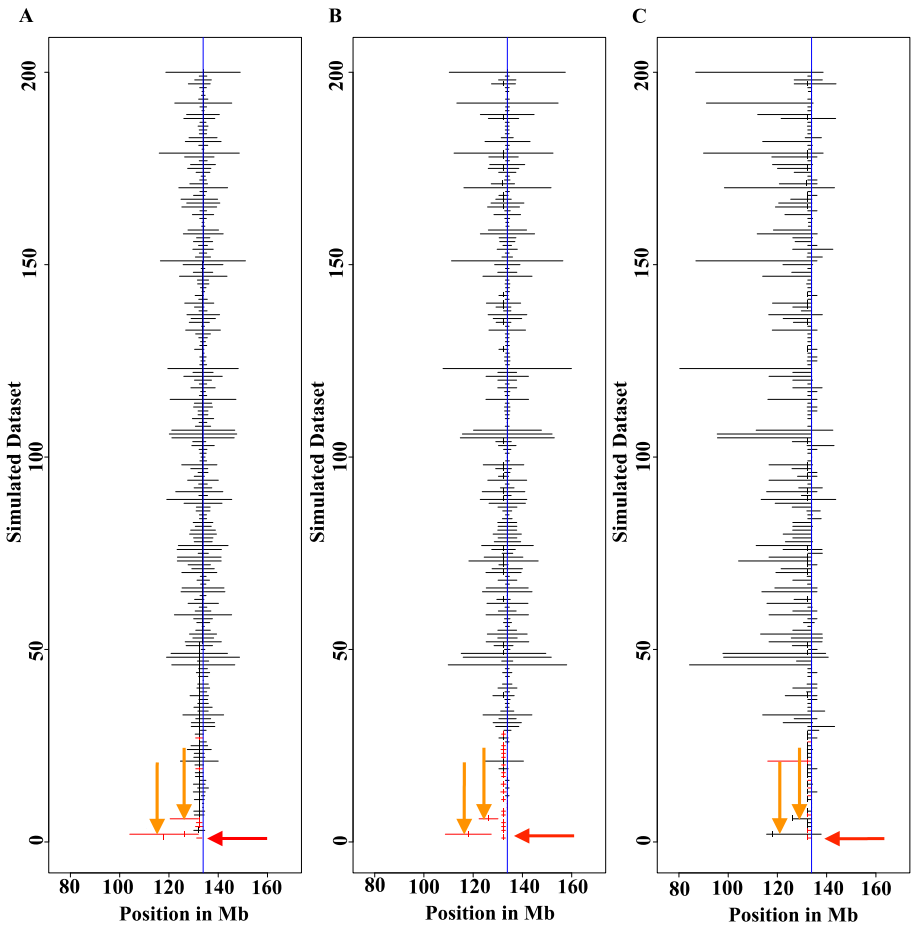

Figure 2.2 Comparison of the coverage between the three methods. The y axes represent each plasmode dataset and on the x axes is the physical position in Mb of the QTL and its confidence limits. The vertical blue line is the physical position of the real QTL (133.88 Mb.). The black line is the CI for the simulated dataset which covers the position of the real QTL. The red line is the CI for the simulated data set which does not cover the real QTL. A, CI calculated using the parametric method. B, CI calculated using the non-parametric centered method. C, CI calculated using the non-parametric non-centered method. Red arrows point to datasets whose CI do not cover the true position of the QTL with the three methods. The orange arrows point to the datasets where the estimated QTL peak was more than 10 MB apart from the true QTL peak

Efficient computational implementation of CI calculations

To make the CI computation faster at step 3 only a subset of markers was fitted, instead of performing a GWA with all the SNP in the dataset. This procedure is in agreement with Visscher et al., (1996), who selected markers to increase speed of the calculations. In our case, the GWA performed in each bootstrap will have half of the animals and much fewer SNP than the analysis performed in Step 1. This overall reduction in size will lower the time and memory demanded by the analysis, allowing efficient computation for several data partitions. A caveat of selecting the SNP in the QTL regions, is to add an extra checkpoint to investigate if that selection of SNP is adequate for the analysis. For instance, if the physical position saved in Step 3 are clustered at the boundaries of the selected region, a wider region is needed.

Conclusion

We tested alternative methods to obtain confidence intervals for QTL positions in BLUP-based GWA. The NPC method failed to provide adequate coverage for the nominal 95% CI. The NPNC and the PM methods produced CI with almost equal coverage, but the length of CI from PM method was on average 20% shorter than the length of CI from NPNC. Neither NPNC nor PM were uniformly better, because in a few cases, the CI obtained with NPNC covered the true QTL position while the PM CI did not, and vice-versa. However, on average, both methods had similar coverage and PM produced shorter intervals. Thus, the PM CI is recommended.

LITERATURE CITED

LITERATURE CITED

Akanno, E. C., Plastow, G., B. W. Woodwards, S. Bauck, H. Okut, X.-L. Wu, C. Sun, J. L. Aalhus, S. S. Moore, S. P. Miller, W. Z., 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. http://doi.org/10.2527/jas.2013-7374

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. http://doi.org/10.1186/1471-2156-14-8

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate : a Practical and Powerful Approach to Multiple Testing When researchers tend to select pursuing multiple the (statistically) and support of conclusions . An unguarded use in a greatly results of single-inference inc. J.R Stat. Soc.B 57:289–300. http://doi.org/10.2307/2346101

Bernal Rubio, Y. L., J. L. Gualdrón Duarte, R. O. Bates, C. W. Ernst, D. Nonneman, G. A. Rohrer, A. King, S. D. Shackelford, T. L. Wheeler, R. J. C. Cantet, and J. P. Steibel. 2015. Meta-analysis of genome-wide association from genomic prediction models. Anim. Genet.:36–48. http://doi.org/10.1111/age.12378

Chen, C. Y., I. Misztal, I. Aguilar, S. Tsuruta, the Meuwissen, S. E. Aggrey, T. Wing, and W. M. Muir. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotyping. J. Anim. Sci. 89:23. http://doi.org/10.2527/jas.2010-3071

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. http://doi.org/10.3389/fgene.2011.00018

Edwards, D. B., C. W. Ernst, N. E. Raney, M. . Doumit, M. D. Hoge, and R. O. Bates. 2008. Quantitative trait loci mapping in an F2 Duroc x Pietrain resource population: II. Carcass and meat quality traits. J. Anim. Sci. 86:254:266. http://doi.org/10.2527/jas.2006-626

Forneris, N. S., A. Legarra, Z. G. Vitezica, S. Tsuruta, I. Aguilar, I. Misztal, and R. J. C. Cantet. 2015. Quality control of genotypes using heritability estimates of gene content at the marker. Genetics 199:675–81. http://doi.org/10.1534/genetics.114.173559

Goddard, M. E., and B. J. Hayes. 2007. Genomic selection. J.Anim.Breed.Genet 124:323–330. http://doi.org/10.1080/09064700801959395

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. http://doi.org/10.1186/1471-2156-14-38

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. http://doi.org/10.1186/1471-2105-15-246

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433–443. http://doi.org/10.3168/jds.2008-1646

Hayes, B. J. 2013. Overview of statistical methods for genome-wide association studies. C. Gondor, B. van der Werf, & B. J. Hayes, editors, Genome-wide association studies and genomic prediction. Human Press, New York, NY. p. 156-157.

Janss, L., G. de los Campos, N. Sheehan, and D. Sorensen. 2012. Inferences from genomic models in stratified populations. Genetics 192:693–704. http://doi.org/10.1534/genetics.112.141143

Kim, J. 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. http://doi.org/83/6/1229

Malek, M., J. C. M. Dekkers, H. K. Lee, T. J. Baas, K. Prusa, E. Huff-Lonergan, and M. F. Rothschild. 2001. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. II. Meat and muscle composition. Mamm. Genome 12:637–645. http://doi.org/10.1007/s003350020019

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. http://doi.org/11290733

Ramos, A. M., R. P. M. A. 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. http://doi.org/10.1371/journal.pone.0006524

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478. http://doi.org/10.1186/1471-2164-15-478

Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. 2003.

Storey, J. D. 2002. A direct approach to false discovery rates. J. R. Stat. Soc. Series B. Stat. Methodol.:479–498. http://doi.org/10.1111/1467-9868.00346

VanRaden, P. M. 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91:4414–4423. http://doi.org/10.3168/jds.2007-0980

Visscher, P. M., R. Thompson, and C. S. Haley. 1996. Confidence Intervals in QTL Mapping by Bootstrapping. Genetics 143:1013–1020.

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. http://doi.org/10.1017/s0016672312000274

CHAPTER THREE

Genome-wide association study in an F2 Duroc x Pietrain resource population for economically important meat quality and carcass traits.[1]

S. Casiró*, D. Velez-Irizarry*, C.W. Ernst*, N. Raney*, R.O. Bates*, M.G. Charles* and J.P. Steibel*†[2]

*Department of Animal Science, Michigan State University, East Lansing

†Department of Fisheries and Wildlife, Michigan State University. East Lansing

[1]This research was financially supported by the National Pork Board Grant no. 11–042. Partial support was also provided by US Pig Genome Coordination funds and Fulbright Scholarship

[2]Corresponding author: steibelj@msu.edu

ABSTRACT

Meat quality is essential for consumer acceptance, it ultimately impacts pork production profitability and it is subject to genetic control. The objective of this study was to map genomic regions associated with economically important meat quality and carcass traits. We performed a genome-wide association (GWA) analysis to map regions associated with 38 meat quality and carcass traits recorded for 948 F2 pigs from the Michigan State University Duroc x Pietrain resource population. The F0, F1 and 336 F2 pigs were genotyped with the Illumina Porcine SNP60 BeadChip, while the remaining F2 pigs were genotyped with the GeneSeek Genomic Profiler for Porcine LD chip, and imputed with high accuracy (r^2=0.97). Altogether the genomic dataset comprised 1015 animals and 44,911 SNP. A Gaussian linear mixed model was fitted to estimate the breeding values and the variance components. A linear transformation was then performed to estimate the marker effects and variances. All the procedures were done using the gwaR package. Type I error rate was controlled at a False Discovery Rate of 5%. Seven putative QTL found in this study were previously reported in other studies. Two novel QTL associated with tenderness (TEN) were located on SSC3 (135.6:137.5Mb; FDR<0.03) and SSC5 (67.3:69.1Mb; FDR<0.02). The QTL region identified on SSC15 includes the previously reported candidate gene, Protein Kinase AMP-activated γ 3-subunit gene (*PRKAG3*), which has been associated with 24-h pH (pH24), drip loss (DL) and cook yield (CY). Also, novel candidate genes were identified for TEND in the region on SSC5 [A Kinase (PRKA) Anchor Protein 3 (*AKAP3)*], and for tenth rib backfat thickness (BF10) [Carnitine O-Acetyltransferase (*CRAT*)] in SSC1. The *PRKAG3* gene has been proposed as a candidate gene for meat quality QTL on SSC15. However, there are no SNP for this gene on the chip used, thus we genotyped the animals for two non-synonymous variants (*I199V* and *T30N*). We then performed a GWA

conditioning on the genotype of both SNP and *I199V* was associated with pH24, DL, protein content (PRO) and CY (P<0.004) and T30N was associated with Juiciness, TEND, shear force, pH24, PRO and CY (P<0.04). Finally, we performed a GWA conditioning on the genotype of the SNP peak detected in this study and T30N remained associated only with PRO (P<0.02). Therefore, in this study we identified two novel QTL regions, suggest two novel candidate genes, and conclude that other SNP in PRKAG3 or a variant(s) of another nearby gene(s) explain the observed associations on SSC15 in this population.

Key words: genome-wide association, meat quality and carcass traits, PRKAG3, swine

INTRODUCTION

One of the primary goals of the industry is to produce the highest quantity of meat at a lower cost, thus, they improved the lean growth. However, selecting for lean growth resulted in lower meat quality (Lonergan et al., 2001). Pork quality is judged by consumers, and whether or not they buy the product will depend on their preferences and perception of the meat. Thus, there is a need to improve meat quality in genetic improvement programs.

Genomic selection is widely used in different livestock species to predict genomic breeding values or GEBV (Meuwissen et al., 2001; Goddard & Hayes, 2007,Hayes et al., 2009; Chen et al., 2011; Wang et al., 2012; Akanno et al., 2014). GEBV prediction is usually followed by a Genome-Wide Association (GWA) analysis to identify SNP associated with traits under study. With the development of the Illumina PorcineSNP60 BeadChip (Ramos et al., 2009) several GWA studies were performed for many traits in different pig populations including reproductive traits (Becker et al., 2013), fat-related traits (Ros-Freixedes et al., 2014; Kim et al., 2015), meat

quality and carcass composition (Becker et al., 2013; Ma et al., 2013; Nonneman et al., 2013; Uimari et al., 2013; Sanchez et al., 2014; Bernal Rubio et al., 2015a) and growth traits (Wang et al., 2015, Gualdrón Duarte et al., 2016). Several genomic regions have been characterized for their association with economically important traits and some candidate genes have been proposed for some of these traits. A region on SSC15 identified for many populations to be associated with meat quality traits includes the gene Protein Kinase AMP-activated γ 3-subunit (PRKAG3). A dominant allele (R200Q) PRKAG3 was associated with glycogen content in skeletal muscle affecting the meat quality (Milan et al., 2000). Furthermore, while R200Q is only segregating in the Hampshire breed, other non-synonymous variants in PRKAG3, such as I199V and T30N were associated with improved meat quality in other breeds (Ciobanu et al., 2001). However, none of these variants are included in the Illumina PorcineSNP60 BeadChip. The goal of this study was to use the Michigan State University Duroc X Pietrain Pig Resource Population to identify new genomic regions containing Single Nucleotide Polymorphism (SNP) associated with economically important meat quality and carcass traits. Additionally, we genotyped variants I199V and T30N in PRKAG3 and we tested them for association with meat quality traits.

MATERIALS AND METHODS

Ethical Statement

Animal protocols were approved by the Michigan State University All University Committee on Animal Use and Care (AUF# 09/03-114-00).

Population and phenotypes

The experimental population consisted of an F2 Duroc x Pietrain cross created at Michigan State

University and extensively described by Edwards et al. (2008). Briefly, the F0 generation consisted of four Duroc sires and 15 Pietrain dams, the F1 included 50 dams and six boars and the F2 generation comprised 954 pigs of both sexes. A total of 38 meat quality and carcass composition traits were recorded on the F2 individuals (Edwards et al., 2008). Descriptive statistics of each phenotype analyzed in this study are shown in Table S.1. The sensory panel traits tenderness (TEN) and overall tenderness (OT) were treated as a single trait (Tenderness), due to their high phenotypic correlation ($r>0.97$, results not shown).

Genotyping and genotype editing

Two SNP chips of different densities were used to genotype the experimental population (Gualdrón Duarte et al., 2013). The entire F0, F1 and 336 F2 animals were genotyped with the Illumina PorcineSNP60 BeadChip (Ramos et al., 2009) that contains approximately 62,000 SNP. The remaining 612 F2 animals were genotyped at lower density (8,836 tag SNP) using the GeneSeek Genomic Profiler for Porcine LD (GGP-Porcine LD, GeneSeek a Neogen Company, Lincoln, NE; Badke et al., 2013).

First, 2,277 SNP from the 60K chip were removed due to having genotypes missing in all animals. After that, we performed a model-based Mendelian consistency checking following Forneris et al. (2015) removing 1,155 SNP whose segregation pattern did not fit the expected Mendelian inheritance rules ($p<8.4E^{-8}$). After these minimal edits we proceeded to impute all missing genotypes in the 60K chip , and those not scored in the low density chip (Badke et al., 2013). The imputation was done using the software FImpute (Sargolzaei et al., 2014), with default settings. During the imputation, SNP-specific imputation accuracies were estimated and 712 SNP were removed due to low imputation accuracy ($r^2<0.64$). Overall imputation accuracy

of remaining SNP was r^2=0.97. Furthermore, 101 SNP which had more than 10% of missing genotypes, were also removed after the imputation because their imputation accuracy could not be reliably estimated. The imputation algorithm flagged 147 SNP and 9 animals that contained further genotyping errors or inconsistencies. Those SNP and individuals were edited out of the genotype database. The final dataset comprised 948 F2 animals with phenotypic records and genotypes for 44,911 SNP.

Genotyping of I199V and T30N

I199V and T30N are known non-synonymous substitutions from the Protein Kinase AMP-activated γ 3-subunit gene (PRKAG3) associated with meat quality traits such as pH 24 hours post-mortem (pH24), drip loss (DL) and cook yield (CY) (Milan et al., 2000; Ciobanu et al., 2001). Custom Taqman genotyping assays were developed for the I199V and T30N SNP (Table S. 2). All F1 animals where genotyped and all F2 animals were either genotyped or inferred from informative homozygous F1 parents.

Statistical analysis

For the association study, a GBLUP based GWAS analysis was performed (Gualdrón Duarte et al., 2014). First an animal-centric Gaussian linear mixed model was fit.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{a} + \boldsymbol{e}, \tag{3.1}$$

where $\boldsymbol{y}$ is the vector containing the phenotypes, $\boldsymbol{X}$ is the incidence matrix which relates the individual records with the fixed effects of sex, slaughter group and carcass weight in $\boldsymbol{\beta}$ (Edwards et al., 2008), an exception was the number of ribs trait, which had sex as the only fixed effect, $\boldsymbol{a}$~N (0, $\mathbf{G}\sigma^2_A$) is a vector of random breeding values. The matrix **G= ZZ'** is the marker

based relationship matrix and $\mathbf{Z}$ is the standardized allelic dosages (counts of the "B" allele) matrix (VanRaden, 2008):

$$\boldsymbol{Z}_{ij} = \frac{\boldsymbol{M}_{ij} - 2\,p_j}{\sqrt{m\,(2\,p_j\,(1-p_j))}},$$

where $\boldsymbol{M}$ is the allelic dosages matrix, p_j is the allelic frequency at the marker *j* of the F2 animals, *i* is the i^{th} animal and *m* is the number of markers. The marker based relationship matrix was used to account for population substructure (Janss et al., 2012). Finally, $\boldsymbol{e} \sim N(0, \boldsymbol{I}\sigma^2_e)$ is a vector of residuals; where the variance covariance $\boldsymbol{I}$ is an identity matrix. Gualdrón Duarte et al. (2014) and Bernal Rubio et al. (2015b), showed an equivalence between a test based on an animal-centric model (equation 3.1) and a test based on a SNP effects fixed model.

Furthermore, we estimated the SNP effect and its variance with a linear transformation of the estimated breeding values ($\hat{\boldsymbol{a}}$) following Gualdrón Duarte et al. (2014):

$$\hat{\boldsymbol{g}} = \boldsymbol{Z}'\boldsymbol{G}^{-1}\,\hat{\boldsymbol{a}}\,, \tag{3.2}$$

$$Var(\hat{\boldsymbol{g}}) = \boldsymbol{Z}'\boldsymbol{G}^{-1}\boldsymbol{Z}\,\sigma^2_A - \boldsymbol{Z}'\boldsymbol{G}^{-1}\boldsymbol{C}^{aa}\,\boldsymbol{G}^{-1}\,\boldsymbol{Z}, \tag{3.3}$$

where all the terms have been described previously in equation (3.1) except $\boldsymbol{G}^{-1}$ which is the inverse of the marker based relationship matrix and $\boldsymbol{C}^{aa}$ which is the portion of the inverse of the mixed model equations that correspond to animal effects.

We standardized the SNP effects to obtain the test statistics:

$$t_j = \frac{\hat{g}_j}{\sqrt{var(\widehat{g_j})}}, \tag{3.4}$$

where the subscript *j is* the j^{th} SNP. The p-values were obtained from the Gaussian distribution:

$$p - value_j = 2(1 - \phi\,(|t_j|)), \tag{3.5}$$

were, $\emptyset(x)$ is the cumulative density function of the normal distribution. All computations were implemented with the gwaR package (https://github.com/steibelj/gwaR.git) in the R environment (https://cran.r-project.org). A False Discovery Rate (Benjamini & Hochberg, 1995; Storey, 2002; Storey & Tibshirani, 2003) of 5% was used as significance criteria to control for multiple tests.

Number of QTL per genomic region and confidence interval of peak position

In some cases, a significant genomic region seemed to include multiple Quantitative Trait Loci (QTL) peaks. For instance, the trait tenth-rib backfat thickness (BF10) has three QTL regions on chromosomes SSC1, SSC6 and SSC15 (Fig. 3.1 A). However, the region in SSC6 shows four putative QTL peaks (red arrows in Fig. 3.1 B). In those cases, knowing the number of QTL per genomic region is necessary to calculate the 95% confidence interval of each peak. To determine the number of QTL peaks, we repeated the GWA scan but we included the genotypes of the peak SNP as a fixed effect (Fig. 3.1 C). If after fitting a SNP as a fixed effect, all other SNP in the region do not exhibit significant association, this is a strong indication of a single QTL peak in the region (Fig. 3.1 D). Additionally, if a single SNP association in another chromosome vanished when a SNP genotype in another region was included in the model (compare Fig. 3.1 A to C, green arrow), this is an indication that a single SNP is not in LD with neighboring SNP but it is in LD with many SNP on another chromosome. We did not consider such SNP for further analysis.

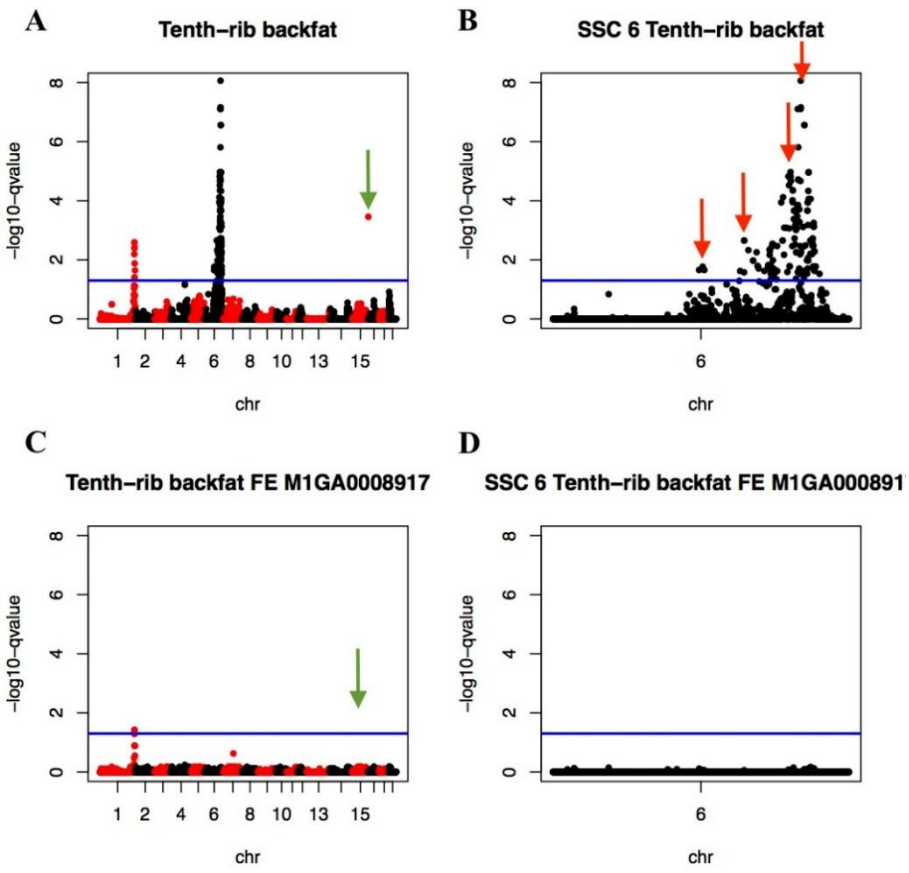

Figure 3.1 Manhattan plot for SNP association with tenth-rib backfat fitting model (1) with sex, slaughter group and carcass weight as fixed effects (FE). (A): Considering Autosomes. (B): Considering SSC6. (C): Considering autosomes and using the marker M1GA0008917 as a FE. (D): Considering SSC6 and using the marker M1GA0008917 as FE, red arrows point at four possible QTL in SSC6, green arrow shows peak in SSC15. Genome-wide significance threshold is shown with the blue line (FDR<0.05)

The 95% CI of the QTL peak position for each genomic region was computed using a method proposed by Hayes (2013). The algorithm based on cross validation comprises the following steps:

Step 1: Perform a GWA and obtain QTL peak position: p (corresponding to the SNP with the smallest p-value in a genomic region).

Step 2: Randomly assign the animals in the dataset to two sets (x_1, x_2).

Step 3: Perform a GWA analyses for x_1 and x_2, separately.

Step 4: Store the physical position of the most significant SNP in the region from x_1 and x_2 into the vectors v_1 and v_2 respectively.

Step 5: Repeat n times in order to fill v_1, v_2.

Step 6: Calculate the Standard Error of the difference between the positions in v_1 and v_2 following:

$$se(\bar{x}) = \sqrt{\frac{1}{4n}\sum_{k=1}^{n}(v_{1k} - v_{2k})^2}$$

where k corresponds to the position of the most significant SNP in the k^{th} repetition.

Step 7: Calculate the 95% CI using p from step 1, z value which corresponds to 97.5 percentile in a normal Gaussian distribution (z=1.96) and the $se(\bar{x})$ from step 6:

$$CI = p \pm z_{97.5} se(\bar{x})$$

Percentage of total variance explained by the SNP

The percentage of variance explained by the peak SNP was calculated by re-fitting model (3.1) including the genotypes ($\boldsymbol{Z_{peak}}$) of the most significant SNP as a fixed effect (already described in previous section). The estimated effect of this marker was used to estimate the variance accounted for by the marker using equation (3.6).

$$\widehat{Var(q)} = b^2 var(\boldsymbol{Z_{peak}}) \tag{3.6}$$

where $\widehat{Var(q)}$ is the estimated variance associated with the marker effect, $\boldsymbol{Z_{peak}}$ is the genotype of the most significant marker, and b is the estimated effect of the marker.

The percentage of variance explained by the marker in study can be calculated:

$$\frac{\widehat{Var(q)}}{\widehat{\sigma_A^2} + \widehat{\sigma_e^2} + \widehat{Var(q)}} \tag{3.7}$$

Where $\widehat{Var(q)}$ was calculated in (3.6), $\widehat{\sigma_A^2}$ is the estimated additive genetic variance and $\widehat{\sigma_e^2}$ is the estimated error variance using the model explained in this section. The results obtained with this procedure were roughly equal to the computationally more involved methods to estimate percentage of variance explained by a QTL presented elsewhere (Hayes et al., 2010; Gualdrón Duarte et al., 2014)

Statistical analysis for SNP in I199V and T30N

For the association study of the two SNP, I199V and T30N from the PRKAG3 gene, an animal-centric Gaussian linear mixed model was fitted:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{I199V}\beta_2 + \boldsymbol{T30N}\beta_3 + \boldsymbol{a} + \boldsymbol{e}, \quad (3.8)$$

where $\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}$ $\boldsymbol{a}$ and $\boldsymbol{e}$ were previously described in model (3.1) and $\boldsymbol{I199V}$ and $\boldsymbol{T30N}$, are the vector of genotypes of both non-synonymous variants, expressed as the allelic dosages; counts of G (I199V) and C (T30N) alleles and β_2 and β_3 are the fixed effects of the markers, respectively. With this model we tested fixed SNP effects and we performed a GWA by transforming the animal effects as described in equations 3.2 to 3.5.

Additionally, for the association study of the two variants conditional on the peak SNP genotype on SSC15 we fitted:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{I199V}\,\beta_2 + \boldsymbol{T30N}\,\beta_3 + \boldsymbol{Z_{peak}}\,\beta_4 + \boldsymbol{a} + \boldsymbol{e}, \quad (3.9)$$

where $\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{a}$ and $\boldsymbol{e}$ were previously described in model (3.1), β_2, β_3 $\boldsymbol{I99V}, \boldsymbol{T30N}$ were previously described in model (3.8) and $\boldsymbol{Z_{peak}}$ is a vector containing the genotypes of the SNP with the smallest q-value on SSC15 after fitting model (3.1) and β_4 is the marker fixed effect. Furthermore, after fitting equation 3.8, we estimated the marker effect and variance components doing a linear transformation (equations 3.2 to 3.5). The Type I Error Rate considered for this

analysis was $\alpha = 0.05$ for testing the fixed effects. As previously mentioned, to account for multiple testing in the GWA scan (equation 3.8 only) a genome-wide significant threshold was determined using FDR=5%.

Post -GWAS analysis

The genomic region used for the identification of candidate genes was defined by the 95% CI constructed around the peak. Annotated genes within those genomic regions were identified with the ENSEMBL annotation of *Sus scrofa* 10.2.83 (December 2015) assembly (http://useast.ensembl.org/biomart/martview/). We used the PigQTL database Release 28 December 2015 database (Hu et al., 2015) to approximately locate the low resolution linkage QTL detected in previous studies.

RESULTS

The genome-wide association analysis found 20 putative QTL (FDR<0.05) for 15 traits. The Manhattan plots for the significant GWA analyses can be seen in Fig. 3.2.

Every region that is reported in Table 3.1 showed a single QTL peak. Some single SNP were significant (FDR<0.05), but they were not studied in more detail because they showed LD with distant QTL peaks as explained in the methods section

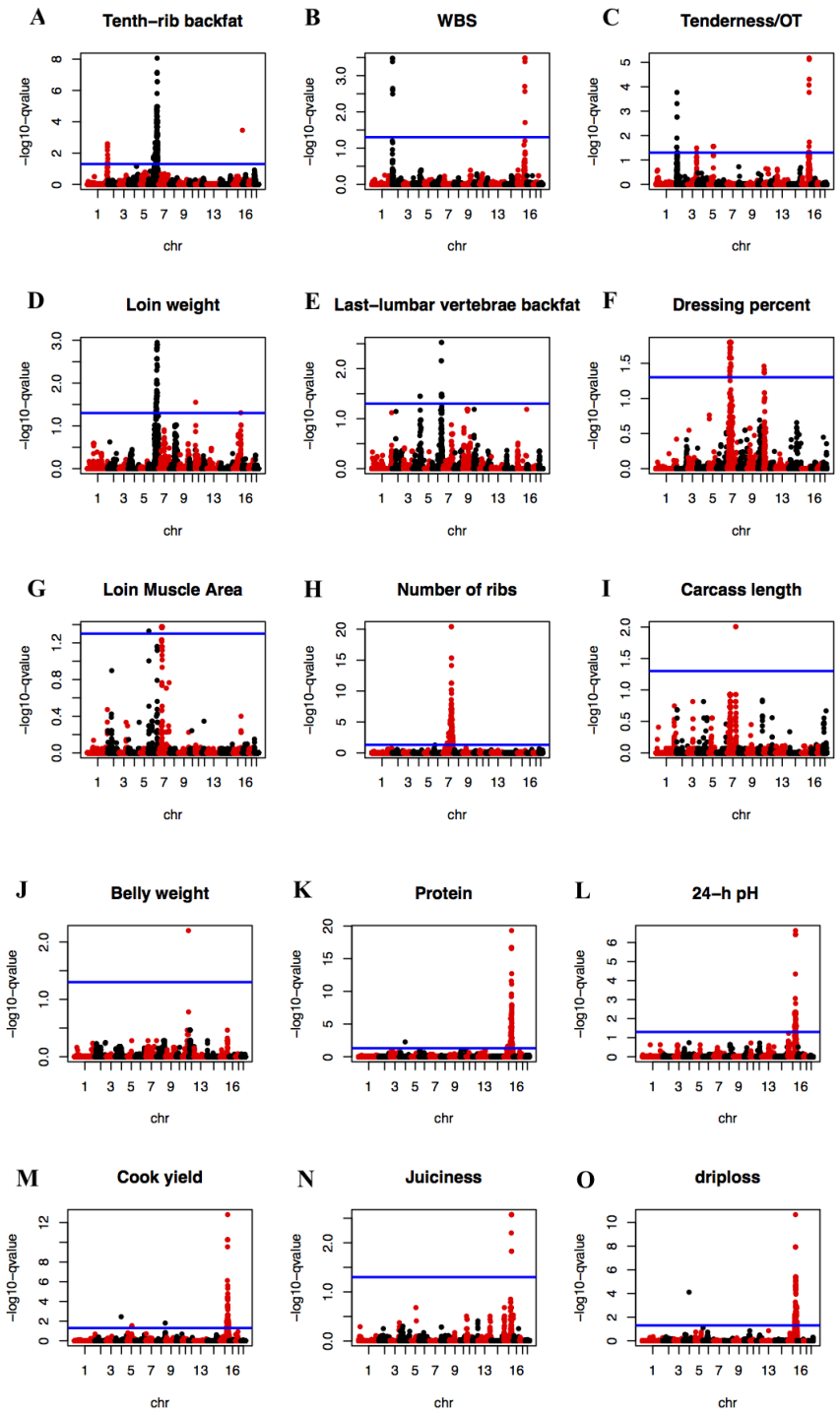

Figure 3.2 Manhattan plots for evaluated traits exhibiting significant QTL. A, Tenth rib backfat; B, WBS; C, Tenderness/OT (Overall Tenderness); D, Loin weight; E, Last lumbar vertebra backfat; F, Dressing percentage; G, Loin muscle area; H, Number of ribs; I, Carcass length; J, Belly weight; K, Protein; L, pH 24 hours post-mortem; M, Cook yield; N, Juiciness; M, Drip loss. $-Log_{10}$(Q-value) (y-axis) vs. SNP position (ordered within chromosome on the x-axis). The blue horizontal line marks the genome-wide significance threshold (FDR=5%)

Table 3.1 Summary of the Quantitative Trait Loci regions

Trait	Marker	SSC[a]	Pos[b]	q-value[c]	Effect[d]	% Var [e]	95% lower pos [f]	95% upper pos [g]	Genes in Region [h]
Tenth-rib backfat	ASGA0008074	1	305.0	2.5E-03	-	3.0	302.9	307.1	80
WBS	M1GA0002229	2	2.9	3.3E-04	-	4.3	1.0	4.9	196
Tenderness/OT	H3GA0005676	2	5.9	1.69E-04	+	4.8	4.0	7.7	
Tenderness/OT	H3GA0011017	3	136.5	3.21E-02	+	3.4	135.6	137.5	4
Tenderness/OT	H3GA0016570	5	68.2	2.77E-02	+	3.2	67.3	69.1	16
Tenth-rib backfat	M1GA0008917	6	133.9	8.65E-09	-	12.6	129.5	138.3	
Loin weight	ASGA0029651	6	133.9	1.10E-03	-	6.5	127.6	140.2	64
Last-lumbar vertebrae Backfat	ALGA0122657	6	136.1	2.90E-03	+	5.1	131.4	140.8	
Dressing Percent	MARC0033464	7	35.2	1.50E-02	+	5.4	34.0	36.3	96
Loin Muscle Area	ASGA0032589	7	36.4	4.60E-02	-	4.5	32.7	40.0	
Carcass length	ASGA0035535	7	104.0	9.80E-03	-	4.9	103.7	104.4	57
Number of ribs	ALGA0043983	7	104.4	3.93E-12	-	11.7	102.5	106.2	
Belly weight	M1GA0015491	11	84.4	6.30E-03	-	4.5	83.6	85.2	10
Juiciness	MARC0047188	15	135.2	2.60E-03	+	4.1	133.4	137.0	
Tenderness/OT	MARC0047188	15	135.2	6.71E-06	+	7.2	133.8	136.6	
WBS	MARC0047188	15	135.2	3.3E-04	-	5.6	134.1	136.4	
24-h pH	MARC0093624	15	135.5	2.36E-07	+	9.4	134.0	137.1	59
Drip loss	MARC0093624	15	135.5	2.20E-11	-	12.8	134.9	136.1	
Protein	MARC0093624	15	135.5	4.95E-20	+	21.0	135.1	135.9	
Cook yield	MARC0093624	15	135.5	1.55E-13	+	14.9	135.2	135.8	

[a] Sus Scrofa Chromosome. [b] Peak position expressed in Megabase. [c] SNP q-value. [d] Additive effect of the SNP [e] Percentage of variance explained by the SNP. [f] Lower boundary of the 95% CI in Megabase. [g] Upper boundary of the 95% CI in Megabase. [h] Number of annotated genes in the region.

A SNP located at position 305 Mb on SSC1 was significantly associated with BF10. This marker explained 3% of the trait variance, with the B allele associated with less backfat thickness. The genomic region defined by the 95% CI comprised 80 genes (Table S.3). A putative candidate gene is described in the discussion section.

Markers in a region on SSC2 (1.0 MB-7.7MB) were associated with two related traits: Warner-Bratzler shear force (WBS) and TEN. For WBS SNP M1GA0002229 is the most significant marker (FDR<0.001) explaining 4.3% of the phenotypic variance. The B allele was associated with lower values of the trait (B allele = less force needed to cut the chop), and the A allele was fixed in Duroc grandparents. For TEN, the QTL peak corresponded to H3GA0005676, which was 3 Mb upstream from the WBS peak, but in linkage disequilibrium (LD) with M1GA0002229 (r^2=0.41) Furthermore, the 95% confidence interval for the two QTL peaks overlapped each other (Table 3.1). The B allele of the most significant marker had a positive effect on the trait (B allele = more tender chop) and the genotypes for this SNP explained 4.8% of the phenotypic variation. The whole QTL region (defined by the 95% CI) contains 196 genes (Table S.4), including three genes that were previously proposed as putative candidates: cystatin E/M (*CST6*; SSC2: 5.395 to 5.396 Mb), cathepsin W (*CTSW*; SSC2:5.550 to 5.554 Mb) and calpain-1 catalytic subunit (*CAPN1*; SSC2: 6.12 to 6.15).

In addition to the peak on SSC2, there were other association peaks for TEN. For instance, a peak at 136.5 Mb (SNP H3GA0011017) on SSC3 (FDR<0.05) explained 3.4% of the phenotypic variance for the trait (Table 3.1). For the SNP H3GA0011017, the B allele was associated with more tender meat, and the 95% CI for the peak extended from 135.6 to 137.5 Mb (Table 3.1). This segment contained only 3 genes and one uncharacterized protein: membrane bound O-acyltransferase domain containing 2 (*MBOAT2*; SCC3:135.64-135.68Mb), ribonuclease L

(*RNASEL*; SSC3:135.75-135.77Mb), DNA-binding protein inhibitor ID-2 (*ID2*;135.801-135.804Mb). Another novel QTL peak for TEN at position 68.2Mb on SSC5 (FDR<0.05) coincided with the SNP H3GA0016570 and it explained 3.2% of the phenotypic variance (Table 3.1). In this case the B allele was associated with increased tenderness, and it was fixed in F0 Duroc sires (*f*(B)=1), but was segregating in F0 Pietrain dams (*f*(B)=0.86). Furthermore, the 95% confidence interval region comprised 16 genes (Table S.5), including A kinase (PRKA) anchor protein 3 (*AKAP3*; SSC5: 68 to 68.02 Mb) located 0.2 Mb downstream the QTL peak. This gene is expressed in the longissimus dorsi muscle from the pigs in this population (data not shown). The possible relationship between this gene and TEN is further discussed in the next section. Phenotypes of three traits, BF10, last-lumbar vertebrae backfat thickness (LLBF) and loin weight (LW), were associated with SNP genotypes on SSC6. The 95% CI for those QTL peaks overlapped each other and defined a large QTL region extended between 127.6 and 140.8 Mb of SSC6 (Table 3.1). The B alleles of ASGA0029651, M1GA0008917 and ALGA0122657 were associated with a lighter LW, reduction of BF10 and increased LLBF, respectively. With regard to BF10, genotypes of SNP M1GA0008917 explained 12.6 % of the phenotypic variance. To find a SNP explaining such large proportion of variance is unusual and we performed further analyses for this SNP. The SNP M1GA0008917 was fixed in Duroc and segregating in Pietrain grandparents (*f*(B)=0 and *f*(B)=0.86 respectively) and we observed that the 25% of animals with the least backfat thickness had genotype BB with frequency 0.36 (both alleles came from the Pietrain granddames), while only 7% of the 25% of animals with the thickest backfat were BB for M1GA0008917. The Pietrain breed is well known to have less backfat than Duroc, thus these frequencies are consistent with a SNP where a common allele of Pietrain origin exerts a strong

effect on the phenotype. The genomic region defined for SSC6 encompasses 64 genes (Table S.6). Some candidate genes are further discussed in the next section.

Chromosome 7 had two regions associated with four traits. One region spanning from 32.7 Mb to 40 Mb contained 96 genes (Table S.7) and two markers (MARC0033464, ASGA0032589) associated (FDR<0.05) with dressing percentage (DRESS%) and loin muscle area (LMA), respectively. The B allele of the SNP MARC0033464 is associated (FDR<0.02) with higher DRESS%, while the B allele of ASGA0032589 is associated (FDR<0.05) with a smaller LMA (Table 3.1). The B allele of ASGA0032589 is fixed in Pietrain animals. Another region on SSC7 located between 102.5Mb and 106.2 Mb contained SNP associated with number of ribs (NR) and carcass length (CL). In particular, the B allele of ALGA0043983 located in this region was associated with fewer ribs.

Furthermore, genotypes of a nearby SNP ASGA003535 were associated with CL. These two markers are in LD (r^2=0.6). The SNP ALGA0043983 explained 11.7% of the phenotypic variance for NR. This large proportion of explained variance was further investigated. Among the animals with 16 or more ribs, genotype AA was predominant (*f*(AA)= 0.64), but among animals with 13 ribs or fewer, genotype AA was the least common (*f*(AA)=0.10). This region contains 57 genes (Table S.8) and the QTL peak is located 1 Mb downstream of the gene vertebrae development associated gene (*VRTN*) that may have a substantial impact on thoracic vertebrae development, affecting the discrete trait, NR.

On SSC11, SNP located between 83.6 Mb and 85.2 Mb were associated with belly weight (FDR<0.007). The B allele of the peak SNP (M1GA0015491) was associated with lower belly weight and it explained 4.5 percent of the phenotypic variation (Table 3.1). Ten genes are located in the 95% CI for this QTL peak: myosin XVI (*MYO16*; SSC11:83.4-83.5Mb), collagen, type IV

alpha 1 and alpha 2 (*COL4A1*; SSC11:84.37-84.43Mb and *COL4A2***;** SSC11:84.61:84.67), RAB20, member RAS oncogene family (*RAB20***;** SSC11:84.68:84.7Mb) and testis expressed 29 (*TEX29***;** SSC11:85.00:85.02Mb), and five uncharacterized proteins. Potentially relevant genes are discussed in the next section.

Chromosome 15 has a QTL region that contains markers associated with seven traits. Even though the peak SNP varied across the seven traits, the 95% confidence interval of the QTL peaks overlapped each other. Thus, we considered a single genomic region spanning from 133.4 to 137.1 Mb. The QTL peak (FDR<0.003) for juiciness (JUI), TEN and WBS corresponds to MARC0047188 (Table 3.1) where the B allele is associated with juicier and more tender meat. The marker MARC0093624 at position 135.5 Mb is associated with four traits (FDR<2.36E-07), where the B allele is associated with higher pH at 24 hours post-mortem, protein content (PRO) and CY, and with reduced drip loss (Table 3.1). The MARC0093624 SNP had breed specific allelic frequencies (Duroc: *f*(B)=1, Pietrain *f*(B)=0.86) and the percentage of variance explained by this SNP for these traits varies from 9.4% to 21% (Table 3.1). This could be explained by the genotype frequencies of phenotypically extreme animals, e.g.: the top 25% of animals with more protein were practically all of BB genotype (frequency= 95%), whereas only 38% of the bottom 25% (animals with less protein content) had BB genotypes, the most common genotype for the bottom 25% of animals with less protein content was AB (frequency = 58%). Similarly, 93% of the top 25% of animals for CY (higher cook yield) had genotype BB, while for the bottom 25% (lower cook yield), genotype BB had a frequency of 45%. Furthermore, 92% of the animals with less drip loss (bottom 25%) had a BB genotype and only 49 % of the top 25% (animals with more drip loss) were BB. Finally, this genomic region contains 59 genes (Table S.9) including

the Protein Kinase AMP-activated γ 3-subunit gene (*PRKAG3*; SSC15: 133.8 Mb), that we discuss in the next section, is comprised in that list.

Finally, we fit the model using equation (3.8) to determine if candidate PRKAG3 SNP were associated with the traits and if the GWA scan would still produce a genome-wide significant association peak in the region when candidate SNP genotypes are included as fixed effects in the model (Table 3.2). We found that the T30N SNP was associated with JUI, TEN, WBS, pH24, PRO and CY ($P<0.01$), while the I199V SNP was only associated with pH24, DL, PRO and CY ($P<0.01$). Furthermore, when genotypes of those candidate SNP were included as fixed effects in the GWA scan, the observed QTL peak was replicated, in some cases, in the exact position of the previous QTL peak and in other cases in a very close position with a SNP that was in high LD ($r^2=0.8$) with the SNP in the original peak.

Table 3.2 Comparison between the Genome-wide Association (GWA) peak and GWA considering I199V and T30N as fixed effects for the traits significant on SSC15

Trait	Peak GWA SNP equation (3.1) [a]	Peak GWA SNP equation (3.8) [b]	q-value peak GWA SNP equation (3.8) [c]	p-value I199V [d]	p-value T30N [d]
Juiciness	MARC0047188	MARC0047188	2.1E-02	1.30E-01	4.10E-02
Tenderness/OT	MARC0047188	MARC0047188	2.6E-04	1.17E-01	1.50E-03
WBS	MARC0047188	MARC0093624	4.2E-03	6.23E-02	1.00E-02
24-h pH	MARC0093624	DIAS0000678	8.8E-06	4.00E-05	3.00E-02
Drip loss	MARC0093624	MARC0093624	1.3E-09	6.50E-05	7.70E-02
Protein	MARC0093624	DIAS0000678	1.7E-14	2.60E-04	7.40E-09
Cook yield	MARC0093624	MARC0093624	3.6E-09	4.80E-03	4.50E-03

[a] Name of the peak SNP when fitting the GWA without the SNP of PRKAG3 as fixed effects.
[b] Name of the peak SNP when fitting the GWA considering I199V and T30N as fixed effects.
[c] q-value of the peak SNP, genome-wide significance level (FDR<0.05).
[d] p-value of PRKAG3 SNP, significant threshold (p<0.05).

After fitting the model using equation (3.9) and testing the significance of previously proposed SNP in PRKAG3 (Table 3.3), the only trait significantly associated with the non-synonymous variant T30N was PRO ($P<0.05$). The rest of the traits were not associated with either of the two non-synonymous variants of PRKAG3 evaluated ($P>0.05$). Moreover, including the peak SNP of the GWA explained an equal or larger proportion of the phenotypic variation than the candidate SNP (Table 3.3). For instance, without including genotype having the peak SNP from the GWA in the model (equation 3.8), the candidate SNP explained anywhere between 0.4 to 7.5% of the phenotypic variance (Table 3.3). However, once the SNP of the QTL peak was included in the model (equation 3.9), it explained from 4 to 18 % of the variation and the candidate SNP explained virtually no phenotypic variation at all (Table 3.3). In summary, we believe that the candidate SNP in PRKAG3 are not responsible for phenotypic variation for these traits in this population and that there must be other SNP in PRKAG3, or in other genes that are in LD with the peak SNP in our GWA

Table 3.3 Comparison of the results for the Single Nucleotide Polymorphism (SNP) peak on SSC15 and the two Protein Kinase AMP-activated γ 3-subunit SNP fitting equations (3.8) and (3.9)

Trait	% Var. explained by peak GWA SNP equation (3.9) [a]	p-value I199V equation (3.9) [b]	% Var. explained by I199V equation (3.8) [c]	% Var. explained by I199V equation (3.9) [d]	p-value T30N equation (3.9) [b]	% Var. explained by T30N equation (3.8) [c]	% Var. explained by T30N equation (3.9) [d]
Juiciness	4.2	9.99E-01	0.4	0.0	7.10E-01	0.7	0.0
Tenderness/OT	6.9	7.85E-01	0.5	0.0	1.59E-01	2.3	0.4
WBS	5.2	7.38E-01	0.7	0.0	4.10E-01	1.2	0.2
24-h pH	9.5	1.39E-01	3.2	0.4	5.71E-01	1.0	0.1
Drip loss	14.0	2.76E-01	3.1	0.2	1.65E-01	0.7	0.4
Protein	18.2	5.65E-01	2.5	0.1	2.00E-02	7.5	1.1
Cook yield	14.1	5.63E-01	2.4	0.1	8.44E-01	2.8	0.0

[a] Percentage of phenotypic variance explained by the peak SNP after performing a GWA using the model which has the SNP peak genotype and SNP PRKAG3 non-synonymous variants as a fixed effect. [b] p-value of the PRKAG3 SNP after fitting the model which has SNP peak and SNP PRKAG3 genotypes as fixed effects. [c] Percentage of phenotypic variance explained by the SNP in PRKAG3 after fitting the model which has SNP peak and SNP PRKAG3 genotypes as fixed effects. [d] Percentage of phenotypic variance explained by the SNP in PRKAG3 after fitting the model with the genotypes of SNP in PRKAG3 as fixed effects.

DISCUSSION

The genomic region associated with BF10 on SSC1 (302.9-307.1 Mb), was previously reported in a Meishan x White composite population using a low resolution linkage map (Rohrer & Keelen, 1998). In this study we replicated the finding in a different population and we map it to a 3.2 Mb region using a physical position map. We found that the B allele of ASGA0008074 is associated with reduced 10^{th} rib backfat thickness, and it is fixed in Pietrain granddams. This is consistent with the reports that Pietrain sired animals have less backfat thickness than Duroc sired animals (Edwards et al., 2003). This region harbors the gene Carnitine O-Acetyltransferase (*CRAT*; SSC1:303.4-303.41 Mb), which is an enzyme that catalyzes a fully reversible exchange of acyl groups between coenzyme A and carnitine without energy consumption (Jogl et al., 2004). A previous study in beef cattle showed that the Barros
ã breed had higher mRNA expression levels of *CRAT* than Alentejana breed in subcutaneous adipose tissue (da Costa et al., 2013). This could be associated with the storage/removal ratio of triacylglycerol (TAG) affecting fat deposition. Further studies should be carried out in order to validate *CRAT* as a candidate gene for this QTL.

Two SNP associated with two traits related to meat tenderness were located at the proximal end of SSC2 (1-7.7 Mb). A QTL peak for WBS (H3GA0005672, 5.90 Mb) has previously been described in a Landrace-Duroc-Yorkshire population (Nonneman et al., 2013). The SNP H3GA0005672 is located 20 Kb upstream from the peak SNP found in this study (H3GA0005676, 5.88Mb). Few studies have included sensory panel phenotypes as we have for our population, and there are no reports of QTL for TEN overlapping the region we identified on SCC2. The genomic region (1-7.7 Mb) contains previously described candidate genes including calpain-1 catalytic subunit (*CAPN1*) (Nonneman et al., 2013; Bernal Rubio et al., 2015a) ,

Cystatin E/M and Cathepsin W (Bernal Rubio et al., 2015a). The CAPN1 gene is considered as a likely candidate gene for this region (Nonneman et al., 2013). Calpain has a crucial role in post mortem changes as meat ages, degrading five key myofibrillar and cytoskeletal proteins which can contribute to post-mortem tenderization processes (Goll et al., 1992; Koohmaraie, 1992; Huff-Lonergan et al., 1996).

A novel QTL region on SSC3 (135.6-137.5Mb) containing SNP associated with TEN was detected. The three genes contained in the region are: *MBOAT2, RNASEL and ID2.* However, there is no evidence of a biological link or apparent biological mechanism connecting these three genes and sensory panel tenderness. Hence, the genetic cause of this association peak remains unknown.

Another novel QTL region containing SNP associated with TEN was identified on SSC5 (67.3-69.1 Mb). Sixteen genes are annotated in this region including the A kinase (PRKA) anchor protein 3 (*AKAP3)* gene that is located 0.2 Mb downstream of the QTL peak. This gene belongs to the AKAPs (A-Kinase anchoring protein) family, which includes proteins that bind to the regulatory subunit of the adenosine monophosphate activated protein kinase (AMPK), also known as PKA (Wong and Scott, 2004). This gene has been studied mainly in sperm, and testicular/ovarian cancer. However, AKAP3 is expressed in skeletal muscle, and it is expressed in longissimus dorsi muscle in pigs in this population. The PKA enzyme plays a crucial role in glucose, glycogen and fat metabolism, and variants in the gene PRKAG3 encoding a regulatory subunit unit have been associated with meat quality traits in swine (Milan et al., 2000; Ciobanu et al., 2001; Ryan et al., 2012; Uimari & Sironen, 2014). Thus, AKAP3 is a potential candidate gene for tenderness traits. However further studies must be carried out to confirm if variants in AKAP3 cause variation in meat tenderness.

The QTL identified at position 127.6-140.8 Mb on SSC6 has been widely studied and well characterized for affecting not only backfat thickness, but also loin weight traits. The BF10 QTL (129.5-138.3Mb) has been previously reported in low resolution linkage analysis in different populations (Malek et al., 2001; Kim et al., 2005) and in this population (Edwards et al., 2008; Choi et al., 2011). In this study we confirmed and mapped it to an 8.8 Mb region of SSC6. The QTL for loin weight has previously been reported using linkage analysis (Edwards et al., 2008; Steibel et al., 2011) and here we confirmed it and mapped it to a 12.6 Mb region of SSC6 (127.6-140.2 Mb). The last lumbar vertebrae back fat thickness QTL located between 131.4 and 140.8 Mb has not been reported before. Furthermore, the genomic region defined by the three confidence intervals (127.6-140.8 Mb) includes the interval (134.6-135 Mb) described by Sanchez et al. (2014), associated with backfat traits. The genomic region defined in this study contains 64 annotated genes, where one of the most relevant genes appears to be the Leptin Receptor Overlapping Transcript (*LEPROT*; SSC6: 135.37-135.38), which negatively regulates leptin cell surface exposed receptors (Couturier et al., 2007). Leptin hormone has crucial roles in feed intake, growth and backfat traits. In swine, it has been shown that the serum concentrations of leptin were positively correlated with backfat thickness and negatively correlated with carcass muscle content (Berg et al., 2003). Moreover, several studies showed associations between polymorphisms in the leptin receptor (*LEPR)* gene and carcass measurements, including backfat thickness and loin weights (Ovilo et al., 2005; Muñoz et al., 2009; Muñoz et al., 2011;Uemoto et al., 2012). The *LEPR* gene which encodes multiple isoforms of the leptin receptor (Tartaglia, 1997), is located on an unassigned contig in the currently available pig genome assembly (version 10.2.83). Thus, LEPR is not included in the list of annotated genes in the QTL region (Table S.6), and it is not proposed as a candidate gene for this QTL. However, earlier studies

have shown that LEPR maps to this region of SSC6 (Ernst et al., 1996). Further analysis using improved genome assemblies and annotations will be needed to determine if the causative gene behind the reported QTL is LEPR, LEPROT, both of them, or another gene(s) in this region. Finally, the results of the SNP effects for M1GA0008917 (*f*(A)=1 in Duroc) and ALGA0122657 (*f*(B)=1 in Duroc) are in agreement with a previous study showing that Pietrain-sired pigs have less backfat thickness than Duroc-sired pigs (Edwards et al., 2003).

A QTL region for LMA and DRESS%, located on SSC7 (32.70-40 Mb) has already been observed this population using linkage analysis (Edwards et al., 2008; Choi et al., 2011), but results of this study refine the position of the QTL narrowing it from 27 cM and 14 cM, respectively, in the low resolution linkage maps to a specific 7.3 Mb genomic segment. The QTL for LMA has also been reported in Meishan x Pietrain and Meishan x Duroc populations using linkage analysis (Geldermann et al., 2003; Sato and Oyamada, 2003). The B allele of ASGA0032589 is fixed in Duroc grandsires and its negative substitution effect is in agreement with our previous report Edwards et al. (2003), which showed that Duroc-sired pigs had less loin muscle area than Pietrain-sired pigs. The QTL region (32.70-40 Mb) comprises 96 annotated genes, where there is no obvious biological link between those genes and the traits in this study.

The NR and CL are economically important traits. Carcass length is mainly determined by the number and length of thoracolumbar vertebras, while the number of ribs is defined by the number of thoracic vertebras. Having one more vertebrae adds on average 15 mm to the carcass length (King and Roberts, 1960), and if the extra vertebrae is thoracic it can add an extra rib, therefore the carcass will have more value than carcasses without the additional vertebra. The QTL region identified for NR and CL in this study (SSC7, 102.5-106.2 Mb) was previously reported in our linkage analysis (Edwards et al., 2008; Choi et al., 2011). According to the

positions reported in the PigQTL database (Hu et al., 2015) several QTL using low resolution linkage maps for carcass length (Liu et al., 2007; Uemoto & Nagamine, 2008; Yoo et al., 2014) and number of ribs (Zhang et al., 2007) in different populations including Western, Chinese and Korean breeds were reported in this region. Also, a genome-wide association study was performed by Sanchez et al. (2014) defining a region with SNP associated with carcass length between 101.1 and 105.3 Mb, which partially overlaps with the QTL region identified in this study. Finally, of the 57 genes annotated in the QTL segment, vertebrae development associated gene (*VRTN*: SCC7:103.45-103.46 Mb) is a strong candidate because the number of vertebras will affect carcass length and number of ribs at the same time. Fan et al. (2013) performed a GWA study with 3 populations (White Duroc x Erhualian F2, Sutai and Erhualian x Tongcheng F2) and reported this QTL region (SSC7, 103.37-104.31 Mb) associated only with the number of the thoracic vertebrae. Additionally, these authors found 2 SNP in complete LD residing in an active promoter corresponding to two transcription binding sites in VRTN and determined that those variants were associated with the number of thoracic vertebras (Fan et al., 2013). Also, its has been shown that an insertion/deletion in the gene contributed to the carcass length and thoracic number of vertebras in a Duroc purebred population (Nakano et al., 2015). Therefore, this gene is not only affecting the number of thoracic vertebras and the carcass length, but it is also a putative candidate gene for number of ribs.

Belly weight is another economically important trait, because from this primal cut, packers obtain the bacon. Thus having heavier bellies will add economic value to the carcass. We found a QTL on SCC11 (83.6-85.2 Mb) for belly weight. Our results agree with those from Milan et al. (2002), who performed a linkage analysis in a Large White x Meishan population, and reported a QTL peak in this region. We mapped this QTL region to 1.6 Mb genomic segment on SSC11.

The QTL region (SSC11, 83.6-85.2Mb) contains 10 genes, including *COL4A1* and *COL4A2* that synthetize the α1 and α2 chains of type IV collagen (Kühn, 1995). These two chains are the main component of the basement membranes (Van Der Rest and Garrone, 1991). Additionally, this region includes the *MYO16* which encodes Myosin XVI, involved in brain development (Patel et al., 2001). Therefore, none of the three genes in the region have been attributed functions related to muscle or fat development, or mechanisms with an obvious association with belly weight, thus the genetic cause of this association peak remains unknown.

On SSC15, QTL for seven traits were found. Because these traits are correlated, we report a single region consisting of the overlapped 95% confidence intervals (133.4-137.1 Mb). This QTL has been widely studied in different swine populations due to its relation to meat quality traits based on low resolution linkage analysis (Thomsen et al., 2004; Rohrer et al., 2005; Edwards et al., 2008; Li et al., 2010; Choi et al., 2011), GWA (Nonneman et al., 2013; Zhang et al., 2015) and in a recent meta-analysis (Bernal Rubio et al., 2015a). The QTL peak found in this study corresponding to juiciness, tenderness and WBS (135.2Mb) replicates our previous results using a low resolution map (Edwards et al., 2008; Choi et al., 2011). Additionally Thomsen et al. (2004) reported a QTL for TEN in a linkage analysis for a F2 Berkshire x Yorkshire population. A QTL peak for WBS was detected in a three way cross Duroc x Landrace x Large White, between 133-134 Mb (Zhang et al., 2015). The genome-wide study using the SNP chip allowed us to refine the QTL regions for our population, for tenderness to a 2.8 Mb region (133.8-136.6 Mb), and for juiciness to a 2.6 Mb region (133.4-137 Mb).

A QTL region associated with PRO, DL and CY has been previously reported in a linkage study for our population (Choi et al., 2011) and for CY (Edwards et al., 2008). Also, a QTL has been reported for DL with a low resolution linkage map (Li et al., 2010), with a GWA (Zhang et al.,

2015) and with a meta analysis using this population (Bernal Rubio et al., 2015a) and for CY with a low resolution map (Rohrer et al., 2005) and with an association study of SNP of PRKAG3 (Rohrer et al., 2012) and with a GWA (Nonneman et al., 2013). Although we did not refine the regions associated with WBS, PRO, DL and CY, our results add support for the region being a true QTL affecting these traits. Our report of a negative effect for DL is consistent with a pervious study showing that Duroc-sired pigs have less drip loss than Pietrain-sired pigs (Edwards et al., 2003).

The pH of the loin muscle at 24 hours post-mortem (pH24) has been widely studied, because this trait can drastically affect meat quality. The QTL for pH24 on SSC15 has been reported by several groups using low resolution linkage analysis (Edwards et al., 2008; Li et al., 2010; Choi et al., 2011), GWA (Nonneman et al., 2013; Zhang et al., 2015) and in a recent meta-analysis conducted by our group, that included the data used for this study (Bernal Rubio et al., 2015a). Some of these previous studies have proposed PRKAG3 (Choi et al., 2011; Nonneman et al., 2013; Bernal Rubio et al., 2015a; Zhang et al., 2015) as the likely candidate gene for this QTL. Moreover, specific variants within PRKAG3 have been proposed (Milan et al., 2000; Ciobanu et al., 2001; Ryan et al., 2012; Uimari and Sironen, 2014). Among the proposed candidate SNP, we genotyped I199V and T30N because according to Ciobanu et al. (2001), these two non-synonymous variants are more common to be segregating in Duroc in comparison with the G52S variant, while I199V has been associated with glycolytic potential traits in Pietrain (Ryan et al., 2012). In this study we showed that, genotypes of I199V and T30N do not fully explain the observed QTL variance. Furthermore, there are other variants in PRKAG3 not in LD with I199V and T30N that have been reported associated to pH 24 (Ryan et al., 2012; Uimari and Sironen, 2014). Therefore, other SNP in PRKAG3, or SNP in other genes in LD with MARC0093624 are

the causal variants behind observed phenotypic variation in pH 24 hours post-mortem and related traits in this population.

CONCLUSION

We performed a GWA and used statistical support intervals to map QTL and to define genomic segments with high likelihood of containing the causative genes. We found nine QTL peaks associated with 15 traits. Two QTL associated with tenderness on SSC3 and SSC5 are novel findings. One novel candidate gene, AKAP3, was proposed for the QTL on SSC5. AKAP3 is expressed in the skeletal muscle and it binds to the regulatory subunit of PKA, thus, affecting glycogen content in the skeletal muscle, which after post mortem modification in muscle could potentially lead to inferior meat quality. The gene CRAT on SSC1, which is related to lipid metabolism was proposed as a candidate gene for BF10. Finally, we showed that the known variants I199V and T30N in the PRKAG3 gene do not fully explain the QTL found on SSC15 for pH24 and related traits.

APPENDIX

Table S.**1** Number of observations, phenotypic mean, phenotypic SD and heritability of the traits analyzed in this study.

Trait	***n***	**Mean**	**SD**	**h^2**
Carcass Measures				
Off-farm body weight, kg	934	112.1	8.58	0.23
Hot carcass weight, kg	934	81.85	6.85	0.16
Dressing percent, %	934	73	2.12	0.24
45-min temperature, °C	933	39.42	2.16	0.07
24-h temperature, °C	931	2.9	1.19	0.15
45-min pH	920	6.37	0.22	0.09
24-h pH	913	5.51	0.14	0.18
45-min to 24-h pH decline	900	0.86	0.22	0.06
Carcass length, cm	933	78.73	2.53	0.48
Number of ribs	655	14.83	0.85	0.38
First-rib backfat, mm	845	40.62	7.06	0.23
Last-rib backfat, mm	933	28.66	6.44	0.25
Last-lumbar vertebrae backfat, mm	932	22.23	6.25	0.41
Tenth-rib backfat, mm	927	24.14	7.32	0.45
Loin Muscle Area, cm^2	928	40.61	4.73	0.59
Primal cut weight				
Belly weight, kg	933	5.03	0.68	0.19
Boston shoulder weight, kg	933	3.9	0.56	0.24
Ham weight, kg	933	9.63	0.77	0.5
Loin weight, kg	933	8.29	0.84	0.3
Picnic shoulder weight, kg	933	3.72	0.57	0.15
Spare rib weight , kg	930	1.53	0.2	0.38
Meat quality evaluation				
a*	887	17.27	1.83	0.63
b*	887	9.1	1.61	0.2
L*	887	53.77	2.25	0.36
Color, 1 to 6	931	3.16	0.82	0.26
Firmness, 1 to 5	918	2.86	0.79	0.13
Marbling, 1 to 10	932	2.82	0.85	0.4
Proximate Analysis				
Fat, %	922	3.18	1.4	0.54
Moisture, %	922	73.94	1.54	0.39
Protein, %	921	23.44	1.13	0.39

Table S.1 (cont'd)

Laboratory analyses				
Cook yield	924	77.27	2.84	0.3
Drip loss, %	932	1.83	1.17	0.27
Warner-Bratzler shear force (WBS), kg	923	3.21	0.69	0.26
Sensory panel analyses				
Connective tissue, 1 to 8	928	6.38	0.39	0.1
Juiciness, 1 to 8	928	5.23	0.59	0.07
Off -flavor, 1 to 8	928	1.14	0.21	0.05
Overall tenderness (OT), 1 to 8	928	5.63	0.55	0.28
Tenderness, 1 to 8	928	5.55	0.61	0.28

Table S.2 Primer and reporter sequences used to genotype PRKAG3 SNP T30N and I199V.

Sequence	**T30N**	**I199V**
Forward Primer	TGTAACCACCAGCTCAGAAAGAAG	ACACCATGCTGGAGATCAAGAA
Reverse Primer	CATCCTCCTGCCTTGTCCAT	TGCTTCTTGCTGTCCCACAAA
Reporter Sequence 1	TAGAGGCCTTGTTCCCCT	CCAACGGCATCCGAG
Reporter Sequence 2	AGGCCTTGGTCCCCT	CCAACGGCGTCCGAG

Table S.3 List of genes in the SSC1QTL region for BF10ordered according to increasing start position expressed in bp.

ENSEMBLE ID	Gene Name	Start position	End Position
ENSSSCG00000005651	ODF2	302,880,019	302,916,932
ENSSSCG00000005650	CERCAM	302,890,305	302,898,623
ENSSSCG00000005652	GLE1	302,917,898	302,944,879
ENSSSCG00000005654	SPTAN1	302,971,843	303,025,990
ENSSSCG00000005655	WDR34	303,025,978	303,054,229
ENSSSCG00000005656	SET	303,074,195	303,088,325
ENSSSCG00000005657	PKN3	303,093,222	303,107,386
ENSSSCG00000005658	ZDHHC12	303,105,455	303,107,386
ENSSSCG00000005659	ZER1	303,110,520	303,142,117
ENSSSCG00000005660	TBC1D13	303,151,697	303,165,931
ENSSSCG00000005661	ENDOG	303,170,034	303,176,319
ENSSSCG00000005662	C9orf114	303,173,471	303,183,167
ENSSSCG00000005663	CCBL1	303,188,689	303,225,560
ENSSSCG00000005664	LRRC8A	303,225,709	303,256,403
ENSSSCG00000005665	PHYHD1	303,258,825	303,270,695
ENSSSCG00000020404	U5	303,267,226	303,267,313
ENSSSCG00000005666	DOLK	303,273,660	303,275,276
ENSSSCG00000005667	NUP188	303,275,683	303,329,083
ENSSSCG00000005668	SH3GLB2	303,329,913	303,348,472
ENSSSCG00000005669	FAM73B	303,351,066	303,380,151
ENSSSCG00000005670	DOLPP1	303,391,083	303,400,174
ENSSSCG00000005671	CRAT	303,403,630	303,417,444
ENSSSCG00000005672	PPP2R4	303,417,952	303,447,712
ENSSSCG00000005673		303,470,306	303,471,521
ENSSSCG00000005678	NTMT1	303,687,311	303,696,507
ENSSSCG00000005679	ASB6	303,696,543	303,701,842
ENSSSCG00000005680	PRRX2	303,720,888	303,770,893
ENSSSCG00000005688	PTGES	303,779,651	303,792,393
ENSSSCG00000005674	C9orf50	303,804,123	303,806,765
ENSSSCG00000005683	TOR1B	303,864,106	303,869,260
ENSSSCG00000025964	TOR1A	303,871,932	303,882,873
ENSSSCG00000030426	C9orf78	303,883,684	303,892,770
ENSSSCG00000024341	USP20	303,892,886	303,935,551

Table S.3 (cont'd)

ENSSSCG00000005689	FNBP1	303,945,075	304,034,969
ENSSSCG00000005698	NCS1	304,186,077	304,192,568
ENSSSCG00000005699		304,346,787	304,445,413
ENSSSCG00000005701	ASS1	304,454,129	304,508,998
ENSSSCG00000005702	FUBP3	304,575,334	304,625,560
ENSSSCG00000005703	PRDM12	304,646,521	304,662,114
ENSSSCG00000005704	EXOSC2	304,671,176	304,679,921
ENSSSCG00000005706	ABL1	304,685,889	304,829,672
ENSSSCG00000005705	QRFP	304,834,824	304,835,228
ENSSSCG00000005707	FIBCD1	304,844,674	304,879,017
ENSSSCG00000005710	LAMC3	304,979,408	305,098,340
ENSSSCG00000005711	NUP214	305,131,025	305,241,772
ENSSSCG00000028172	FAM78A	305,269,556	305,284,180
ENSSSCG00000005713	PLPP7	305,296,657	305,297,166
ENSSSCG00000022130		305,352,962	305,353,855
ENSSSCG00000005715		305,377,245	305,439,325
ENSSSCG00000029718	SNORD62	305,426,341	305,426,426
ENSSSCG00000021682	SNORD62	305,430,346	305,430,431
ENSSSCG00000005716	POMT1	305,440,958	305,456,203
ENSSSCG00000005717	UCK1	305,455,972	305,462,254
ENSSSCG00000027241		305,471,374	305,486,208
ENSSSCG00000005719	RAPGEF1	305,502,184	305,633,144
ENSSSCG00000005720	MED27	305,733,695	305,937,230
ENSSSCG00000005721		306,021,149	306,021,451
ENSSSCG00000005723	NTNG2	306,096,986	306,137,454
ENSSSCG00000005724	SETX	306,154,772	306,228,908
ENSSSCG00000005725	TTF1	306,231,645	306,244,293
ENSSSCG00000026458		306,284,353	306,285,021
ENSSSCG00000005727	CFAP77	306,333,931	306,368,383
ENSSSCG00000005730	BARHL1	306,405,795	306,412,475
ENSSSCG00000005729	DDX31	306,414,455	306,486,475
ENSSSCG00000005731	GTF3C4	306,486,802	306,506,690
ENSSSCG00000005732	AK8	306,530,099	306,530,326
ENSSSCG00000023474		306,637,564	306,641,266
ENSSSCG00000027595		306,696,779	306,698,101
ENSSSCG00000005738	RALGDS	306,700,486	306,714,090
ENSSSCG00000028671	CEL	306,719,763	306,726,596
ENSSSCG00000005739	GTF3C5	306,729,965	306,927,497

Table S.3 (cont'd)

ENSSSCG00000022250		306,779,660	306,788,316
ENSSSCG00000005733		306,794,635	306,820,070
ENSSSCG00000005737	GFI1B	306,853,465	306,857,748
ENSSSCG00000005735		306,897,871	306,903,460
ENSSSCG00000005734		306,919,984	306,923,953
ENSSSCG00000023173	REXO4	306,985,377	306,995,916
ENSSSCG00000021241	ADAMTS13	306,998,075	307,029,937
ENSSSCG00000021113	CACFD1	307,030,425	307,039,181
ENSSSCG00000024166	SLC2A6	307,041,086	307,049,206

Table S.4 List of genes in the SSC2 QTL region for WBS and TEN ordered according to increasing start position expressed in bp.

ENSEMBLE ID	Gene Name	Start position	End Position
ENSSSCG00000022736		1,051,307	1,071,874
ENSSSCG00000012864		1,228,857	1,235,210
ENSSSCG00000030888	FADD	1,388,619	1,390,090
ENSSSCG00000022579		1,469,314	1,477,594
ENSSSCG00000022123		1,487,597	1,488,346
ENSSSCG00000021469		1,508,154	1,518,080
ENSSSCG00000029328		1,563,459	1,567,755
ENSSSCG00000024158		1,575,575	1,605,841
ENSSSCG00000029868	ANO1	1,621,019	1,651,965
ENSSSCG00000012869		1,866,961	1,902,581
ENSSSCG00000012872	FGF3	1,965,237	1,973,613
ENSSSCG00000012870	FGF4	2,004,484	2,006,245
ENSSSCG00000012871	FGF19	2,060,948	2,064,674
ENSSSCG00000012874	ORAOV1	2,095,043	2,103,105
ENSSSCG00000012873	CCND1	2,342,270	2,343,202
ENSSSCG00000012875	TPCN2	2,708,675	2,727,191
ENSSSCG00000012878	IGHMBP2	2,751,383	2,774,853
ENSSSCG00000012879	MRPL21	2,773,574	2,786,091
ENSSSCG00000012880	CPT1A	2,815,220	2,874,163
ENSSSCG00000012881		2,828,172	2,838,805
ENSSSCG00000012882	MTL5	2,878,052	2,920,326
ENSSSCG00000012883	GAL	2,931,347	2,937,425
ENSSSCG00000012884	PPP6R3	2,974,853	3,097,590
ENSSSCG00000012885	LRP5	3,168,524	3,213,765
ENSSSCG00000012886		3,247,137	3,257,948
ENSSSCG00000012887		3,294,895	3,332,625
ENSSSCG00000012889	CHKA	3,416,545	3,473,335
ENSSSCG00000012890	TCIRG1	3,476,378	3,488,420
ENSSSCG00000023420		3,498,054	3,507,821
ENSSSCG00000028501		3,518,994	3,536,523
ENSSSCG00000012893		3,539,587	3,551,073
ENSSSCG00000026349		3,563,353	3,568,662
ENSSSCG00000012895		3,594,982	3,598,550

Table S.4 (cont'd)

ENSSSCG00000012896	NDUFV1	3,598,806	3,604,500
ENSSSCG00000012897	GSTP1	3,618,668	3,620,934
ENSSSCG00000012900	CABP2	3,621,643	3,624,689
ENSSSCG00000012901		3,634,981	3,637,655
ENSSSCG00000028107		3,724,619	3,727,474
ENSSSCG00000020764		3,734,833	3,739,060
ENSSSCG00000012902	CDK2AP2	3,744,053	3,746,289
ENSSSCG00000012903	PITPNM1	3,748,579	3,760,628
ENSSSCG00000012904	AIP	3,761,548	3,767,520
ENSSSCG00000012905	TMEM134	3,775,765	3,782,603
ENSSSCG00000012906	CABP4	3,787,774	3,790,770
ENSSSCG00000012920	CORO1B	3,799,100	3,804,062
ENSSSCG00000012909	PTPRCAP	3,806,257	3,806,883
ENSSSCG00000012910	RPS6KB2	3,807,219	3,813,390
ENSSSCG00000012911	CARNS1	3,814,979	3,821,634
ENSSSCG00000012912	TBC1D10C	3,830,957	3,837,376
ENSSSCG00000012914	RAD9A	3,838,643	3,848,285
ENSSSCG00000012913	PPP1CA	3,838,643	3,843,037
ENSSSCG00000012915	CLCF1	3,885,576	3,888,922
ENSSSCG00000012916		3,944,557	3,947,680
ENSSSCG00000027956	SSH3	3,968,534	3,976,217
ENSSSCG00000012917	ANKRD13D	3,976,503	3,990,125
ENSSSCG00000012918	ADRBK1	3,992,424	4,000,951
ENSSSCG00000029637		4,021,430	4,051,218
ENSSSCG00000021760		4,156,989	4,182,781
ENSSSCG00000022726	RHOD	4,233,085	4,238,597
ENSSSCG00000025443		4,350,835	4,351,810
ENSSSCG00000024837	SYT12	4,385,769	4,409,341
ENSSSCG00000028474		4,514,069	4,515,044
ENSSSCG00000012925		4,576,317	4,585,943
ENSSSCG00000012926		4,604,883	4,609,731
ENSSSCG00000012927	RCE1	4,612,347	4,615,044
ENSSSCG00000029902	C11orf80	4,615,256	4,674,876
ENSSSCG00000029317		4,752,997	4,792,821
ENSSSCG00000012946		4,793,528	4,810,745
ENSSSCG00000021891		4,816,099	4,825,400
ENSSSCG00000025136		4,826,321	4,839,557
ENSSSCG00000021340		4,839,809	4,844,997

Table S.4 (cont'd)

ENSSSCG00000012933	CCDC87	4,864,531	4,867,104
ENSSSCG00000012934		4,867,366	4,875,351
ENSSSCG00000027368		4,930,949	4,936,147
ENSSSCG00000030087		4,936,399	4,949,636
ENSSSCG00000012947		4,950,491	4,957,808
ENSSSCG00000025339		4,965,113	4,982,609
ENSSSCG00000030415		4,983,355	5,022,966
ENSSSCG00000012944	PELI3	5,026,250	5,036,536
ENSSSCG00000012943	MRPL11	5,054,876	5,057,918
ENSSSCG00000012942	NPAS4	5,067,608	5,072,359
ENSSSCG00000012941	SLC29A2	5,104,241	5,111,869
ENSSSCG00000012940	B4GAT1	5,120,057	5,122,388
ENSSSCG00000012939	BRMS1	5,122,632	5,130,266
ENSSSCG00000012950	RIN1	5,131,242	5,135,843
ENSSSCG00000029949	CD248	5,146,889	5,149,543
ENSSSCG00000012952	TMEM151A	5,166,663	5,171,830
ENSSSCG00000012953	YIF1A	5,173,983	5,178,254
ENSSSCG00000012954	CNIH2	5,178,535	5,184,499
ENSSSCG00000012955	KLC2	5,189,928	5,198,860
ENSSSCG00000012956	PACS1	5,210,217	5,246,316
ENSSSCG00000012957	SF3B2	5,351,240	5,366,039
ENSSSCG00000028015	GAL3ST3	5,373,567	5,378,222
ENSSSCG00000012959	CATSPER1	5,384,927	5,393,145
ENSSSCG00000012960	CST6	5,395,320	5,396,971
ENSSSCG00000012961		5,405,034	5,405,903
ENSSSCG00000027334		5,408,048	5,411,734
ENSSSCG00000029016		5,420,795	5,425,387
ENSSSCG00000027377		5,437,528	5,438,439
ENSSSCG00000012962		5,440,257	5,441,959
ENSSSCG00000012963	SART1	5,458,386	5,477,036
ENSSSCG00000012964	TSGA10IP	5,478,348	5,491,342
ENSSSCG00000012965	DRAP1	5,513,474	5,515,508
ENSSSCG00000012966	C11orf68	5,515,886	5,518,146
ENSSSCG00000012967	FOSL1	5,534,025	5,541,866
ENSSSCG00000012968	CCDC85B	5,541,055	5,541,663
ENSSSCG00000012969	FIBP	5,544,493	5,553,916
ENSSSCG00000012970	CTSW	5,550,132	5,554,197
ENSSSCG00000012971	EFEMP2	5,558,977	5,570,871

Table S.4 (cont'd)

ENSSSCG00000012973	MUS81	5,565,606	5,572,868
ENSSSCG00000012974	CFL1	5,572,133	5,579,066
ENSSSCG00000012975	SNX32	5,580,145	5,584,189
ENSSSCG00000027361	OVOL1	5,630,671	5,642,984
ENSSSCG00000012977		5,646,350	5,646,901
ENSSSCG00000012978	AP5B1	5,652,042	5,655,474
ENSSSCG00000012980	KAT5	5,666,205	5,676,732
ENSSSCG00000012979	RNASEH2C	5,667,175	5,668,321
ENSSSCG00000012981	RELA	5,698,668	5,707,943
ENSSSCG00000012982		5,760,306	5,768,449
ENSSSCG00000012983	PCNXL3	5,771,503	5,791,589
ENSSSCG00000012984	SCYL1	5,830,476	5,846,197
ENSSSCG00000012985	LTBP3	5,840,778	5,861,803
ENSSSCG00000012988	EHBP1L1	5,865,073	5,880,838
ENSSSCG00000012986		5,881,656	5,884,689
ENSSSCG00000020737		5,904,731	5,907,469
ENSSSCG00000012992	FRMD8	5,955,700	5,976,369
ENSSSCG00000012993	SLC25A45	5,984,015	5,990,284
ENSSSCG00000012994	TIGD3	5,999,046	6,000,491
ENSSSCG00000012995	DPF2	6,003,983	6,020,495
ENSSSCG00000012996	CDC42EP2	6,030,298	6,036,364
ENSSSCG00000012997	POLA2	6,054,435	6,081,792
ENSSSCG00000012998		6,102,436	6,127,700
ENSSSCG00000012999	CAPN1	6,129,549	6,155,373
ENSSSCG00000030977	CU457406.2	6,143,457	6,143,931
ENSSSCG00000019315	U6	6,182,339	6,182,442
ENSSSCG00000027057	SYVN1	6,192,447	6,200,950
ENSSSCG00000013001	MRPL49	6,198,564	6,203,040
ENSSSCG00000013002		6,203,200	6,204,847
ENSSSCG00000013003	ZNHIT2	6,207,436	6,208,686
ENSSSCG00000013004	TM7SF2	6,208,815	6,213,504
ENSSSCG00000013005	VPS51	6,213,604	6,226,834
ENSSSCG00000013006	TMEM262	6,230,658	6,231,296
ENSSSCG00000013007	ZFPL1	6,231,271	6,235,566
ENSSSCG00000013008	CDCA5	6,235,700	6,251,633
ENSSSCG00000013010	NAALADL1	6,259,685	6,269,977
ENSSSCG00000013011	SAC3D1	6,270,450	6,273,674
ENSSSCG00000013012	SNX15	6,275,115	6,287,159

Table S.4 (cont'd)

ENSSSCG00000013013	ARL2	6,290,009	6,299,042
ENSSSCG00000023178	BATF2	6,312,310	6,321,081
ENSSSCG00000013015	GPHA2	6,378,546	6,379,201
ENSSSCG00000013016	PPP2R5B	6,379,401	6,386,952
ENSSSCG00000023890	ATG2A	6,393,343	6,413,661
ENSSSCG00000019906	ssc-mir-194a	6,416,776	6,416,859
ENSSSCG00000018349	ssc-mir-192	6,416,980	6,417,059
ENSSSCG00000013017	EHD1	6,430,537	6,452,599
ENSSSCG00000013018	CDC42BPG	6,460,653	6,480,139
ENSSSCG00000013019	MEN1	6,490,721	6,497,546
ENSSSCG00000024968		6,497,782	6,497,852
ENSSSCG00000013020	MAP4K2	6,497,805	6,512,966
ENSSSCG00000013021	SF1	6,524,030	6,536,326
ENSSSCG00000013022	PYGM	6,542,313	6,553,842
ENSSSCG00000013023	RASGRP2	6,557,414	6,573,068
ENSSSCG00000013024	NRXN2	6,587,432	6,684,684
ENSSSCG00000013025	SLC22A12	6,688,308	6,696,328
ENSSSCG00000020831	SLC22A11	6,777,940	6,791,460
ENSSSCG00000013030	PRDX5	6,864,027	6,867,851
ENSSSCG00000013028	ESRRA	6,867,951	6,878,759
ENSSSCG00000013029	TRMT112	6,868,094	6,869,375
ENSSSCG00000013027	TEX40	6,881,439	6,883,031
ENSSSCG00000013031	KCNK4	6,883,965	6,890,656
ENSSSCG00000013032	GPR137	6,893,208	6,896,116
ENSSSCG00000013033	BAD	6,897,945	6,909,698
ENSSSCG00000013034	PLCB3	6,911,684	6,927,124
ENSSSCG00000013035	PPP1R14B	6,932,322	6,934,762
ENSSSCG00000013036	FKBP2	6,935,189	6,936,674
ENSSSCG00000013037	VEGFB	6,939,770	6,942,952
ENSSSCG00000013038	DNAJC4	6,943,440	6,947,235
ENSSSCG00000013039	NUDT22	6,947,812	6,950,624
ENSSSCG00000013040	TRPT1	6,950,712	6,953,739
ENSSSCG00000013041	FERMT3	6,953,588	6,976,696
ENSSSCG00000021620		6,978,383	7,040,372
ENSSSCG00000013042	CCDC88B	6,992,733	7,010,294
ENSSSCG00000028806	RPS6KA4	7,012,544	7,025,897
ENSSSCG00000013043	MACROD1	7,054,730	7,204,542
ENSSSCG00000013044	FLRT1	7,095,830	7,097,924

Table S.4 (cont'd)

ENSSSCG00000013045	OTUB1	7,204,713	7,212,472
ENSSSCG00000013046		7,218,448	7,220,218
ENSSSCG00000022687	NAA40	7,282,375	7,290,873
ENSSSCG00000013049	RCOR2	7,350,357	7,352,786
ENSSSCG00000013050	MARK2	7,355,161	7,364,707
ENSSSCG00000013048	C11orf84	7,464,512	7,506,772
ENSSSCG00000013051		7,516,650	7,517,021
ENSSSCG00000013053	C11orf95	7,550,065	7,555,127
ENSSSCG00000013052	RTN3	7,559,962	7,714,023
ENSSSCG00000013054		7,615,311	7,649,588
ENSSSCG00000026914	PLA2G16	7,653,511	7,681,456

Table S.5 List of genes in the SSC5 QTL region for TEN ordered according to increasing start position expressed in bp.

ENSEMBLE ID	Gene Name	Start position	End Position
ENSSSCG00000021596	KCNA5	67,653,414	67,655,144
ENSSSCG00000000716	KCNA1	67,768,663	67,770,150
ENSSSCG00000000717		67,863,129	67,864,757
ENSSSCG00000000718	GALNT8	67,903,769	67,941,362
ENSSSCG00000000719		67,965,977	67,994,531
ENSSSCG00000000720	AKAP3	68,003,279	68,020,465
ENSSSCG00000000721	DYRK4	68,022,319	68,047,436
ENSSSCG00000000722	RAD51AP1	68,079,112	68,103,087
ENSSSCG00000000723	C12orf4	68,104,334	68,145,457
ENSSSCG00000000724	FGF6	68,198,018	68,207,512
ENSSSCG00000029028	FGF23	68,250,561	68,261,760
ENSSSCG00000024219	TIGAR	68,266,578	68,280,421
ENSSSCG00000000727	CCND2	68,314,898	68,331,943
ENSSSCG00000000728	PARP11	68,750,885	68,791,407
ENSSSCG00000000730	PRMT8	68,942,252	69,061,177
ENSSSCG00000000732	CRACR2A	69,023,638	69,052,093

Table S.6 List of genes in the SSC6 QTL region for BF10, LLBF and LW ordered according to increasing start position expressed in bp.

ENSEMBLE ID	Gene Name	Start position	End Position
ENSSSCG00000003777	SLC44A5	127,567,588	127,970,326
ENSSSCG00000020641	U6	127,659,262	127,659,367
ENSSSCG00000003778	LHX8	128,034,488	128,059,465
ENSSSCG00000003779	TYW3	128,286,279	128,301,919
ENSSSCG00000003780	CRYZ	128,304,889	128,327,585
ENSSSCG00000003781	ERICH3	128,400,155	128,476,446
ENSSSCG00000003782		128,673,923	128,673,994
ENSSSCG00000003783	FPGT	128,715,728	128,722,791
ENSSSCG00000003784	LRRIQ3	128,721,115	128,766,182
ENSSSCG00000020493	5S_rRNA	129,093,391	129,093,527
ENSSSCG00000025085	NEGR1	130,680,416	130,880,673
ENSSSCG00000003787	ZRANB2	131,542,884	131,563,408
ENSSSCG00000019065	ssc-mir-186	131,558,966	131,559,047
ENSSSCG00000003788	PTGER3	131,574,481	131,616,324
ENSSSCG00000003789	CTH	132,027,412	132,050,286
ENSSSCG00000003790	ANKRD13C	132,080,136	132,189,620
ENSSSCG00000003791	SRSF11	132,193,667	132,240,405
ENSSSCG00000003792	LRRC40	132,240,462	132,288,460
ENSSSCG00000003793	LRRC7	132,309,803	132,436,579
ENSSSCG00000023754		132,544,632	132,612,363
ENSSSCG00000003794	RPE65	133,492,012	133,513,894
ENSSSCG00000003795	WLS	133,752,325	133,820,951
ENSSSCG00000003797	DIRAS3	133,865,007	133,867,486
ENSSSCG00000003798	SERBP1	134,068,990	134,081,998
ENSSSCG00000003799	IL12RB2	134,096,226	134,161,817
ENSSSCG00000003800		134,197,114	134,197,933
ENSSSCG00000003801	IL23R	134,219,683	134,283,657
ENSSSCG00000003802	SLC35D1	134,374,824	134,382,782
ENSSSCG00000003803	C1orf141	134,410,380	134,464,826
ENSSSCG00000030423	MIER1	134,486,591	134,497,982
ENSSSCG00000023741		134,537,789	134,545,301
ENSSSCG00000003805	PDE4B	134,870,923	134,913,618

Table S.6 (cont'd)

	LEPROT	135,379,661	135,387,507
ENSSSCG00000003806			
ENSSSCG00000003807	DNAJC6	135,397,188	135,560,153
ENSSSCG00000003808		135,587,479	135,663,171
ENSSSCG00000018551	5S_rRNA	135,686,780	135,686,910
ENSSSCG00000019990	ssc-mir-101-2	135,736,114	135,736,196
ENSSSCG00000019115		135,812,759	135,812,840
ENSSSCG00000003809	JAK1	135,899,356	135,917,892
ENSSSCG00000025672	RAVER2	135,984,034	136,062,365
ENSSSCG00000004829	CACHD1	136,116,938	136,181,931
ENSSSCG00000003810	UBE2U	136,685,425	136,756,970
ENSSSCG00000003811	ROR1	136,765,918	136,909,446
ENSSSCG00000028540	U6	136,874,221	136,874,327
ENSSSCG00000003812	PGM1	137,171,271	137,233,581
ENSSSCG00000003814	EFCAB7	137,259,293	137,316,747
ENSSSCG00000020427	SNORA16	137,320,904	137,321,016
ENSSSCG00000003815	ALG6	137,377,946	137,434,247
ENSSSCG00000003816	ATG4C	137,803,101	137,886,719
ENSSSCG00000030607	U6	138,028,937	138,029,043
ENSSSCG00000003818	DOCK7	138,042,658	138,172,957
ENSSSCG00000003819	ANGPTL3	138,101,691	138,111,290
ENSSSCG00000003821	KANK4	138,309,636	138,383,020
ENSSSCG00000023243		138,764,544	138,851,088
ENSSSCG00000003823	C1orf87	139,665,483	139,746,337
ENSSSCG00000003824		139,867,650	139,892,000
ENSSSCG00000021940	CYP2J34	139,902,194	139,929,805
ENSSSCG00000023035	U6	139,902,298	139,902,404
ENSSSCG00000003825	CYP2J2	139,936,092	139,965,989
ENSSSCG00000019515	U6	139,946,008	139,946,112
ENSSSCG00000003826	HOOK1	140,117,181	140,190,970
ENSSSCG00000022526		140,248,872	140,253,789
ENSSSCG00000003827		140,363,964	140,393,806
ENSSSCG00000003828		140,421,878	140,676,185

Table S.7 List of genes in the SSC7 QTL region associated with loin muscle area and dressing percentage, ordered according to increasing start position expressed in bp.

ENSEMBLE ID	Gene Name	Start position	End Position
ENSSSCG00000001493		32750198	33022968
ENSSSCG00000019940	U6	32784377	32784487
ENSSSCG00000029539	RAB23	33091878	33119792
ENSSSCG00000001494	BAG2	33129268	33140672
ENSSSCG00000026207		33140918	33141047
ENSSSCG00000001497	ZNF451	33144581	33226714
ENSSSCG00000001498	BEND6	33231047	33273159
ENSSSCG00000025172		33274353	33374712
ENSSSCG00000001499	DST	33504832	33751812
ENSSSCG00000020421	SNORA72	33632171	33632302
ENSSSCG00000001500	COL21A1	33918801	34009970
ENSSSCG00000001501	VPS52	34128551	34142395
ENSSSCG00000001502	RPS18	34142621	34146677
ENSSSCG00000001503	B3GALT4	34146955	34148615
ENSSSCG00000001504	WDR46	34148893	34157078
ENSSSCG00000001505	PFDN6	34157443	34158709
ENSSSCG00000001506	RGL2	34159507	34166143
ENSSSCG00000001507	TAPBP	34167701	34176411
ENSSSCG00000001508	ZBTB22	34177375	34180363
ENSSSCG00000001509	DAXX	34180626	34184970
ENSSSCG00000001510		34198532	34213249
ENSSSCG00000001517		34273783	34278594
ENSSSCG00000001516	BAK1	34279881	34288116
ENSSSCG00000001515	ZBTB9	34338560	34339966
ENSSSCG00000001513	SYNGAP1	34342914	34371345
ENSSSCG00000001512	CUTA	34373453	34375042
ENSSSCG00000001511	PHF1	34375137	34380476
ENSSSCG00000028301		34381247	34387660
ENSSSCG00000001518	ITPR3	34443056	34510838
ENSSSCG00000030668	UQCC2	34506319	34537625
ENSSSCG00000001520	LEMD2	34641299	34656755
ENSSSCG00000001521	MLN	34663313	34672854
ENSSSCG00000001523	GRM4	34839241	34928071
ENSSSCG00000001526	HMGA1	34984737	34990089

Table S.7 (cont'd)

		34990224	34992729
ENSSSCG00000023160			
ENSSSCG00000001528	RPS10	35109882	35117436
ENSSSCG00000024531	SPDEF	35214365	35233219
ENSSSCG00000027053	PACSIN1	35236544	35244143
ENSSSCG00000001527	C6orf106	35362835	35410240
ENSSSCG00000001531	SNRPC	35451546	35466721
ENSSSCG00000001532	UHRF1BP1	35474279	35538654
ENSSSCG00000001533	TAF11	35543524	35555308
ENSSSCG00000001534	ANKS1A	35556631	35766117
ENSSSCG00000001535	TCP11	35783270	35807767
ENSSSCG00000027696	U6	35795518	35795624
ENSSSCG00000001536	SCUBE3	35854337	36008752
ENSSSCG00000001544	TEAD3	36085701	36100686
ENSSSCG00000001543		36106602	36108831
ENSSSCG00000001546	FANCE	36111076	36123998
ENSSSCG00000025377		36127213	36135077
ENSSSCG00000001539	PPARD	36141606	36215260
ENSSSCG00000001538	DEF6	36220769	36243212
ENSSSCG00000001537	ZNF76	36245632	36254505
ENSSSCG00000001549		36373069	36478480
ENSSSCG00000001550	ARMC12	36518830	36529646
ENSSSCG00000001551	CLPSL2	36533600	36535664
ENSSSCG00000001552	CLPS	36546797	36549010
ENSSSCG00000001553	LHFPL5	36552501	36560440
ENSSSCG00000001554	SRPK1	36575699	36638641
ENSSSCG00000022412		36597506	36600159
ENSSSCG00000001555	SLC26A8	36660474	36714021
ENSSSCG00000001556	MAPK14	36725707	36795310
ENSSSCG00000020803		36804604	36865913
ENSSSCG00000001559	PNPLA1	36888272	36925502
ENSSSCG00000001560	C6orf222	36937623	36949206
ENSSSCG00000001561	ETV7	36990013	37008146
ENSSSCG00000001562	KCTD20	37046645	37093046
ENSSSCG00000001563	STK38	37106646	37154544
ENSSSCG00000001564		37191603	37198743
ENSSSCG00000022111		37258053	37267163
ENSSSCG00000001566	RAB44	37291881	37305207
ENSSSCG00000001567	CPNE5	37314126	37411217

Table S.7 (cont'd)

	PPIL1	37424434	37448185
ENSSSCG00000030390			
ENSSSCG00000001569	C6orf89	37451398	37483062
ENSSSCG00000001570	PI16	37504337	37514234
ENSSSCG00000001571	MTCH1	37520856	37539035
ENSSSCG00000001572	FGD2	37551887	37578005
ENSSSCG00000001573	PIM1	37691963	37697467
ENSSSCG00000001574		37721577	37722092
ENSSSCG00000001575		37724093	37724653
ENSSSCG00000001576	TMEM217	37728884	37729462
ENSSSCG00000001577		37742829	37818510
ENSSSCG00000001578	RNF8	37836027	37867217
ENSSSCG00000028436	SNORA70	37891821	37891951
ENSSSCG00000001579		37905566	37968899
ENSSSCG00000027292		37970118	37990894
ENSSSCG00000001582	MDGA1	38114185	38140436
ENSSSCG00000001583		38173326	38173520
ENSSSCG00000001584		38746485	38878917
ENSSSCG00000021215	ZFAND3	38909262	38909459
ENSSSCG00000030073		38947185	39089845
ENSSSCG00000027778	GLO1	39241001	39271606
ENSSSCG00000001588	DNAH8	39286129	39569290
ENSSSCG00000001589	GLP1R	39595053	39623436
ENSSSCG00000025423	KCNK5	39704035	39743285
ENSSSCG00000001592	KCNK17	39796871	39811137

Table S.8 List of genes in the SSC7 QTL region associated with number of ribs and carcass length, ordered according to increasing start position expressed in bp.

ENSEMBLE ID	Gene Name	Start position	End Position
ENSSSCG00000022178		102416956	102504747
ENSSSCG00000002344		102610127	102612130
ENSSSCG00000002345		102665568	102680830
ENSSSCG00000002346	PNMA1	102682182	102683240
ENSSSCG00000002347	DNAL1	102691235	102733635
ENSSSCG00000002348	ACOT6	102742330	102748593
ENSSSCG00000030559		102770509	102771156
ENSSSCG00000002349	ACOT4	102786805	102791429
ENSSSCG00000002350		102828131	102831810
ENSSSCG00000020600		102900731	102900857
ENSSSCG00000002351		102928581	102950219
ENSSSCG00000002352	ZNF410	102952601	103001300
ENSSSCG00000002353	FAM161B	102997112	103012716
ENSSSCG00000002354	COQ6	103012820	103025093
ENSSSCG00000002355	ENTPD5	103028636	103046142
ENSSSCG00000002356	BBOF1	103070839	103114503
ENSSSCG00000002357	ALDH6A1	103117991	103136318
ENSSSCG00000027564	LIN52	103136846	103337802
ENSSSCG00000024942		103192746	103202335
ENSSSCG00000002359		103231353	103248739
ENSSSCG00000002363	VSX2	103375383	103394310
ENSSSCG00000002362		103411781	103428283
ENSSSCG00000002361	VRTN	103457506	103467076
ENSSSCG00000025000	SYNDIG1L	103502244	103522505
ENSSSCG00000002366	NPC2	103573074	103582623
ENSSSCG00000002367	ISCA2	103582741	103584289
ENSSSCG00000002368	LTBP2	103590368	103694062
ENSSSCG00000002370	AREL1	103736066	103783771
ENSSSCG00000021442	FCF1	103783793	103805586
ENSSSCG00000002372		103818763	103886143
ENSSSCG00000002373	PROX2	103908499	103918733
ENSSSCG00000002373	PROX2	103908499	103918733
ENSSSCG00000002374	DLST	103934687	103955706
ENSSSCG00000002375	RPS6KL1	103957391	103974114

Table S.8 (cont'd)

	PGF	103994328	104006524
ENSSSCG00000002376			
ENSSSCG00000002376	PGF	103994328	104006524
ENSSSCG00000002377	EIF2B2	104047923	104054504
ENSSSCG00000002378	MLH3	104064666	104089446
ENSSSCG00000002379	ACYP1	104093822	104109273
ENSSSCG00000002380	ZC2HC1C	104110332	104117479
ENSSSCG00000002381	NEK9	104120031	104166205
ENSSSCG00000002382	TMED10	104173906	104206702
ENSSSCG00000002383	FOS	104293657	104297121
ENSSSCG00000021606		104424137	104479243
ENSSSCG00000030582	BATF	104503137	104524774
ENSSSCG00000024255		104608351	104625059
ENSSSCG00000002384		104685217	104811317
ENSSSCG00000029832	7SK	104854284	104854566
ENSSSCG00000025984	TTLL5	104990613	105175649
ENSSSCG00000025984	TTLL5	104990613	105175649
ENSSSCG00000029713	7SK	105034004	105034286
ENSSSCG00000002385	TGFB3	105178633	105206982
ENSSSCG00000002385	TGFB3	105178633	105206982
ENSSSCG00000002386	IFT43	105268438	105299304
ENSSSCG00000002387	GPATCH2L	105370887	105416373
ENSSSCG00000002388	ESRRB	105628140	105743716
ENSSSCG00000018299		105886390	105886534
ENSSSCG00000002389	VASH1	105960537	105978240
ENSSSCG00000002390	ANGEL1	105992708	105999604
ENSSSCG00000002391	LRRC74A	106005464	106033717
ENSSSCG00000002392	IRF2BPL	106151786	106154137

Table S.9 List of genes in the SSC15 QTL region for juiciness, tenderness, WBS, 24-h pH, protein content, CY and drip loss ordered according to increasing start position expressed in bp.

ENSEMBLE ID	Gene Name	Start position	End Position
ENSSSCG00000020722	SNORA42	133,359,013	133,359,150
ENSSSCG00000029002		133,367,666	133,370,494
ENSSSCG00000016185	ARPC2	133,377,438	133,389,903
ENSSSCG00000021228	SNORA42	133,415,454	133,415,591
ENSSSCG00000016186		133,425,398	133,432,043
ENSSSCG00000022483		133,442,237	133,443,117
ENSSSCG00000016189		133,444,024	133,447,622
ENSSSCG00000025058		133,452,329	133,456,736
ENSSSCG00000016187		133,462,315	133,468,029
ENSSSCG00000016196	VIL1	133,479,159	133,507,423
ENSSSCG00000016194	USP37	133,512,791	133,595,116
ENSSSCG00000016195		133,583,661	133,584,166
ENSSSCG00000016193	RQCD1	133,604,447	133,636,861
ENSSSCG00000016192		133,641,605	133,669,136
ENSSSCG00000016191		133,667,702	133,729,988
ENSSSCG00000026964		133,753,315	133,852,318
ENSSSCG00000016201	TTLL11	133,768,369	133,815,768
ENSSSCG00000016200	PRKAG3	133,800,248	133,807,019
ENSSSCG00000028455	TTLL4	133,817,190	133,818,795
ENSSSCG00000016199		133,828,770	133,831,213
ENSSSCG00000016202		133,834,386	133,837,207
ENSSSCG00000023145		133,855,102	133,856,393
ENSSSCG00000021584	CDK5R2	133,925,849	133,927,094
ENSSSCG00000026963		133,957,554	133,957,623
ENSSSCG00000026958		134,033,487	134,033,948
ENSSSCG00000016203	CCDC108	134,087,004	134,103,779
ENSSSCG00000016204	IHH	134,122,695	134,129,391
ENSSSCG00000016205	NHEJ1	134,145,208	134,232,523
ENSSSCG00000016213	GLB1L	134,233,624	134,318,259
ENSSSCG00000029694	SLC23A3	134,234,854	134,240,678
ENSSSCG00000016206	CNPPD1	134,243,394	134,248,180
ENSSSCG00000016207	FAM134A	134,249,648	134,255,270
ENSSSCG00000016208	ZFAND2B	134,277,911	134,280,873
ENSSSCG00000016210	ABCB6	134,280,965	134,288,809

Table S.9 (cont'd)

	ATG9A	134,289,150	134,300,597
ENSSSCG00000016211			
ENSSSCG00000018394		134,294,632	134,294,714
ENSSSCG00000016212	ANKZF1	134,300,649	134,309,215
ENSSSCG00000016214	STK16	134,320,058	134,323,470
ENSSSCG00000016215	TUBA4A	134,325,130	134,328,954
ENSSSCG00000016216		134,341,043	134,347,416
ENSSSCG00000016217	DNAJB2	134,355,300	134,360,929
ENSSSCG00000016218	PTPRN	134,364,671	134,384,701
ENSSSCG00000018628		134,369,134	134,369,275
ENSSSCG00000016219	RESP18	134,398,121	134,404,728
ENSSSCG00000022460		134,407,248	134,407,824
ENSSSCG00000016220	DNPEP	134,425,137	134,437,835
ENSSSCG00000030434	ssc-mir-4334	134,433,005	134,433,073
ENSSSCG00000020771	INHA	134,507,304	134,511,349
ENSSSCG00000028052	OBSL1	134,512,277	134,533,002
ENSSSCG00000027541	TMEM198	134,534,143	134,587,352
ENSSSCG00000020785	DES	134,560,460	134,567,338
ENSSSCG00000021610	CHPF	134,588,465	134,594,046
ENSSSCG00000029968	ASIC4	134,594,938	134,619,791
ENSSSCG00000023935	GMPPA	134,627,636	134,635,226
ENSSSCG00000018784	U6	135,384,999	135,385,103
ENSSSCG00000020038	SNORA31	135,665,635	135,665,753
ENSSSCG00000026566		135,772,108	135,772,198
ENSSSCG00000016230	EPHA4	136,746,506	136,874,945
ENSSSCG00000016231		136,930,815	136,936,376

LITERATURE CITED

LITERATURE CITED

Akanno, E. C., Plastow, G., B. W. Woodwards, S. Bauck, H. Okut, X.-L. Wu, C. Sun, J. L. Aalhus, S. S. Moore, S. P. Miller, W. Z., 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. http://doi.org/10.2527/jas.2013-7374

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. http://doi.org/10.1186/1471-2156-14-8

Becker, D., K. Wimmers, H. Luther, A. Hofer, and T. Leeb. 2013. A Genome-Wide Association Study to Detect QTL for Commercially Important Traits in Swiss Large White Boars. PLoS One 8:1–6. http://doi.org/10.1371/journal.pone.0055951

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate : a Practical and Powerful Approach to Multiple Testing When researchers tend to select pursuing multiple the (statistically) and support of conclusions . An unguarded use in a greatly results of single-inference inc. J.R Stat. Soc.B 57:289–300. http://doi.org/10.2307/234610

Berg, E. P., E. L. McFadin, K. R. Maddock, R. N. Goodwin, T. J. Baas, and D. H. Keisler. 2003. Serum concentrations of leptin in six genetic lines of swine and relationship with growth and carcass characteristics. J. Anim. Sci. 81:167–171.

Bernal Rubio, Y. L., J. L. Gualdrón Duarte, R. O. Bates, C. W. Ernst, D. Nonneman, G. A. Rohrer, A. King, S. D. Shackelford, T. L. Wheeler, R. J. C. Cantet, and J. P. Steibel. 2015a. Implementing meta-analysis from genome-wide association studies for pork quality traits 1. J. Anim. Sci. 93:5607–5617. http://doi.org/10.2527/jas2015-950

Bernal Rubio, Y. L., J. L. Gualdrón Duarte, R. O. Bates, C. W. Ernst, D. Nonneman, G. A. Rohrer, A. King, S. D. Shackelford, T. L. Wheeler, R. J. C. Cantet, and J. P. Steibel. 2015b. Meta-analysis of genome-wide association from genomic prediction models. Anim. Genet.:36–48. http://doi.org/10.1111/age.12378

Chen, C. Y., I. Misztal, I. Aguilar, S. Tsuruta, the Meuwissen, S. E. Aggrey, T. Wing, and W. M. Muir. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotyping. J. Anim. Sci. 89:23. http://doi.org/10.2527/jas.2010-3071

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. http://doi.org/10.3389/fgene.2011.00018

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 3 -Subunit Gene Associated With Low Glycogen Content in Pig Skeletal Muscle and Improved Meat Quality. Genetics 159:1151–1162.

Couturier, C., C. Sarkis, K. Séron, S. Belouzard, P. Chen, A. Lenain, L. Corset, J. Dam, V. Vauthier, A. Dubart, J. Mallet, P. Froguel, Y. Rouillé, and R. Jockers. 2007. Silencing of OB-RGRP in mouse hypothalamic arcuate nucleus increases leptin receptor signaling and prevents diet-induced obesity. Proc. Natl. Acad. Sci. U. S. A. 104:19476–19481. http://doi.org/10.1073/pnas.0706671104

da Costa, a. S. H., V. M. R. Pires, C. M. G. a. Fontes, and J. a. M. Prates. 2013. Expression of genes controlling fat deposition in two genetically diverse beef cattle breeds fed high or low silage diets. BMC Vet. Res. 9:118. http://doi.org/10.1186/1746-6148-9-118

Edwards, D. B., C. W. Ernst, N. E. Raney, M. . Doumit, M. D. Hoge, and R. O. Bates. 2008. Quantitative trait loci mapping in an F2 Duroc x Pietrain resource population: II. Carcass and meat quality traits. J. Anim. Sci. 86:254:266. http://doi.org/10.2527/jas.2006-626

Edwards, D., R. Bates, and W. Osburn. 2003. Evaluation of Duroc-vs. Pietrain-sired pigs for carcass and meat quality measures. J. Anim. Sci.:1895–1899.

Ernst, C. W., P. A. Kapke, M. Yerle, and M. F. Rothschild. 1996. The leptin receptor gene (LEPR) maps to porcine chromosome 6. Mamalian genome 8:226.

Fan, Y., Y. Xing, Z. Zhang, H. Ai, Z. Ouyang, J. Ouyang, M. Yang, P. Li, Y. Chen, J. Gao, L. Li, L. Huang, and J. Ren. 2013. A further look at porcine chromosome 7 reveals VRTN variants associated with vertebral number in Chinese and Western pigs. PLoS One 8:e62534. http://doi.org/10.1371/journal.pone.0062534

Forneris, N. S., A. Legarra, Z. G. Vitezica, S. Tsuruta, I. Aguilar, I. Misztal, and R. J. C. Cantet. 2015. Quality control of genotypes using heritability estimates of gene content at the marker. Genetics 199:675–81. http://doi.org/10.1534/genetics.114.173559

Geldermann, H., E. Muller, G. Moser, G. Reiner, H. Bartenschlager, S. Cepica, A. Stratil, J. Kuryl, C. Moran, R. Davoli, and C. Brunsch. 2003. Genome-wide linkage and QTL mapping in porcine F2 families generated from Pietrain, Meishan and Wild Boar crosses. J. Anim. Breed. Genet. 120:363–393. http://doi.org/10.1046/j.0931-2668.2003.00408

Goddard, M. E., and B. J. Hayes. 2007. Genomic selection. J.Anim.Breed.Genet 124:323–330. http://doi.org/10.1080/09064700801959395

Goll, D. E., V. F. Thompson, R. G. Taylor, and T. Zalewska. 1992. Is calpain activity regulated by membranes and autolysis or by calcium and calpastatin? Bioessays 14:549–56. http://doi.org/10.1002/bies.950140810

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. http://doi.org/10.1186/1471-2156-14-38

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. http://doi.org/10.1186/1471-2105-15-246

Gualdrón Duarte, J. L.; Cantet, R. J. C.; Bernal Rubio, Y. L.; Bates, R. O.; Ernst, C. W.; Raney,

N. E.; Rogberg-Muñoz, A.; Steibel, J. P. 2016. Refining genome wide association for growth and fat deposition traits in an F2 pig population1. Journal of Animal Science 94 http://doi.org/10.2527/jas2015-0182

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433–443. http://doi.org/10.3168/jds.2008-1646

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic Prediction: Coat colour, Milk-fat percentage, and type in holstein cattle as contrasting model traits. PLoS Genet. 6. http://doi.org/10.1371/journal.pgen.1001139

Hayes, B. J. 2013. Overview of statistical methods for genome-wide association studies. C. Gondor, B. van der Werf, & B. J. Hayes, editors, Genome-wide association studies and genomic prediction. Human Press, New York, NY. p. 156-157.

Hu, Z.-L., C. A. Park, and J. M. Reecy. 2015. Developmental progress and current status of the Animal QTLdb. Nucleic Acids Res. 44:D827–833. http://doi.org/10.1093/nar/gkv1233

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. http://doi.org//1996.745993x

Janss, L., G. de los Campos, N. Sheehan, and D. Sorensen. 2012. Inferences from genomic models in stratified populations. Genetics 192:693–704. http://doi.org/10.1534/genetics.112.141143

Jogl, G., Y. S. Hsiao, and L. Tong. 2004. Structure and function of carnitine acyltransferases. Ann. N. Y. Acad. Sci. 1033:17–29. http://doi.org/10.1196/annals.1320.002

Kim, E. S., R. Ros-Freixedes, R. N. Pena, T. J. Bass, J. Estany, and M. F. Rothschild. 2015. Identification of signatures of selection for intramuscular fat and backfat thickness in teo Duroc populations. J. Anim. Sci. 93:3293–3302. http://doi.org/10.2527/jas2015-8879

Kim, J. 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. http://doi.org/83/6/1229

King, J. W. B., and R. C. Roberts. 1960. Carcass length in the bacon pig; its association with vertebrae numbers and prediction from radiographs of the young pig. Anim. Prod. 2:59–65. http://doi.org/10.1017/S0003356100033493

Koohmaraie, M. 1992. Ovine skeletal muscle multicatalytic proteinase complex (proteasome): purification, characterization, and comparison of its effects on myofibrils with mu-calpains. J. Anim. Sci. 70:3697–3708. http://doi.org/1474009

Kühn, K. 1995. Basement membrane (type IV) collagen. Matrix Biol. 14:439–445.

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. 88:2904–2912. http://doi.org/10.2527/jas.2009-2590

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. http://doi.org/10.1111/j.1365-2052.2007.01592.x

Lonergan, S. M., E. Huff-Lonergan, L. J. Rowe, D. L. Kuhlers, and S. B. Jungst. 2001. Selection for lean growth efficiency in Duroc pigs influences pork quality. J. Anim. Sci. 79:2075–2085. http://doi.org//2001.7982075x

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. http://doi.org/10.1371/journal.pone.0064047

Malek, M., J. C. M. Dekkers, H. K. Lee, T. J. Baas, and M. F. Rothschild. 2001. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition. Mamm. Genome 12:630–636. http://doi.org/10.1007/s003350020018

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. http://doi.org/11290733

Milan, D., J. P. Bidanel, N. Iannuccelli, J. Riquet, Y. Amigues, J. Gruand, P. Le Roy, C. Renard, and C. Chevalet. 2002. Detection of quantitative trait loci for carcass composition traits in pigs. Genet. Sel. Evol. 34:705–728. http://doi.org/10.1051/gse

Milan, D., J. T. Jeon, C. Looft, V. Amarger, A. Robic, M. Thelander, C. Rogel-Gillard, 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. http://doi.org/10.1126/science.288.5469.1248

Muñoz, G., E. Alcazar, A. Fernández, C. Barragán, A. Carrasco, E. de Pedro, L. Silio, J. L. Sanchez, and M. C. Rodriguez. 2011. Effects of porcine MC4R and LEPR polymorphisms, gender and Duroc sire line on economic traits in Duroc??Iberian crossbred pigs. Meat Sci. 88:169–173. http://doi.org/10.1016/j.meatsci.2010.12.018

Muñoz, G., C. Ovilo, L. Silló, 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. 87:459–468. http://doi.org/10.2527/jas.2008-1127

Nakano, H., S. Sato, Y. Uemoto, T. Kikuchi, T. Shibata, H. Kadowaki, E. Kobayashi, and K. Suzuki. 2015. Effect of VRTN gene polymorphisms on Duroc pig production and carcass

traits, and their genetic relationships. Anim. Sci. J. 86:125–131. http://doi.org/10.1111/asj.12260

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 D. J. Nonneman, S. D. Shackelford, D. A. King, T. L. Wheeler, R. T. Wiedmann, W. M. Snelling and G. A. Rohrer. J. Anim. Sci.:4043–4050. http://doi.org/10.2527/jas2013-6255

Ovilo, C., A. Fernández, J. L. Noguera, C. Barragán, R. Letón, C. Rodríguez, A. Mercadé, E. Alves, J. M. Folch, L. Varona, and M. Toro. 2005. Fine mapping of porcine chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross. Genet. Res. 85:57–67. http://doi.org/10.1017/S0016672305007330

Patel, K. G., C. Liu, P. L. Cameron, and R. S. Cameron. 2001. Myr 8, a novel unconventional myosin expressed during brain development associates with the protein phosphatase catalytic subunits 1alpha and 1gamma1. J. Neurosci. 21:7954–7968.

Ramos, A. M., R. P. M. A. 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. http://doi.org/10.1371/journal.pone.0006524

Rohrer, G. a, and J. W. Keelen. 1998. Identification of quantitative trait loci affecting Carcass Composition in swine: I. Fat Deposition Traits. J. Anim. Sci. 76:2247–2254.

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. 92:511–518. http://doi.org/10.1016/j.meatsci.2012.05.020

Rohrer, G. a., R. M. Thallman, S. Shackelford, T. Wheeler, and M. Koohmaraie. 2005. A genome scan for loci affecting pork quality in a Duroc-Landrace F2 population. Anim. Genet. 37:17–27. http://doi.org/10.1111/j.1365-2052.2005.01368.x

Ros-Freixedes, R., S. Gol, R. N. Pena, M. Tor, J. C. M. Dekkers, and J. Estany. 2014. Proceedings, 10. In: Genome-Wide Association Study for Intramuscular Fat content and Composition in Duroc Pigs.

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. http://doi.org/10.1186/1471-2156-13-66

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. http://doi.org/10.1186/1297-9686-46-12

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478. http://doi.org/10.1186/1471-2164-15-478

Sato, S., and Y. Oyamada. 2003. Quantitative trait loci analysis for growth and carcass traits in a Meishan× Duroc F2 resource population. J. Anim. …:2938–2949.

Steibel, J. P., R. O. Bates, G. J. M. Rosa, R. J. Tempelman, V. D. Rilington, A. Ragavendran, N. E. Raney, A. M. Ramos, F. F. Cardoso, D. B. Edwards, and C. W. Ernst. 2011. Genome-wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs. PLoS One 6. http://doi.org/10.1371/journal.pone.0016766

Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. 2003.

Storey, J. D. 2002. A direct approach to false discovery rates. J. R. Stat. Soc. Series B. Stat. Methodol.:479–498. http://doi.org/10.1111/1467-9868.00346

Tartaglia, L. a. 1997. The leptin receptor. J. Biol. Chem. 272:6093–6096. http://doi.org/10.1074/jbc.272.10.6093

Thomsen, H., H. K. Lee, M. F. Rothschild, M. Malek, and J. C. M. Dekkers. 2004. Characterization of quantitative trait loci for growth and meat quality in a cross between commercial breeds of swine 1. :2213–2228.

Uemoto, Y., T. Kikuchi, H. Nakano, S. Sato, T. Shibata, H. Kadowaki, K. Katoh, E. Kobayashi, and K. Suzuki. 2012. Effects of porcine leptin receptor gene polymorphisms on backfat thickness, fat area ratios by image analysis, and serum leptin concentrations in a Duroc purebred population. Anim. Sci. J. 83:375–385. http://doi.org/10.1111/j.1740-0929.2011.00963.x

Uemoto, Y., and Y. Nagamine. 2008. Quantitative trait loci analysis on Sus scrofa chromosome 7 for meat production, meat quality, and carcass traits within a Duroc purebred population. J. Anim. …. http://doi.org/10.2527/jas.2007-0293

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. http://doi.org/10.2527/jas.2012-5811

Uimari, P., and A. Sironen. 2014. A combination of two variants in PRKAG3 is needed for a positive effect on meat quality in pigs. BMC Genet. 15:29. http://doi.org/10.1186/1471-2156-15-29

VanRaden, P. M. 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91:4414–4423. http://doi.org/10.3168/jds.2007-0980

Van Der Rest, M., and R. Garrone. 1991. Collagen family of proteins. FASEB J. 5:2814–2823.

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. http://doi.org/10.1017/s0016672312000274

Wang, K., D. Liu, J. Hernandez-Sanchez, J. Chen, C. Liu, Z. Wu, M. Fang, and N. Li. 2015. Genome Wide Association Analysis Reveals New Production Trait Genes in a Male Duroc Population. PLoS One 10:e0139207. http://doi.org/10.1371/journal.pone.0139207

Wong, W., and J. . Scott. 2004. AKAP signalling complexes: focal points in space and time. Nat. Rev. 5:959–970.

Yoo, C.-K., I.-C. Cho, J.-B. Lee, E.-J. Jung, H.-T. Lim, S.-H. Han, S.-S. Lee, M.-S. Ko, T. Kang, J.-H. Hwang, Y. S. Park, and H.-B. Park. 2014. QTL analysis of body weight and carcass body length traits in an F2 intercross between Landrace and Korean native pigs. Anim. Genet. 45:589–592. http://doi.org/10.1152/physiolgenomics.00172.2011

Zhang, C., Z. Wang, H. Bruce, R. A. Kemp, P. Charagu, Y. Miar, T. Yang, and G. Plastow. 2015. Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs. BMC Genet. 16:1–12. http://doi.org/10.1186/s12863-015-0192-1

Zhang, J. h., Y. Z. Xiong, B. Zuo, M. G. Lei, F. E. Li, and J. L. Li. 2007. Detection of Quantitative Trait Loci Associated with Live Measurement Traits in Pigs. Agric. Sci. China 6:863–868. http://doi.org/10.1016/S1671-2927(07)60123-0

CHAPTER FOUR

Conclusion

GOALS AND CONTRIBUTIONS OF THIS STUDY

Meat quality traits are subject to genetic control, and are negatively correlated with lean growth traits (van Wijk et al., 2005). For a long time, the swine industry selected animals for lean growth and efficiency traits, inducing a softer and exudative meat (Lonergan et al., 2001). Additionally, using traditional breeding methods such as selection index (Hazel, 1943) is challenging for traits expressed later in life. With the development of statistical models for genomic selection (Meuwissen et al., 2001; Goddard & Hayes, 2007), genomic breeding values can be estimated, and animals can be selected earlier in life. However, genomic selection relies mostly on LD between markers and causative variants [such as Quantitative Trait Nucleotides (QTN)] as opposed to directly exploiting QTN (Weller & Ron, 2011). Detecting QTN has potential value in implementing selection across multiple population or for simultaneous multi-trait selection. More recently, the importance of knowing QTN has been highlighted by the prospect of using genome editing (Jenko et al., 2015). To detect a QTN, a Genome-Wide Association (GWA) is typically performed first, but in livestock sizable genomic regions are usually defined by GWA due to long range persistence of LD. Therefore, there is a need to refine the genomic regions using a method that can maximize the chances of covering the causal variants. The goals of this study were:

1. Implement and test properties of methods for computing the confidence interval of a QTL position in the context of GWA from mixed effects Genomic Best Linear Unbiased Predictor (GBLUP) models.

2. Perform GWA of meat quality and carcass traits in an F2 Duroc x Pietrain resource population using the methods tested under aim 1 and propose candidate genes for further study.

In chapter 2, I used plasmode simulations to test the properties of three methods to compute Confidence Intervals (CI) for the position of Quantitative Trait Loci (QTL). Hayes, (2013) proposed a parametric method (PM) and I proposed two related non-parametric alternatives: one where the CI is centered around the QTL peak (NPC), and another one that produced asymmetric CI, non centered around the QTL peak (NPNC). I focused on two key properties of CI: Realized statistical coverage and length of the interval. The realized coverage of an optimal CI has to reach its nominal level using the shortest interval possible. The NPC failed to provide adequate coverage for the 95% CI for the QTL position. The NPNC and PM had very similar coverage, close to the nominal level (96% and 96.5% respectively). The asymmetry of the CI obtained with NPNC made this method adaptable to the density of significant SNP around the QTL peak at the expense of a longer CI length. None of the NPNC and PM were uniformly better than the other. In some cases, the 95% CI derived with the NPNC covered the true QTL position when the 95% CI derived with PM do not, and vice-versa. However, PM produced intervals on average 20% shorter than those from the NPNC. Therefore, I recommended to use the PM method for the calculation of CI for the QTL position in GWA and I used the method extensively in chapter 3. In chapter 3 I performed a GBLUP-based GWA in an F2 Duroc x Pietrain Resource Population for 38 meat quality and carcass traits and used statistical support intervals to map QTL. I found nine QTL associated with 15 traits on 8 chromosomes. Seven QTL had been previously reported. From those seven QTL, three (one on SSC1, tenth rib backfat thickness; one on SSC7, dressing percentage and loin muscle area; and one on SSC11, belly weight) had been previously mapped

only using low resolution linkage analyses. In this work, for the first time, those three QTL have been physically mapped to specific genomic segments. As a result, I proposed the gene Carnitine O-Acetyltransferase (*CRAT*; SSC1:303.4-303.41 Mb), which is related to lipid metabolism, as a candidate for tenth rib carcass backfat thickness.

Two QTL associated with sensory panel tenderness on SSC3 and on SSC5 are novel findings of this study. The gene A kinase (PRKA) anchor protein 3 (AKAP3; SSC5: 68 to 68.02 Mb) is a novel candidate gene proposed for tenderness in this thesis. AKAP3 can bind to the regulatory subunit of PKA affecting the glycogen content in muscle, affecting the quality of the meat.

Finally, I studied an association peak for pH 24 hours post-mortem, drip loss and cook yield located close to the well-known candidate gene Protein Kinase AMP-activated γ 3-subunit (*PRKAG3*). My follow-up analysis focused on the association to two well-known non-synonymous variants (I199V and T30N) of PRKAG3 that have been proposed as candidate variants for pH 24 hours post mortem, drip loss and cook yield. In this thesis, I showed that in this population, those variants do not fully explain the genotypes of the QTL found on SSC15, associated with juiciness, tenderness, drip loss, pH 24 hours, WBS, cook yield and protein content. This means that the MSUPRP population remains a valuable resource to further discover more causative variants for pH 24 hours and related traits, either within PRKAG3 or in its close vicinity.

FUTURE RESEARCH DIRECTIONS:

In this study I evaluated the properties of CI (length and coverage) considering single association peaks on one chromosome. However, there might be multiple peaks on one chromosome associated with one trait. Visscher et al., (1996), addressed this problem and proposed to calculate the CI separately, but they never tested if the properties of the CI were similar to the ones for a single QTL. This study did not test Visscher's recommendation, because I only simulated one QTL per plasmode. Thus testing the properties of CI for multiple QTL should be studied. The hypothesis of the proposed study would be that the CI obtained with data partition methods applied independently to multiple QTL will retain the desirable properties shown in Chapter 2 for single QTL. To test the hypothesis, a new plasmode study, simulating two or more QTL peaks in different positions on one chromosome should be performed. The 95% CI of each QTL peak position in a plasmode dataset must be computed using the three proposed methods. Furthermore, coverage should be computed following methods presented in Chapter 2 to determine if the realized coverage of CI is equal to their nominal level. After confirming that the coverage is adequate, the average length of the CI obtained with each method should be compared.

According to the GWA results from this study, further work needs to be done for a) resequencing certain regions to find causal variants and improve the annotation, b) validating the proposed candidate genes. Therefore, I propose:

1. Validating the candidate gene CRAT.

   This study found a QTL located on SSC1 between 302.9-307.1 Mb. Carnitine O-Acetyltransferase (CRAT) is one of the functional genes annotated in that genomic region. This enzyme is involved in lipid metabolism and it was shown in beef cattle to be

differentially expressed in subcutaneous tissue (da Costa et al., 2013). Therefore, I propose a three tier analysis to validate this candidate. First, a differential gene expression analysis can be done with the data of the 176 F2 Duroc x Pietrain animals that is already available in the fat tissue, to compare the expression level of the gene in animals with extreme fat deposition phenotypes. In addition to the in-silico validation, an in vitro experiment could be conducted using adipocytes to assess the potential roll of CRAT in adipocyte growth. Finally, if results of the other two studies are promising, an in vivo study with knockout mouse could confirm the gene's role.

2. Resequencing the genomic region on SSC6 associated with fat traits.

   The genomic region between 127.6-140.8 Mb. on SSC6 associated with backfat traits, contains the gene Leptin Receptor Overlapping Transcript (LEPROT), which negatively regulates leptin cell surface exposed receptor (Couturier et al., 2007). Moreover, the Leptin Receptor gene (LEPR) encodes for multiple isoforms of the leptin receptor (Tartaglia, 1997). Causative variants in LEPR have also been associated with carcass measurements such as backfat traits (Ovilo et al., 2005; Muñoz et al., 2009; Muñoz et al., 2011; Uemoto et al., 2012). Although it has been reported that LEPR maps to SSC6 (Ernst et al., 1996), we were not able to find this gene, because it is located on an unassigned contig in the current pig genome assembly (version 10.2.83) and thus it is not annotated in the current assembly. Therefore, resequencing this genomic region could be beneficial to a) annotate the LEPR gene in that region, b) discover SNP in the gene, c) confirm if those SNP are associated with the phenotypes of interest.

LITERATURE CITED

LITERATURE CITED

Costa, a. S. H., Pires, V. M. R., Fontes, C. M. G. a., & Prates, J. a. M. (2013). Expression of genes controlling fat deposition in two genetically diverse beef cattle breeds fed high or low silage diets. *BMC Veterinary Research*, *9*, 118. http://doi.org/10.1186/1746-6148-9-118

Couturier, C., Sarkis, C., Séron, K., Belouzard, S., Chen, P., Lenain, A., … Jockers, R. (2007). Silencing of OB-RGRP in mouse hypothalamic arcuate nucleus increases leptin receptor signaling and prevents diet-induced obesity. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19476–19481. http://doi.org/10.1073/pnas.0706671104

Ernst, C. W., Kapke, P. A., Yerle, M., & Rothschild, M. F. (1996). The leptin receptor gene (LEPR) maps to porcine chromosome 6. *Mamalian Genome*, *8*, 226.

Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *J.Anim.Breed.Genet*, *124*, 323–330. http://doi.org/10.1080/09064700801959395

Hayes, B. J. (2013). *Overview of statistical methods for genome-wide association studies. C. Gondor, B. van der Werf, & B. J. Hayes, editors, Genome-wide association studies and genomic prediction. Human Press, New York, NY. p. 156-157*.

Hazel, L. N. (1943). The Genetic Basis for Constructing Selection Indexes. *Genetics*, *28*(6), 476–490.

Jenko, J., Gorjanc, G., Cleveland, M. A., Varshney, R. K., Whitelaw, C. B. A., Woolliams, J. A., & Hickey, J. M. (2015). Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. *Genetics, Selection, Evolution : GSE*, *47*(1), 55. http://doi.org/10.1186/s12711-015-0135-3

Lonergan, S. M., Huff-Lonergan, E., Rowe, L. J., Kuhlers, D. L., & Jungst, S. B. (2001). Selection for lean growth efficiency in Duroc pigs influences pork quality. *Journal of Animal Science*, *79*(8), 2075–2085. http://doi.org//2001.7982075x

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829. http://doi.org/11290733

Muñoz, G., Alcazar, E., Fernández, A., Barragán, C., Carrasco, A., de Pedro, E., … Rodriguez, M. C. (2011). Effects of porcine MC4R and LEPR polymorphisms, gender and Duroc sire line on economic traits in Duroc??Iberian crossbred pigs. *Meat Science*, *88*(1), 169–173. http://doi.org/10.1016/j.meatsci.2010.12.018

Muñoz, G., Ovilo, C., Silló, L., Tomás, A., Noguera, J. L., & Rodríguez, M. C. (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. *Journal of Animal Science*, *87*(2), 459–468. http://doi.org/10.2527/jas.2008-1127

Ovilo, C., Fernández, A., Noguera, J. L., Barragán, C., Letón, R., Rodríguez, C., … Toro, M. (2005). Fine mapping of porcine chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross. *Genetical Research*, *85*(1), 57–67. http://doi.org/10.1017/S0016672305007330

Tartaglia, L. a. (1997). The leptin receptor. *J. Biol. Chem.*, *272*, 6093–6096. http://doi.org/10.1074/jbc.272.10.6093

Uemoto, Y., Kikuchi, T., Nakano, H., Sato, S., Shibata, T., Kadowaki, H., … Suzuki, K. (2012). Effects of porcine leptin receptor gene polymorphisms on backfat thickness, fat area ratios by image analysis, and serum leptin concentrations in a Duroc purebred population. *Animal Science Journal*, *83*(5), 375–385. http://doi.org/10.1111/j.1740-0929.2011.00963.x

van Wijk, H. J., Arts, D. J. G., Matthews, J. O., Webster, M., Ducro, B. J., & Knol, E. F. (2005). Genetic parameters for carcass composition and pork quality estimated in a commercial production chain. *J. Anim. Sci*, *83*, 324–333.

Visscher, P. M., Thompson, R., & Haley, C. S. (1996). Confidence Intervals in QTL Mapping by Bootstrapping. *Genetics*, *143*, 1013–1020.

Weller, J. I., & Ron, M. (2011). Invited review: quantitative trait nucleotide determination in the era of genomic selection. *Journal of Dairy Science*, *94*(3), 1082–1090. http://doi.org/68/jds.2010-3793