Genome-wide mapping and functional analysis of genes determining the meat quality in pigs

DISSERTATION

zur Erlangung des akademischen Grades

Doctor scientiarum agriculturae

(Dr. sc. agr.)

vorgelegt

dem Institut für Tierhaltung und Tierzüchtung

der Universität Hohenheim

von

Patrick Stefan Stratz

Dipl. Agr. Biol.

aus Furtwangen

Baden-Württemberg

Hohenheim, 2014

Dekanin: Prof. Dr. M. Brockmeier

1. Prüfer: Berichterstatter: Prof. Dr. J. Bennewitz

2. Prüfer: Mitberichterstatter: Prof. Dr. K. Wimmers

3. Prüfer: Prof Dr. A. E. Melchinger

Tag der mündlichen Prüfung: 11.07.2014

Die Dissertation wurde mit dankenswerter Unterstützung der Carl-Zeiss Stiftung angefertigt.

Table of Contents

General Summary

The main objective of the research project was genome-wide mapping and functional analysis of genes determining meat quality in pigs.

In **chapter one** QTL were mapped and tested for pairwise epistasis for meat quality traits in three connected porcine F_2 crosses comprising around 1000 individuals. The crosses were derived from Chinese Meishan, European Wild Boar and Piétrain. The animals were genotyped genomewide for approximately 250 genetic markers and phenotyped for seven meat quality traits. QTL mapping was done using a multi-QTL multi-allele model. It considered additive (*a*), dominance (*d*) and imprinting (*i*) effects. The major gene *RYR1:G.1843C>T* affecting the meat quality was included as a cofactor in the model. The mapped QTL were tested for possible epistatic effects between the main effects, leading to nine orthogonal forms of epistasis (*aa*, *ad*, *da*, *di*, *id*, *ai*, *ia*, *dd* and *ii*). Numerous QTL were found; the most interesting are located on chromosome SSC6. Epistasis was significant (FDR *q-value*<0.2) for the pairwise QTL on SSC12 and SSC14 for pH 24 h after slaughter and for the QTL on SSC2 and SSC5 for rigour.

In **chapter two** around 500 progeny tested Piétrain sires were genotyped with the PorcineSNP60 BeadChip. After data filtering around 48k SNPs were useable in this sample. These SNPs were used to conduct a genome-wide association analysis for growth, muscularity and meat quality traits. Because it is known, that a mutation in the RYR1 gene located on chromosome 6 shows a major effect on meat quality, this mutation was included in the models. Single-marker and multi-marker association analysis were performed. The results revealed between one and eight significant associations per trait with *P-value*<0.00005. Of special interest are SNPs located on SSC6, 10 and 15.

In **chapter three** a literature search was conducted to search putative candidate genes in the vicinity of significant SNPs found in the association analysis. *MYOD1* was suggested as putative candidate gene. The expression of *MYOD1* was measured in muscle tissue from 20 Piétrain sires. Growth, muscularity and meat quality traits were available. DNA was isolated out of blood tissue to genotype the SNP ASGA0010149:g. 47980126G>A. Significant Correlations (FDR *q-value*<0.15) between the expression of *MYOD1* and growth and muscularity traits were found. Association between the traits, respectively *MYOD1*, and

ASGA0010149:g. 47980126G>A was tested, but was only significant (FDR *q-value*<0.15) for two muscularity traits.

In **chapter four** the LD structure in the genome of the Piétrain pigs was characterized using data from the PorcineSNP60 BeadChip. The Relative Extended Haplotype Homozygosity test was conducted genome-wide to search for selection signatures using core haplotypes above a frequency of 0.25. The test was also conduct in targeted regions, where significant SNPs were already found in association analysis. A small subdivision of the population with regard to the geographical origin of the individuals was observed. As a measure of the extent of linkage disequilibrium, r^2 was calculated genome-wide for SNP pairs with a distance ≤ 5Mb and was on average 0.34. Six selection signatures having a *P-value*<0.001 were genome-wide detected, located on SSC1, 2, 6 and 17. In targeted regions, it was possible to successfully annotate nine SNPs to core regions. Strong evidence for recent selection was not found in those regions. Three selection signatures with *P-value*<0.1 were detected on SSC2, 5 and 16.

To reduce the costs of genomic selection, selection candidates can be genotyped with an SNP panel of reduced density (384 SNPs). The aim of **chapter five** was to investigate two strategies for the selection of SNPs to be considered in the above mentioned SNP panel, using 895 progeny tested and genotyped German Piétrain boars. In the first strategy equal spaced SNPs were selected, which were used to impute the high density genotypes. In the second strategy SNPs were selected based on results of association analysis. Direct genomic values were estimated with GBLUP from deregressed estimated breeding values. Accuracies of direct genomic values for the two strategies were obtained from cross validation. A regression approach to correct for the upward bias of the cross validation accuracy of the direct genomic values was used. The first strategy resulted in more accurate direct genomic values. This implies that imputation is beneficial even if only 384 SNPs are genotyped for the selection candidates.

This thesis ends with a general discussion which addresses substantive terms once again to associate them with the most recent research.

Allgemeine Zusammenfassung

Das Ziel dieses Forschungsprojektes war es Gene, welche die Fleischqualität beim Schwein determinieren, genomweit zu kartieren und funktionell zu analysieren.

In **Kapitel eins** wurden QTL für Fleischqualitätsmerkmale in drei verbundenen F_2 Kreuzungen bestehend aus ca. 1000 Individuen kartiert und auf paarweise Epistasie getestet. Die Kreuzungen sind aus den Ausgangsrassen Chinesisches Meishanschwein, Europäisches Wildschwein und dem Piétrainschwein hervorgegangen. Die Tiere wurden genomweit an ca. 250 genetischen Markern genotypisiert und für sieben Fleischqualitätsmerkmale phänotypisiert. Die QTL Kartierung erfolgte mit Hilfe eines multi-QTL multi-Allel Modells. Dabei wurden additiv- (*a*), dominanz- (*d*) und imprinting Effekte (*i*) berücksichtigt. Das Majorgen *RYR1:G.1843C>T*, das die Fleischqualität beeinflusst, wurde als Kofaktor ins Modell mit aufgenommen. Die kartierten QTL wurden auf mögliche paarweise epistatische Effekte zwischen den Haupteffekten untersucht, was zu neun orthogonalen Formen der Epistasie führte (*aa*, *ad*, *da*, *di*, *id*, *ai*, *ia*, *dd* und *ii*). Zahlreiche QTL konnten gefunden werden; die Interessantesten davon befanden sich auf SSC6. Die Epistasie war für die paarweisen QTL auf SSC12 und SSC14 für den pH-Wert, gemessen 24h nach der Schlachtung, und für die QTL auf SSC2 und SSC5 für Rigor signifikant (FDR *q-Wert*<0.15).

In **Kapitel zwei** wurden ca. 500 Nachkommen geprüfte Eber mit Hilfe des PorkinenSNP60 BeadChip genotypisiert; davon waren in dieser Stichprobe nach Datenfilterung ca. 48k SNPs zu gebrauchen. Mit Hilfe dieser SNPs wurden genomweite Assoziationsstudien für die Merkmale des Wachstums, der Bemuskelung und der Fleischqualität durchgeführt. Da bekannt ist, dass eine Mutation im *RYR1* Gen, lokalisiert auf SSC6, einen Majoreffekt auf die Fleischqualität zeigt, wurde diese Mutation in die Modelle mit aufgenommen. Einzel- und multi-Marker Assoziationsstudien wurden durchgeführt. Mit einer Irrtumswahrscheinlichkeit von *P-Wert*<0.00005 konnten zwischen einer und acht signifikante Assoziationen pro Merkmal gefunden werde. Von besonderer Bedeutung sind dabei SNPs auf SSC6, 10 und 15.

In **Kapitel drei** wurde in unmittelbarer Nachbarschaft zu signifikanten SNPs aus den genomweiten Assoziationsstudien, mit Hilfe der Literatur, nach putativen Kandidatengenen gesucht. *MYOD1* wurde als putatives Kandidatengen vorgeschlagen. Die Expression von

MYOD1 wurde im Muskelgewebe von 20 Piétrain Ebern gemessen. Wachstums-, Muskel- und Fleischqualitätsmerkmale waren verfügbar. Die DNA wurde aus dem Blut isoliert und der SNP ASGA0010149:g. 47980126G>A genotypisiert. Signifikante Korrelationen (FDR *q-Wert*<0.15) der Expression von *MYOD1* mit Wachstums- und Bemuskelungsmerkmalen wurden gefunden. (FDR *q-Wert*<0.15). Die Assoziation zwischen den Merkmalen, respektive der Expression von *MYOD1* und dem SNP ASGA0010149:g. 47980126G>A, wurde getestet. Dieser war jedoch nur für zwei Bemuskelungsmerkmale signifikant (FDR *q-Wert*<0.15).

In **Kapitel vier** wurde die LD Struktur im Genom der Piétrain Schweine mit Hilfe der PorkinenSNP60 BeadChip Daten charakterisiert. Der Relative Extended Haplotype Homozygosity Test wurde genomweit zur Suche nach Selektionssignaturen durchgeführt. Dabei wurden nur Kernhaplotypen oberhalb einer relativen Häufigkeit von 0.25 betrachtet. Der Test wurde zusätzlich in Zielregionen angewandt in denen bereits in Assoziationsstudien signifikante SNPs gefunden wurden. Im Hinblick auf den geographischen Ursprung der Individuen konnte eine geringfügige Unterteilung der Population festgestellt werden. Als Maß des Kopplungsungleichgewichtes wurde r^2, genomweit für SNP Paare die $\leq$ 5MB auseinander liegen, berechnet. Dieser lag im Durchschnitt bei 0.34. Genomweit konnten sechs Selektionssignaturen mit einem *P-Wert*<0.001 detektiert werden. Diese sind auf SSC1, 2, 6 und 17 lokalisiert. In Zielregionen konnten neun SNPs erfolgreich Kernregionen zugeordnet werden. Diese wiesen jedoch keine eindeutigen Anzeichen von jüngster Selektion auf. Drei Selektionssignaturen auf SSC2, 5 und 16 hatten einen *P-Wert*<0.1.

Um die Kosten der genomischen Selektion zu reduzieren können Selektionskandidaten mit einem SNP Panel von reduzierter Dichte (384 SNPs) genotypisiert werden. Das Ziel von **Kapitel fünf** war es zwei Strategien zur Auswahl von SNPs für das oben genannte Panel anhand 895 nachkommengeprüfter und genotypisierter Piétrain Ebern zu untersuchen. In Strategie eins wurden abstandsgleiche SNPs ausgewählt um Genotypen mit hoher Dichte zu imputieren. In Strategie zwei wurden SNPs basierend auf den Ergebnissen der Assoziationsanalysen ausgewählt. Direkte Genomische Werte wurden mittels GBLUP aus deregressierten Zuchtwerten geschätzt. Die Genauigkeiten dieser wurden mittels Kreuzvalidierung ermittelt. Zur Korrektur der nach oben verzerrten kreuzvalidierten Genauigkeiten der Direkten Genomischen Werte kam ein Regressionsansatz zur Anwendung. Die erste Strategie führte zu genauer geschätzten Direkten Genomischen Werten. Dieses

impliziert, dass die Imputierung sogar dann von Vorteil ist, wenn nur 384 SNPs der Selektionskandidaten genotypisiert werden.

Die Dissertationsschrift endet mit einer kapitelübergreifenden Diskussion. In dieser werden inhaltliche Aspekten nochmals aufgegriffen, um sie in Zusammenhang mit dem neusten Stand der Forschung zu bringen.

General Introduction

Domestication of the majority of today's livestock species took place during the Neolithic period in different regions around the world. A continuous selection of the livestock populations and their adaption for human purposes created traits which would not have been incurred in the natural environment. Targeted artificial selection led to anatomical and functional differences between productive livestock and wild types. In China, pigs were traditionally selected for increased subcutaneous and intramuscular fat which resulted in a higher fertility. In Europe, pigs were bred to grow fast and produce meat, which in turn affected the meat quality. Different factors influencing meat quality have been discussed (Huff-Lonergan and Lonergan 2007; Fischer 2007), and heritability estimates of 0.2-0.4 indicate considerable genetic variance (e.g. De Vries et al., 1994, Borchers et al., 2007).

In the past, quantitative trait loci (QTL) were mapped using experimental designs from genetically divergent and outbred founder pig breeds. The founder breeds were frequently chosen from the Asian and European type of breeds, because they revealed distinct lineages (Frantz et al., 2013). The disadvantage of linkage based approaches is that the power to map QTL is limited, because of the limited number of individuals in a typical F_2 cross and the associated limited number of usable meisosis. A potential remedy would be to pool the data from different F_2 crosses and to analyse them jointly (Rückert and Bennewitz, 2010). Some authors suggested the inclusion of epistasis in QTL mapping (Carlborg and Haley 2004).

With the advent of SNP chip technology in livestock species it became possible to conduct genome-wide association studies (GWAS) also within populations (Goddard and Hayes 2009). In GWAS the LD between SNPs and the causative mutation within a population is taken into account and also historical meiosis were utilised. To dissect the genetics of complex traits like meat quality, analysis of transcription profiles of genes is suggested by Ponsuksili et al. (2010). In contrast to association studies, likely targets of past selection, so called selection signatures, can also be identified without having knowledge about the traits they regulate. Such approaches reveal loci with outlier pattern of variation (Qanbari et al., 2010). Genomic selection uses a large number of SNP spread across the genome for breeding value estimation in a reference population and for subsequent selection of candidates based on gnomically enhanced breeding values (Meuwissen et al., 2001; Goddard and Hayes, 2009). Wellmann et al. (2013) suggested to decrease genotyping costs for selection candidates by using an SNP panel of reduced density. Beneath others the authors proposed a method for marker selection based on the results of GWAS.

The aim of this thesis was the genome-wide mapping and functional analysis of genes determining the meat quality in pigs. The present thesis is divided into nine chapters, whereas this chapter gives a general introduction.

In **chapter one** a two-step approach was applied to a powerful porcine experimental design consisting of three connected F_2-crosses derived from the three genetically different founder breeds Piétrain, Chinese Meishan and European Wild Boar (Geldermann et al., 2003). 966 performance tested F_2-animals were analyzed. In the first step, QTL for meat quality traits were mapped considering additive, dominant and imprinting effects and by applying a multi-QTL multi-allele model. The QTL mapped in the first step were tested for pairwise epistatic effects in the second step. Several QTL that partly showed pairwise epistatic effects were identified, also for imprinting which needs confirmation and further investigation.

The aim of **chapter two** was to conduct a genome-wide association study (GWAS) using 500 progeny tested German Piétrain boars. Individuals were genotyped with the Illumina PorcineSNP60 BeadChip (Ramos et al., 2009). Growth, muscularity, and meat quality traits were collected from progeny. Yield deviations were calculated from progeny and were used as trait variable for association studies. SNPs affecting growth, muscularity, and meat quality traits within Piétrain pigs were identified.

In **chapter three** genes being involved in trait determining pathways or metabolisms were searched in the vicinity of SNPs with non-negligible effects, which were identified in GWAS. Only one gene called *MYOD1* was found in the neighbourhood of the SNP ASGA0010149:g. 47980126G>A. 20 German Piétrain boars were used to study the expression of the putative candidate gene *MYOD1* in muscle tissue. The animals were also genotyped for the SNP ASGA0010149:g. 47980126G>A. Observations for growth, muscularity and meat quality traits were available for the boars. An association analysis was conducted by regressing the observations and expression levels on the number of A-alleles at the SNP_*MYOD1*. No association between the expression and the number of A-alleles was found.

Searching for substructures among 849 German Piétrain boars and characterizing their LD structure, based on r^2, was the aim of **chapter four**. Selection signatures were searched over the genome and in targeted regions, where SNPs with non-negligible effects were detected in

GWAS. Therefore the Extended Haplotype Homozygosity statistics (*EHH*, Sabeti et al., 2002) was applied. Illumina PorcineSNP60 BeadChip (Ramos et al., 2009) genotypes were abailible from the individuals. The population structure indicates that there is no specific modelling of the geographic origin of the animals required in GWAS or genomic selection. The low r^2 for all possible marker pairs estimated in this study indicates a high diversity in the Piétrain pig population. Strong evidence for recent selection was neither found over the genome nor in targeted regions.

The aim of **chapter five** was to apply two strategies for the selection of SNPs to be considered in a very low density SNP panel (384 SNPs). Illumina PorcineSNP60 BeadChip (Ramos et al., 2009) genotypes from 895 German Piétrain boars were availible. The boars were progeny tested for growth, muscularity and meat quality traits and split into a training and validation set. The validation set for imputation and genomic selection consisted of 100 boars, which were the youngest animal with high-density (HD) genotyped sires. The remaining 795 boars were included in the training set. In the first strategy, equal spaced SNPs were selected, which were used to impute the high density genotypes. In the second strategy SNPs were selected based on results of GWAS. Accuracies of direct genomic values for the two strategies were obtained from cross validation. In the first strategy more accurate direct genomic values were obtained than with the second strategy.

This thesis ends with a general discussion.

References

Carlborg, O., and C. S. Haley (2004). Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics* **5**: 618-25.

Frantz, L. A. F., J. G. Schraiber, O. Madsen, H .J. Megens, M. Bosse, Y. Paudel, G. Semiadi, E. Meijaard, N. Li, R. P. M. A. Crooijmans, A. L. Archibald, M. Slatkin, L. B. Schook, G. Larson, M. A. M. Groenen (2013). Genome sequencing reveals fine scale diversification and reticulation history during speciation. *Genome Biology* **14**: R107.

Fujii, J., K. Otsu, F. Zorzato, S. DeLeon, V. K. Khanna, J. E. Weiler, P.J. O'Brien, D. H. McLennan (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253**: 448-451.

Geldermann, H., E. Müller, G. Moser, G. Reiner, H. Bartenschlager, S. Cepica, A. Stratil, J. Kuryl, C. Moran, R. Danoli, C. Brunsch (2003). Genome-wide linkage and QTL mapping in porcine F2 families generated from Piétrain, Meishan and Wild Boar crosses. *Journal of Animal Breeding and Genetics* **120**: 363-393.

Goddard, M. E., and B. J. Hayes (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10**: 381–391.

Ponsuksili, S., E. Murani, M. Schwerin, K. Schellander, K. Wimmers (2010). Identification of expression QTL (eQTL) of genes expressed in porcine M. longissimus dorsi and associated with meat quality traits. *BMC Genomics* **11**: 572.

Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi, H. Simianer (2010). A Genome-Wide Scan for Signatures of Recent Selection in Holstein Cattle. *Animal Genetics* **44**: 377-389.

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. V. Tassell, J. F. Taylor, R. T. Wiedmann, L. B. Schook, 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.

Rückert, C., and Bennewitz J. (2010). Joint QTL analysis of three connected F2-crosses in pigs. *Genetics Selection Evolution* **42**: 40.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson, G. J. McDonald, H. C. Ackerman, S. J. Campbell, D. Altshuler, R. Cooper, D. Kwiatkowski, R. Ward, E. S. Lander (2002). Detecting

recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832–837.

Seidenspinner, T., J. Bennewitz, F. Reinhardt G. Thaller (2009). Need for sharp phenotypes in QTL detection for calving traits in dairy cattle. *Journal of Animal Breeding and Genetics* **126**: 455 – 462.

Wellmann, R., S. Preuß, E. Tholen, J. Heinkel, K. Wimmers, J. Bennewitz (2013). Genomic selection using low density marker panels with application to a sire line in pigs. *Genetics Selection Evolution* **45**: 28.

Chapter one

A two-step approach to map quantitative trait loci for meat quality in connected porcine F2 crosses considering main and epistatic effects

Patrick Stratz[a,*], Christine Baes[a,b], Christine Rückert[a], Siegfried Preuss[a], and Jörn Bennewitz[a]

[a] Institute of Animal Husbandry and Breeding,
University of Hohenheim,
D-70599 Stuttgart, Germany
[b] Swiss College of Agriculture,
Bern University of Applied Science,
3052 Zollikofen, Switzerland

Published in **Animal Genetics** (2012) 44: 14–23
doi: 10.1111/j.1365-2052.2012.02360.x

A two-step approach to map quantitative trait loci for meat quality in connected porcine F2 crosses considering main and epistatic effects

P. Stratz[a,*], C. Baes[a,b], C. Rückert[a], S. Preuss[a] and J. Bennewitz[a]

[a] Institute of Animal Husbandry and Breeding, University of Hohenheim, D-70599 Stuttgart, Germany

[b] Swiss College of Agriculture, Bern University of Applied Science, CH-3052 Zollikofen, Switzerland

*Corresponding author

Tel.: +49 711 459 22677

Fax: +49 711 459 23101

Email addresses:

PS: Patrick.Stratz@uni-hohenheim.de

CB: cba@suisag.ch

CR: christine.rueckert@uni-hohenheim.de

SP: Siegfried.Preuss@uni-hohenheim.de

JB: j.bennewitz@uni-hohenheim.de

Summary

The aim of the present study was to map QTL for meat quality traits in three connected porcine F2 crosses comprising around 1000 individuals. The three crosses were derived from the founder breeds Chinese Meishan, European Wild Boar and Piétrain. The animals were genotyped genomewide for approximately 250 genetic markers, mostly microsatellites. They were phenotyped for seven meat quality traits (pH at 45 minutes and 24 hours after slaughter, conductivity at 45 min and 24 h after slaughter, meat colour, drip loss and rigor). QTL mapping was conducted using a two-step procedure. In the first step the QTL were mapped using a multi-QTL multi-allele model that was tailored to analyse multiple connected F_2 crosses. It considered additive, dominance and imprinting effects. The major gene *RYR1:G.1843C>T* affecting the meat quality on SSC6 was included as a cofactor in the model. The mapped QTL were tested for pairwise epistatic effects in the second step. All possible epistatic effects between additive, dominant, and imprinting effects were considered, leading to nine orthogonal forms of epistasis. Numerous QTL were found. The most interesting chromosome was SSC6. Not all genetic variance of meat quality was explained by *RYR1* C1843T. A small confidence interval was obtained, which facilitated the identification of candidate genes underlying the QTL. Epistasis was significant for the pairwise QTL on SSC12 and SSC14 for pH24 and for the QTL on SSC2 and SSC5 for rigor. Some evidence for additional pairwise epistatic effects was found, although not significant. Imprinting was involved in epistasis.

Keywords: meat quality, pig, QTL, RYR, two-step procedure

Introduction

The majority of today's livestock species were domesticated during the Neolithic period (7,000 to 10,000 years ago) in different regions around the world. Anatomical and functional differences between productive livestock and wild types were accelerated by modern breeding technologies. In pig breeding one of the main selection criteria was the ability of pigs to grow fast and produce meat, which in turn affected the meat quality. In China, pig breeders traditionally concentrated on increasing subcutaneous and intramuscular fat. In contrast, selection for leanness, growth rate and feed efficiency in Europe has led to increased concern about meat quality.

The definition of meat quality is generic and depends, for example, on the proposed use of the meat. Meat quality can be considered as a combination of sensory properties and technological traits. Sensory properties include smell, taste and tenderness, and are usually recorded subjectively by educated tasters. Objectively measurable traits are, for example, pH-value, drip loss, meat colour and intramuscular fat content. Economically, drip loss likely plays the most important role as a measure of meat quality in Germany (Fischer 2007). Different factors influencing the meat quality have been discussed (Huff-Lonergan and Lonergan, 2007; Fischer 2007), and heritability estimates of 0.2-0.4 indicate considerable genetic variance (e.g. De Vries et al. 1994, Borchers et al. 2007).

Meat quality traits are firmly embedded within the breeding goal of many pig breeding programs. It is not easy to determine the genetic foundation of these traits because of complex interactions between muscle fibres, peri- and postmortem energy metabolism and different environmental factors that affect postmortem transformation from muscle to meat (Karlsson et al. 1999). One major gene affecting meat quality is *RYR1* (Fujii et al. 1991). There are multiple isoforms of ryanodine receptors, whereby *RYR1* is primarily expressed in skeletal muscle where it codes for the ryanodine receptor. It is localized to the sarcoplasmatic reticulum and regulates Ca^{2+} release into the cytoplasm. A transition of C to T on position 1843 in the *RYR1* gene leads to a substitution of Arginine with Cysteine at amino acid 615. This is one of the reasons for the disturbed regulation of intracellular Ca^{2+} in pig skeletal muscles, resulting in malignant hyperthermia and reduced meat quality. Several research groups have mapped further QTL for meat quality traits in pigs, and numerous QTL have been identified (see Hu et al. 2005). Several authors suggested the inclusion of epistasis in QTL mapping (Carlborg and Haley, 2004) and Große-Brinkhaus et al. (2010) and Duthie *et al.* (2011) found epistatic effects between QTL affecting the meat quality. Additionally, imprinting seems to be important for some QTL affecting the meat quality (de Koning et al.

2001). Recently, Wolf and Cheverud (2009) developed a framework for QTL mapping combining both epistatic and imprinting effects.

Geldermann et al. (2003) established a large and powerful porcine experimental design consisting of three connected F_2 crosses derived from the three genetically different founder breeds: Chinese Meishan, European Wild Boar and Piétrain. Bennewitz and Meuwissen (2010) showed that numerous QTL for a variety of traits segregate in this design. Rückert et al. (2011) used the design to map QTL for metabolic and cytological fatness traits. The aim of this study was to use this design to map QTL for meat quality traits using a two-step procedure. In the first step, QTL were mapped based on their additive, dominant and imprinting effects with a multi-allele multi-QTL approach tailored to analyse connected F_2 crosses, which was previously introduced from plant breeding by Rückert and Bennewitz (2010). The mapped QTL were tested for pairwise epistatic effects in the second step. All possible epistatic effects between additive, dominant, and imprinting effects were considered, leading to nine orthogonal forms of epistasis for each cross.

Materials and Methods

Animals and traits

Three porcine F_2 crosses were generated from the three founder breeds Meishan, Piétrain and Wild Boar, resulting in Wild Boar x Piétrain (WxP), Wild Boar x Meishan (WxM) and Meishan x Piétrain (MxP) crosses. For details of the experimental design see Geldermann et al. (2003). All 966 F_2 individuals were phenotyped for 46 traits including growth, fattening, fat deposition, muscling, meat quality, stress resistance and body conformation.

Seven meat quality traits were investigated in this study: con45, con24 (conductivity in mS/cm 45 min and 24h post-mortem respectively), meat colour (meat colour, measured 24h post-mortem), pH45 and pH24 (pH-values measured 45min and 24h post-mortem, respectively), rigor (stiffness in units, measured 45min post-mortem) and drip loss [cooling loss in %, describes the difference between warm carcass weight (1 h after slaughter) and cold carcass weight (24 h after slaughter) as percentage of warm carcass weight]. For all meat quality traits except meat colour two repeated measurements per trait were made in the carcass. For meat colour, five repeated measurements at different points were made. All traits except rigor were measured in the muscle longissimus dorsi. Rigor was measured in the M. semimembranosus. The traits con45, con24, meat colour, pH45 and pH24 were measured between the 13th and 14th rib.

The phenotypes were pre-corrected for the effects of sex, litter, season and age at slaughter. Descriptive statistics are given in Table 1. The animals were genotyped for around 250 genetic markers (mostly microsatellites) located across the whole genome.

Statistical analysis

Pedigree and marker data were used to calculate a linkage map across the three crosses (see Rückert and Bennewitz 2010 for details and the genetic map). Genotype probabilities $pr(Q_i^p Q_i^m)$, $pr(Q_j^p Q_i^m)$, $pr(Q_i^p Q_j^m)$ and $pr(Q_j^p Q_j^m)$ were calculated with a modified version of BIGMAP (Reinsch 1999) for each chromosomal position and each F_2 individual of a certain cross, assuming that the founder breeds were homozygous at the QTL. The upper subscript denotes the parental origin, paternal (*p*) or maternal (*m*) derived, and the lower subscript for the breed origin of the alleles (*i* or *j*, with *i*, *j* being Meishan, Piétrain or Wild Boar). These genotype probabilities were used to calculate the probability of an F_2 individual *k* of a certain cross receiving a QTL allele from founder breed *i* from its father as $z_{i,k}^p = pr(Q_i^p Q_i^m) + pr(Q_i^p Q_j^m)$, with *i*, *j* being Meishan, Piétrain or Wild Boar. Similarly, the probability of receiving the founder breed allele from *i* from its mother was calculated as $z_{i,k}^m = pr(Q_i^p Q_i^m) + pr(Q_j^p Q_i^m)$. The same was done for the founder breed allele from breed *j*. The probability of an F_2 individual *k* being heterozygous was calculated as the sum of the two heterozygous genotype probabilities, i.e. $z_{ij,k} = pr(Q_i^p Q_j^m) + pr(Q_j^p Q_i^m)$. A regression model could have been established by regression of pre-corrected phenotypes on these probabilities. However, because the sum of the additive effects within each parental origin is equal to zero such a model would be overparameterized. Therefore, the parameterisation of the model was done as $\tilde{z}_{M,k}^p = z_{M,k}^p - z_{W,k}^p$, $\tilde{z}_{P,k}^p = z_{P,k}^p - z_{W,k}^p$, $\tilde{z}_{M,k}^m = z_{M,k}^m - z_{W,k}^m$, and $\tilde{z}_{P,k}^m = z_{P,k}^m - z_{W,k}^m$ (see Rückert and Bennewitz 2010). The regression model was

$$y_{ijk} = cross_{ij} + RYR_{ij} + a_M^p \tilde{z}_{M,k}^p + a_M^m \tilde{z}_{M,k}^m + a_P^p \tilde{z}_{P,k}^p + a_P^m \tilde{z}_{P,k}^m + d_{MP} z_{MP,k} + d_{WP} z_{WP,k} + d_{WM} z_{WM,k} + e_{ijk}. \quad (1)$$

where the terms $cross_{ij}$ and RYR_{ij} denote the fixed effects of the F_2-cross and the fixed effect of the genotypes at *RYR1:g.1843C>T*. The *RYR1* gene was only included in model (1) if significant ($P<0.05$). Significance was tested using the following model:

$$y_{ijk} = cross_{ij} + RYR_{ij} + e_{ijk} \quad (2)$$

Model (1) produced estimates of the dominance effects ($\hat{d}_{MP}$, $\hat{d}_{WP}$, $\hat{d}_{WM}$) as well as the additive breed effects of breeds Meishan and Piétrain considering the parental origin of the

alleles ($\hat{a}_M^p, \hat{a}_M^m, \hat{a}_P^p, \hat{a}_P^m$). The additive effects of the Wild Boar breeds were estimated as $\hat{a}_W^p = -(\hat{a}_M^p + \hat{a}_P^p)$ and $\hat{a}_W^m = -(\hat{a}_M^m + \hat{a}_P^m)$. The error variance was assumed to be heterogeneous in both models, $e_{ijk} \sim N(0, \sigma_{ij}^2)$; for details, see Rückert and Bennewitz (2010). Combined additive Mendelian effects (i.e. ignoring parental origin of the alleles) were calculated as $\hat{a}_M = \hat{a}_M^p + \hat{a}_M^m$, $\hat{a}_P = \hat{a}_P^p + \hat{a}_P^m$, and $\hat{a}_W = -(\hat{a}_M^p + \hat{a}_M^m + \hat{a}_P^p + \hat{a}_P^m)$. Model (1) was fitted every cM on the autosomes by adapting the *z* terms accordingly. The test statistic was an *F*-test. The *F*-values were converted into LOD-scores as $LOD \approx (np * F)/(2 * \log(10))$, with *np* being the number of estimated QTL effects, i.e. *np* = 7 (four additive and three dominance effects). The global null hypothesis was that every estimated parameter is equal to zero at the chromosomal position with the highest test statistic. The alternative hypothesis was that at least one of the seven parameters is unequal to zero. The 1% and 5% threshold of the test statistic corrected for multiple testing on the chromosome was obtained using the quick method of Piepho (2001). Once the global null hypothesis was rejected, the sub-hypotheses for additive, dominance and imprinting effects were tested at significant chromosomal positions by building linear contrasts. The test of the three sub-hypotheses resulted in the three error probabilities p_{add}, p_{dom}, and p_{imp} for additive, dominance and imprinting QTL, respectively. Additionally, the number of QTL alleles which could be distinguished based on their additive mendelian effects was assessed. This was done by testing QTL segregation in each of the three crosses, considering only additive mendelian effects and ignoring dominance and imprinting. For each significant QTL, a confidence interval was calculated using the one LOD drop method. Multiple QTL were included as cofactors in the model using a forward selection approach. For further details of this model see Rückert and Bennewitz (2010).

Two-locus model for pairwise epistatic interactions

Model (1) cannot be easily expanded to test for epistatis (Bennewitz and Rückert, unpublished results), so the model of Wolf and Cheverud (2009) was used to test for comparison-wise epistatic effects. Therefore, we followed the general F_2 model, in which it is assumed that gene frequencies in a single F_2 population are one half. Further assumptions are Hardy–Weinberg equilibrium and linkage equilibrium between multiple QTL. The starting point is the orthogonal F_2 model derived from two founder breeds *i* and *j* for a locus A,

$$\begin{bmatrix} g_{ii}^{pm} \\ g_{ij}^{pm} \\ g_{ji}^{pm} \\ g_{jj}^{pm} \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & \frac{1}{2} & -1 \\ -1 & -\frac{1}{2} & 0 \end{bmatrix} \times \begin{bmatrix} a \\ d \\ i \end{bmatrix} + \begin{bmatrix} \mu \\ \mu \\ \mu \\ \mu \end{bmatrix} \quad . \qquad (3)$$

The *g* terms denote the genetic value; the upper subscript denotes parental origin and the lower subscript denotes the breed origin. The terms *a*, *d* and *i* denote the additive, dominant and imprinting effects, respectively, and μ denotes the mean of the F_2 cross. The model can be written as $G_A = S_A \times E_A + \mu_A$, where the lower subscript denotes the locus A. S_A is the genetic effect design matrix, which is coded so that elements of each column sum to zero. See also Wolf and Cheverud (2009).

This model can be expanded to include all three F_2 crosses. G_A is a 12x1 vector and S_A is a 12x9 block diagonal matrix with elements in the blocks equal to (3) and zero otherwise. E_A is a 9x1 vector of and contains the *a*, *d* and *i* effects for each cross. μ_A is a 12x3 block matrix with the cross mean in the blocks and zero otherwise. This model differs from model (1), because it treats the F_2 crosses separately and thus produces 9 QTL effects instead of 7 as in model (1). Therefore, when mapping QTL without epistasis, this model would be of reduced power compared to model (1), see also Rückert and Bennewitz (2010). Model (3) can be formulated for a second locus B as $G_B = S_B \times E_B + \mu_B$. If locus A and B interact, the model can be expanded to include both loci and the interaction between them. The two-locus genotypic effect vector, $G_{A,B}$, is of dimension $3x4^2=48$. The square is due to the possible combinations of the four genotypes at locus A and the four at locus B. The combinations are present in all three crosses. The two locus effect vector can be written as $E_{A,B} = \left[E_A \quad E_B \quad E_A \otimes E_B\right]'$, with the dimensions 45x1. Eighteen elements denote the *a*, *d* and *i* effects for two loci and three crosses (3x2x3=18) and the remaining 27 elements denote interaction effects. The Kronecker product is used here to indicate the interaction effects between the additive effects of A with those of B ($aa_{A,B}$ effects), the additive effects of A with dominance effects of B ($ad_{A,B}$), the additive effects of A with imprinting effects of B ($ai_{A,B}$), the dominance effects of A with additive effects of B ($da_{A,B}$), the dominance effects of A with those of B ($dd_{A,B}$), the dominance effects of A with imprinting effects of B ($di_{A,B}$), the imprinting effects of A with additive effects of B ($ia_{A,B}$), the imprinting effects of A with dominance effects of B ($id_{A,B}$) and the imprinting effects of A with those of B ($ii_{A,B}$). The genetic effect design matrix $S_{A,B}$ (dimensions 48 x 45) is structured as

$$S_{A,B} = S_A \otimes S_B .$$

Hence, the two-locus genetic model accounting for epistasis and imprinting becomes

$$G_{A,B} = S_{A,B} \times E_{A,B} + \mu .$$

This model was used to build the following regression model (in matrix notation)

$$\begin{aligned} y = {} & Xb + Z_A a_A + Z_B a_B + W_A d_A + W_B d_B + Q_A i_A + Q_B i_B \\ & + Z_A \otimes Z_B aa_{A,B} + Z_A \otimes W_B ad_{A,B} + Z_A \otimes Q_B ai_{A,B} \\ & + W_A \otimes Z_B da_{A,B} + W_A \otimes W_B dd_{A,B} + W_A \otimes Q_B di_{A,B} \\ & + Q_A \otimes Z_B ia_{A,B} + Q_A \otimes W_B id_{A,B} + Q_A \otimes Q_B ii_{A,B} + e \end{aligned} \quad (4)$$

with *y* being a vector with pre-corrected phenotypic observations in the F_2-populations, *b* is a vector with fixed cross effects and RYR1:g.1843C>T effects (if significant in model (2)). Vector *e* contains random residuals with cross-specific variances. Vector a_A (a_B) contains additive effects of locus A (B), d_A (d_B) is a vector with dominant effects of locus A (B), i_A (i_B) is a vector with imprinting effects of locus A (B), $aa_{A,B}$ is a vector with additive by additive interaction effects, $ad_{A,B}$ ($da_{A,B}$) is a vector with additive by dominant (dominant by additive) interaction effects, $ai_{A,B}$ ($ia_{A,B}$) is a vector with additive by imprinting (imprinting by additive), $dd_{A,B}$ is a vector with dominant by dominant interaction effects, $di_{A,B}$ ($id_{A,B}$) is a vector with dominant by imprinting (imprinting by dominant) and $ii_{A,B}$ is a vector with imprinting by imprinting interaction effects. The dimension of vectors containing QTL effects is always 3x1 (one effect for each cross). X_A, Z_A, W_A, Q_A, X_B, Z_B, W_B and Q_B are known design matrices linking the observations to the fixed effects and to the additive, dominance and imprinting QTL effects of locus A and B, respectively. The elements in *Z*, *W* and *Q* were calculated using the four genotype probabilities of each F_2-individual and using parameterisation as shown in (3).

$$Z_A = \begin{bmatrix} z_{A,MxP} & 0 & 0 \\ 0 & z_{A,WxP} & 0 \\ 0 & 0 & z_{A,WxM} \end{bmatrix}, W_A = \begin{bmatrix} w_{A,MxP} & 0 & 0 \\ 0 & w_{A,WxP} & 0 \\ 0 & 0 & w_{A,WxM} \end{bmatrix}, Q_A = \begin{bmatrix} q_{A,MxP} & 0 & 0 \\ 0 & q_{A,WxP} & 0 \\ 0 & 0 & q_{A,WxM} \end{bmatrix}$$

The *z*, *w* and *q* coefficients are regression variables calculated for each individual and each chromosomal position using the QTL genotype probabilities as

$$z_{A,MxP} = pr(Q_M^p Q_M^m) - pr(Q_P^p Q_P^m), \qquad q_{A,MxP} = pr(Q_M^p Q_P^m) - pr(Q_P^p Q_M^m),$$
$$z_{A,WxP} = pr(Q_W^p Q_W^m) - pr(Q_P^p Q_P^m), \qquad q_{A,WxP} = pr(Q_W^p Q_P^m) - pr(Q_P^p Q_W^m),$$
$$z_{A,WxM} = pr(Q_W^p Q_W^m) - pr(Q_M^p Q_M^m), \qquad q_{A,WxM} = pr(Q_W^p Q_M^m) - pr(Q_M^p Q_W^m),$$
$$w_{A,MxP} = -\frac{1}{2} pr(Q_M^p Q_M^m) + \frac{1}{2} pr(Q_M^p Q_P^m) + \frac{1}{2} pr(Q_P^p Q_M^m) - \frac{1}{2} pr(Q_P^p Q_P^m),$$
$$w_{A,WxP} = -\frac{1}{2} pr(Q_W^p Q_W^m) + \frac{1}{2} pr(Q_W^p Q_P^m) + \frac{1}{2} pr(Q_P^p Q_W^m) - \frac{1}{2} pr(Q_P^p Q_P^m),$$
$$w_{A,WxM} = -\frac{1}{2} pr(Q_W^p Q_W^m) + \frac{1}{2} pr(Q_W^p Q_M^m) + \frac{1}{2} pr(Q_M^p Q_W^m) - \frac{1}{2} pr(Q_P^p Q_P^m).$$

For locus B they were calculated analog to locus A (not shown).

Hypothesis testing

The null hypothesis was that all regression coefficients in the 3x1 vectors displaying epistatic effects ($aa_{A,B}$, $ad_{A,B}$, $da_{A,B}$, $di_{A,B}$, $id_{A,B}$, $ai_{A,B}$, $ia_{A,B}$, $dd_{A,B}$, $ii_{A,B}$) were equal to zero. The alternative hypothesis was that at least one of these 3*9=27 interaction effects was unequal to zero. The test statistic was an *F*-test. In total, 18 tests for pairwise interactions between QTL were conducted, resulting in comparison-wise error probabilities *p*. An experiment-wise error (p_e) was calculated as $p_e = 1-(1-p)^{18}$ using the Bonferroni corrections. To avoid a too stringent significance criterion the false discovery rate (FDR) was additionally applied using the QVALUE software (Storey 2002).

Results

Single QTL analysis

The summary statistics in Table 1 reveal a substantial variation for all traits within and across the three crosses. For conductivity traits the mean of the WxM cross is generally substantially lower than that of the other two crosses. In contrast, mean pH-values and mean of meat colour are highest in the WxM cross compared to the other two crosses. The drip loss mean of the WxP cross is substantially higher than those of the other crosses. For rigor the mean of the WxM cross is lower than in the other crosses. This is in agreement with the breeding history of the Piétrain and the Meishan breeds. The Piétrain breed is a typical sire line and was selected for leanness and growth during the last decades, resulting in reduced meat quality in general. The Meishan breed is known to be a fatty breed. The Wild Boar breed was not subject to artificial selection and hence there was little or no selection pressure on meat quality traits.

As shown in Table S1, all F_2 individuals in the cross WxM are homogeneous for the *RYR1:G.1843C>T* genotypes; individuals in the crosses WxP and MxP show all three genotypes. The *RYR1* gene polymorphism C1843T was highly significant for all traits except drip loss (Table S2). QTL results from model (1) for all meat quality traits are shown in Table 2. The chromosomal position and the upper and lower bounds of the confidence intervals together with the markers are presented for each QTL. For all meat quality traits with a chromosome-wise error probability below 1% (below 5%), 7 (19) QTL were found. Based on the combined additive mendelian effect (i.e. ignoring dominance and impriting), 6 QTL revealed three alleles, 10 QTL revealed 2 alleles and 3 QTL revealed 1 allele. All in all 15 additive (p_{add}<0.05), 12 significant dominant (p_{dom}<0.05) and 5 significant imprinting effects (p_{imp}<0.05) were found for QTL with a 5% chromosome-wise significance level. On SSC6 QTL for meat colour, con24, con45 and ph45 were mapped in the region of the *RYR1:G.1843C>T* gene, although the *RYR1:G.1843C>T* mutation was included in model (1) for all traits except drip loss. For con45 two QTL were mapped on SSC6 (at 100cM and 105cM), see Table 2. Figure 1 shows LOD-score profiles with test statistics plotted against chromosomal position on SSC6 for all traits. On the top, QTL significant for a chromosome-wise error probability above the 1% threshold are depicted, whereas the lower LOD-score profiles depict traits with no significant QTL on SSC6. The plot for the significant QTL was remarkably sharp, leading to small confidence intervals for the QTL. Additionally, two distinct peaks around 100cM and 105cM can be distinguished for the trait con45, supporting the presence of two QTL in this chromosomal region.

Results from the two-locus analysis

The results from the two-locus model using pairwise epistatic effects are shown in Table 3. The interaction between SSC14 and SSC12 for pH24 showed the highest test statistic, followed by the interaction between SSC2 and SSC5 for rigor. Five pairwise epistatic interactions had a false discovery rate below 25% (Table 3); these interactions showed the highest test statistics (*F*-values) and the lowest error probabilities (comparison-wise *p* values). Model (4) was used to calculate additive, dominant and imprinting QTL effects (Table 4) and pairwise epistatic effects (Table 5) for these interactions. Additionally, the heterogeneity in error variance estimates between the crosses is demonstrated in Table 4. A total of 15 single interaction effect estimates were found to be significant (p<0.05), but none of the $aa_{A,B}$ were significant.

Discussion

Two step procedure

In the present study a two-step procedure was used to map and characterize QTL for meat quality traits in a large F_2 experiment. In the first step, QTL were mapped considering additive, dominant and imprinting effects and by applying the multi-QTL multi-allele model. Rückert and Bennewitz (2010) showed the high statistical power of this approach compared to separate analysis of the crosses. The QTL mapped in the first step were tested for pairwise epistatic effects in the second step. Because the model applied in the first step cannot be easily expanded to include multiple interacting loci, the orthogonal model of Wolf and Cheverud (2009) was applied. In principle, this model could have been used to conduct a full two-locus epistatic genome scan, however this would result in a massive multiple testing problem. Even if appropriate correction for multiple testing would have been conducted, the result would be of low experimental power for detection of epistatic effects. Additionally, it is assumed that QTL with real epistatic effects should also show some real single QTL effects (see also Wolf and Cheverud, 2009). Both genes could, for example, be involved in one or two compensatory pathways or metabolisms which determine muscle phenotype. The test for epistasis was done by treating the crosses separately, but applying one linear regression model which included all three crosses (model (4)). This offered the possibility to test for epistasis in all three crosses simultaneously. Alternatively, the model could have been applied separately for each cross, but this would increase the number of tests by the factor three.

General QTL effects

In contrast to Meishans, Piétrain pigs have a dominance of white muscle fibres, connected with a high rate of post-mortem glycogenolysis and an especially high disposition for pale, soft and exudative meat. Therefore a lower pH and a higher conductivity value in the cutlet muscle of Piétrains was expected. This is also supported by the low variation of con45 in the WxM cross (Table 1) and by the low error variance of model (4) shown in Table 4 for this cross. For both pH45 QTL, the Meishan breed allelic effect was higher than the Piétrain breed allelic effect (Table 2). Unexpectedly, for the three pH24 QTL which revealed additive QTL effects on SSC9, 12 and 14, the Meishan breed allelic effect was the same or even lower than the Piétrain breed allelic effect. Due to a negative correlation between pH and conductivity, it was expected that the allelic effect of the QTL in Piétrain is higher compared to the Meishan breed allelic effect. This order of estimated effects ($\hat{a}_P > \hat{a}_M$) holds true for two con24 QTL located on SSC4 and 6 and for one con45 QTL located on SSC6. Meat color is especially

dependent on the amount of myoglobin, which is mainly located in type I (slow-twitch) fibers. These represent the main type of fibers in Wild Boar. Hence, the Wild Boar breed allelic effect for one meat colour QTL located on SSC6 is higher than the other allelic effects. No clear pattern of breed allelic effects was observable for pH24 or the remaining QTL not discussed above.

QTL results

A comparison of the QTL results and other literature results can be done using the pig QTL data base (Hu et al., 2005). In the following, only the most interesting QTL results will be discussed. The *RYR1:G.1843C>T* is a well known major gene affecting meat quality traits (Fujii et al. 1991). The T allele was mainly observed in Piétrain breeds, whereas other breeds are usually homozygous CC. In our study, the genotypes CT and TT were only observed in the WxP and MxP, but not in the WxM cross (Table S1), which indicates that the T allele originates from the Piétrain breed. Interestingly, when *RYR1:G.1843C>T* was included as a fixed effect in model (1), the QTL for pH45, meat colour, con24 and con45 next to the position of the *RYR1* gene at around 100cM remained significant (see Figure 1). This indicates the presence of a further polymorphism with an effect on meat quality within the gene, a second QTL in the same chromosomal region, or both. In the same region, Malek et al. (2001), Markljung et al. (2008) and Cherel et al. (2011) found evidence for meat quality QTL which was not attributable to the *RYR1:G.1843C>T* mutation. This is in agreement with our results. In contrast, Mohrmann et al. (2006) found no evidence for additional QTL closely linked to the *RYR1* polymorphism for stress syndrome traits such as conductivity and pH. Due to the moderate to strong correlations between meat quality traits (De Vries et al., 1994; Borchers et al., 2007), an overlap of confidence intervals across the traits was expected and observed, e.g. for drip loss and rigor on SSC2, for pH45 and con24 on SSC4, for rigor and con24 on SSC5 and for pH45, meat colour, con24 and con45 on SSC6 (Table 4 and Figure 1). Numerous QTL have been found for meat quality traits on SSC2 (Li et al., 2010 ; Qiu et al., 2010; Thomsen et al., 2004; van Wijk et al., 2006) as well as on SSC5 by Srikanchai et al. (2009) and Wimmers et al. (2007). Additionally Wimmers et al. (2007) found QTL on SSC4 for pH45.

Imprinting and pairwise epistatic effects

Few QTL showed significant imprinting effects (Table 2). The imprinted QTL for rigor located in the marker interval [SWR453; SW2] (Table 2) on SSC5 was also found by Thomsen et al. (2004). Duthie et al. (2011) found an imprinting effect for pH45 positioned close to the imprinting QTL on SSC7 for meat colour (Table 2). SSC7 probably contains the orthologue ovine chromosomal region encompassing the callipyge locus (Boysen et al., 2010), which is known to show imprinting effects in sheep. This region is located in the telomere on the *q*-arm of SSC7 and therefore far away from the imprinting QTL confidence interval for meat colour on SSC7 in this study (Table 2). De Koning et al. (2001) detected one imprinted QTL for drip loss on SSC6, whereas in this study imprinted QTL for con45 were found (Table 2).

The application of model (4) enabled us to study epistatic effects involving imprinting, which was proposed by Wolf and Cheverud (2009). The interpretation of the corresponding pairwise epistatic effects shown in Table 5 is complex. However, the results underline the importance of considering imprinting in epistatic analysis. For the two pH24 QTL located on SSC14 and SSC12, the interaction between the imprinting (SSC12) and additive effect (SSC14) was significant in one cross whereas the interaction between both imprinting effects was significant in two crosses. Interestingly, imprinting was not found in the QTL analysis using model (1) for these QTL (Table 2). The importance of imprinting interactions was also observed for the second, pairwise significant epistasis between the two QTL for rigor on SSC2 and SSC5. Imprinting at these QTL is controlled by the second QTL and vice versa. This kind of control of imprinting of a QTL by a second QTL was also found in mice by Wolf and Cheverud (2009), who provided some arguments from an evolution perspective.

Some studies considering epistasis were conducted for muscle-, fat- and meat quality traits. Duthie et al. (2011) found one interaction between QTL located on SSC4 and SSC6 for meat colour, which is in agreement with the interaction found for pH45 (Table 3). Interactions between SSC6 and SSC14 for con45 (Table 3) were also found for microstructural muscle properties (Estelle et al., 2008) and for pH decline (Große-Brinkhaus et al., 2010). Because the power to detect epistatic effects is substantially lower than that of detecting additive effects (Mao and Da, 2005) and the significance of the pairwise epistatic effects is weak for most traits with an FDR *q*-value below 0.25, these results have to be confirmed. Putative interaction effects between the *RYR1:G.1843C>T* genotype and other QTL remain to be investigated.

Conclusions

In the present study three connected F_2 crosses with almost 1000 individuals were analyzed to map QTL for meat quality traits. Several QTL that partly showed pairwise epistatic effects were identified. Epistasic effects were also found for imprinting which need confirmation and further investigation.

Acknowledgements

PS received funding from the Carl-Zeiss-Stiftung, Stuttgart, Germany and CR from the H. Wilhelm Schaumann Stiftung, Hamburg, Germany.

References

Bennewitz, J., and T. H. E. Meuwissen (2010). The distribution of QTL additive and dominance effects in porcine F2 crosses. *Journal of Animal Breeding and Genetics* **127**: 171–179.

Borchers, N., G. Otto, E. Kalm (2007). Genetic relationship of drip loss to further meat quality traits in purebred Piétrains. *Archiv für Tierzucht, Dummerstorf* **50**: 1, 84-91.

Boysen, T. J., J. Tetens, G. Thaller (2010). Detection of a quantitative trait locus for ham weight with polar overdominance near the ortholog of the callipyge locus in an experimental pig F_2 population. *Journal of Animal Science* **88**: 3167-3172.

Carlborg, O., and C. S. Haley (2004). Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics*. **5**: 618–625.

Cherel, P., J. Pires, J. Glénisson, D. Milan, N. Iannuccelli, F. Hérault, M. Damon, P. Le Roy (2011). Joint analysis of quantitative trait loci and major-effect causative mutations affecting meat quality and carcass composition traits in pigs. *BMC Genetics* **12**: 76.

De Koning, D. J., B. Harlizius, A. P. Rattink, M. A. Groenen, E. W. Brascamp, J. A. Van Arendonk (2001). Detection and characterization of quantitative trait loci for meat quality traits in pigs. *Journal of Animal Science* **79***:* 2812–*1*9.

De Vries, A. G., P. G. van der Wal, T. Long, G. Eikelenboom, J. W. M. Merks (1994). Genetic parameters of pork quality and production traits in Yorkshire populations. *Livestock Production Science* **40**: 277–289.

Duthie, C. A., G. Simm, A. Doeschl-Wilson, E. Kalm, P. W. Knap, R. Roehe (2011). Quantitative trait loci for meat quality traits in pigs considering imprinting and epistatic effects. *Meat Science* **87**: 394-402.

Estelle, J., F. Gil, J. M. Vazquez, R. Latorre, G. Ramirez, M. C. Barragan, J. M. Folch, J. L. Noguera, M. A. Toro, M. Perez-Enciso (2008). A quantitative trait locus genome scan for porcine muscle fiber traits reveals overdominance and epistasis. *Journal of Animal Science* **86**: 3290–3299.

Fischer, K. (2007). Drip loss in pork: influencing factors and relation to further meat quality traits. *Journal of Animal Breeding and Genetics* **124**: 12–18.

Fujii, J., K. Otsu, F. Zorzato, S. DeLeon, V. K. Khanna, J. E. Weiler, P. J. O'Brien, D. H. McLennan (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253**: 448-451.

Geldermann, H., E. Müller, G. Moser, G. Reiner, H. Bartenschlager, S. Cepica, A. Stratil, J. Kuryl, C. Moran, R. Danoli, C. Brunsch (2003). Genome-wide linkage and QTL

mapping in porcine F2 families generated from Piétrain, Meishan and Wild Boar crosses. *Journal of Animal Breeding and Genetics* **120**: 363-393.

Große-Brinkhaus, C., E. Jonas, H. Buschbell, C. Phatsara, D. Tesfaye, H. Jüngst, C. Looft, K. Schellander, E. Tholen (2010). Epistatic QTL pairs associated with meat quality and carcass composition traits in a porcine Duroc x Piétrain population. *Genetics Selection Evolution* **42**: 39.

Hu, Z. L., S. Dracheva, W. Jang, D. Maglott, J. Bastiaansen, M. F. Rothschild, J. M. Reecy (2005). A QTL resource and comparison tool for pigs: PigQTLDB. *Mammalian Genome* **16**: 792-800.

Huff-Lonergan, E., and S. M. Lonergan (2007). New frontiers in understanding drip loss in pork: recent insights on the role of postmortem muscle biochemistry. *Journal of Animal Breeding and Genetics* **124**: 19–26.

Karlsson, A. H., R. E. Klont, X. Fernandez (1999). Skeletal muscle fibres as factors for pork quality. *Livestock Production Science* **60**: 255–269.

Li, H. D., O. F. Christensen, V. R. Gregersen, P. Henckel, C. Bendixen (2010). Quantitative trait loci analysis of swine meat quality traits. *Journal of Animal Science* **88**: 2904-2912.

Malek, M., J. C. M. Dekkers, H. K. Lee, T. J. Baas, K. Prusa, E. Huff-Lonergan, M. F. Rothschild (2001). A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. II. Meat and muscle composition. *Mammalian Genome* **12**: 637-645.

Mao, Y., and Y. Da (2005). Statistical power for detecting epistasis QTL effects under the F-2 design. *Genetics Selection Evolution* **37**: 129-150.

Markljung, E., M. Braunschweig, P. Karlskov-Mortensen, C. Bruun, M. Sawera, I. Cho, I. Hedebro-Velander, A. Josell, K. Lundström, G. Seth (2008). Genome-wide identification of quantitative trait loci in a cross between Hampshire and Landrace II: Meat quality traits. *BMC Genetics* **9**: 22.

Mohrmann, M., R. Roehe, P. W. Knap, H. Looft, G. S. Plastow, E. Kalm (2006). Quantitative trait loci associated with AutoFOM grading characteristics, carcass cuts and chemical body composition during growth of Sus scrofa. *Animal Genetics* **37**: 435–43.

Piepho, H. P. (2001). A quick method for computing approximate threshold for quantitative trait loci detection. *Genetics* **157**: 425–432.

Qiu, H., X. Xu, B. Fan, M. F. Rothschild, Y. Martin, B. Liu (2010). Investigation of LDHA and COPB1 as candidate genes for muscle development in the MYOD1 region of pig chromosome 2. *Molecular Biology Reports* **37**: 629–636.

Reinsch, N. (1999). A multiple-species, multiple-project database for genotypes at codominant loci. *Journal of Animal Breeding and Genetics* **116**: 425-435.

Rückert, C., and J. Bennewitz (2010). Joint QTL analysis of three connected F2-crosses in pigs. *Genetics Selection Evolution* **42**: 40.

Rückert, C., P. Stratz, S. Preuss, J. Bennewitz (2011). Mapping QTL for metabolic and cytological fatness traits of connected F2 crosses in pigs. *Journal of Animal Science* **90**: 399–409.

Srikanchai, T., E. Murani, K. Wimmers, S. Ponsuksili (2009). Four loci differentially expressed in muscle tissue depending on water-holding capacity are associated with meat quality in commercial pig herds. *Molecular Biology Reports* **37**: 595–601.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* **64**: 479-498.

Thomsen, H., H. K. Lee, M. F. Rothschild, M. Malek, J. C. M. Dekkers (2004). Characterization of quantitative trait loci for growth and meat quality in a cross between commercial breeds of swine. *Journal of Animal Science* **82***:* 2213–28.

van Wijk, H. J., B. Dibbits, E. E. Baron, A. D. Brings, B. Harlizius, M. A. M. Groenen, E. F. Knol, H. Bovenhuis (2006). Identification of quantitative trait loci for carcass composition and pork quality traits in a commercial finishing cross. *Journal of Animal Science* **84**: 789–799.

Wimmers, K., E. Murani, M. F. Te Pas, K. C. Chang, R. Davoli, J. W. Merks, H. Henne, M. Muraniova, N. da Costa, B. Harlizius, K. Schellander, I. Bll, S. Braglia, A. A. de Wit, M. Cagnazzo, L. Fontanesi, D. Prins, S. Ponsuksili (2007). Associations of functional candidate genes derived from gene-expression profiles of prenatal porcine muscle tissue with meat quality and muscle deposition. *Animal Genetics* **38**: 474–484.

Wolf, J. B., and J.M. Cheverud (2009). A framework for detecting and characterizing genetic background-dependent imprinting effects. *Mammalian Genome* **20**: 681–698.

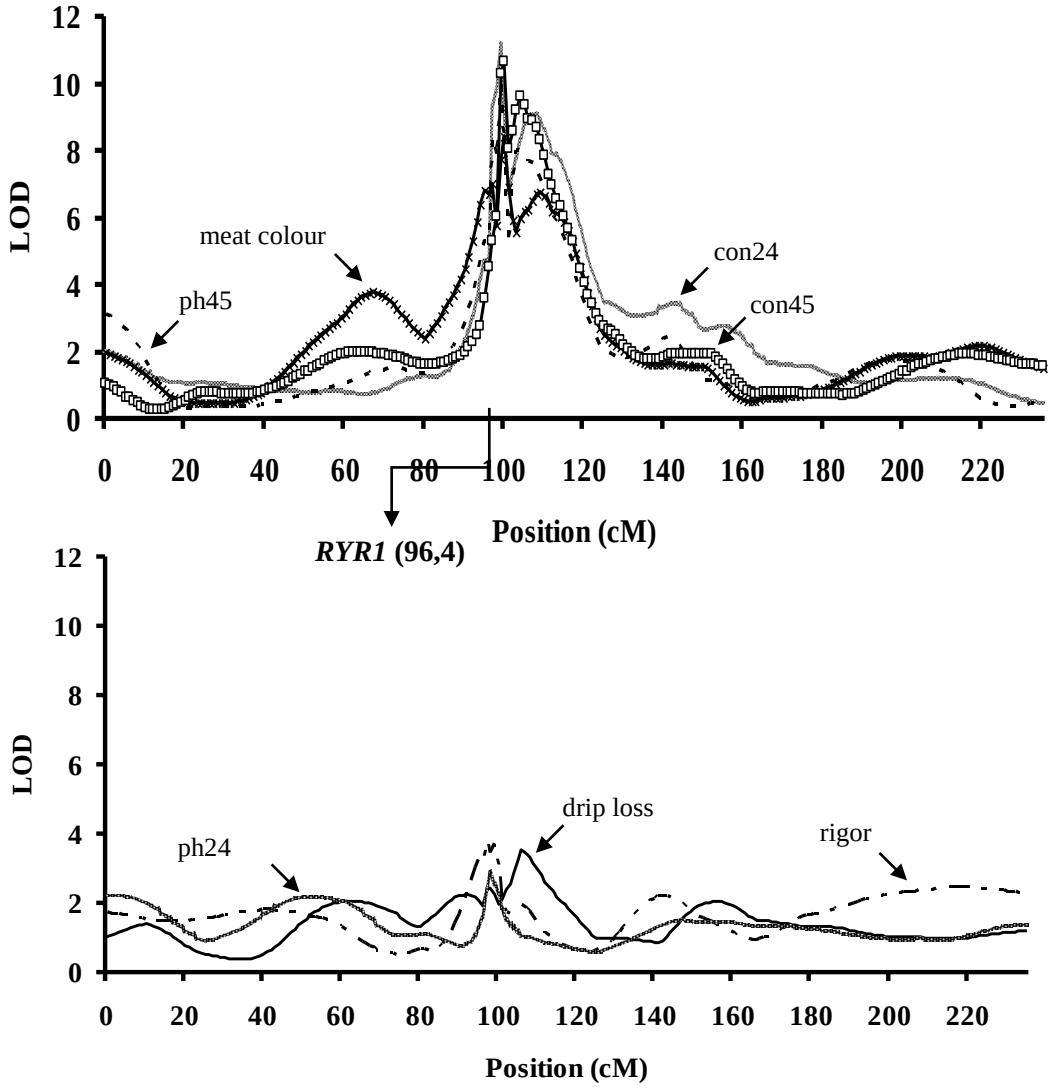


Figure 1 LOD-score profiles on chromosome 6 for con24, con45, pH45 and meat colour (top) and for rigor, drip loss and pH24 (bottom) calculated using model (1) where *RYR1:G.1843C>T* was included for all traits except drip loss. The arrow on the x-coordinate indicates the position of the *RYR1*-locus.

Table 1 Number of observation (N), mean, standard deviation (Sd), minimum (Min) and maximum (Max) of the phenotypic observations and coefficient of variation (CV)

Trait	Cross	N	Mean	Sd	Min	Max	CV
con45	MxP	316	7.50	5.36	2.52	22.78	71.42
	WxP	315	9.57	7.41	2.59	24.90	77.43
	WxM	335	3.59	0.53	2.34	6.30	14.66
	Joint	966	6.82	5.79	2.34	24.90	84.94
con24	MxP	316	7.36	2.78	2.26	13.42	37.81
	WxP	315	8.66	3.03	1.99	14.04	35.00
	WxM	335	3.53	0.94	1.89	8.10	26.69
	Joint	966	6.45	3.26	1.89	14.04	50.54
meat colour	MxP	316	63.87	8.10	45.6	83.30	12.68
	WxP	315	68.48	10.57	44.9	88.00	15.44
	WxM	335	69.59	5.87	48.1	82.50	8.43
	Joint	966	67.36	8.71	44.9	88.00	12.93
pH45	MxP	316	5.68	0.40	5.10	6.56	7.01
	WxP	315	5.59	0.39	5.14	6.54	6.92
	WxM	335	6.01	0.21	5.27	6.38	3.51
	Joint	966	5.77	0.39	5.10	6.56	6.71
pH24	MxP	316	5.48	0.09	5.34	6.22	1.65
	WxP	315	5.47	0.06	5.32	5.69	1.06
	WxM	335	5.54	0.09	5.35	6.00	1.66
	Joint	966	5.5	0.09	5.32	6.22	1.59
rigor	MxP	316	8.94	1.92	3.50	13.00	21.45
	WxP	315	8.80	1.9	5.00	14.00	21.59
	WxM	335	7.5	1.43	4.50	12.00	19.03
	Joint	966	8.39	1.87	3.50	14.00	22.32
drip loss	MxP	316	1.31	0.54	0.27	4.48	40.92
	WxP	315	2.09	1.03	0.58	7.32	49.31
	WxM	335	1.79	0.79	0.59	7.41	44.02
	Joint	966	1.73	0.87	0.27	7.41	50.18

Table 2 QTL results from the joint design using model (1)

Trait	SSC	Pos	CI[1]	*F*-value	p_{add}[2]	p_{dom}[3]	p_{imp}[4]	Order of effects[5]
drip loss	2	58	[S0141;MLP] [39.9;68.0]	3.57*	0.002	0.042	0.062	$\hat{a}_W > \hat{a}_P = \hat{a}_M$
	16	16	[S0111;SW419] [0.0;33.3]	3.08*	0.375	0.005	0.020	$\hat{a}_W = \hat{a}_M = \hat{a}_P$
pH45	4	86	[MEF2D;EAL] [82.5;93.7]	3.35*	0.005	0.022	0.168	$\hat{a}_M > \hat{a}_W = \hat{a}_P$
	6	99	[LIPE;TGFB1] [98.3;99.5]	6.08**	<0.001	<0.001	0.935	$\hat{a}_M > \hat{a}_W > \hat{a}_P$
rigor	2	64	[SW240;MYOD1] [52.9;70.6]	3.32*	0.070	0.001	0.944	$\hat{a}_W = \hat{a}_M = \hat{a}_P$
	5	53	[SWR453;SW2] [39.0;64.4]	3.38*	0.023	0.048	0.009	$\hat{a}_M > \hat{a}_P = \hat{a}_W$
meat colour	6	100	[LIPE; A1BG] [98.3;101.2]	5.30**	<0.001	0.074	0.732	$\hat{a}_W > \hat{a}_M > \hat{a}_P$
	7	47	[S0025;CYPA] [0.0;73.3]	3.25*	0.008	0.237	0.013	$\hat{a}_M = \hat{a}_P > \hat{a}_W$
	15	94	[SW2053;SW1983] [71.9;99.4]	3.32*	<0.001	0.523	0.276	$\hat{a}_P > \hat{a}_M > \hat{a}_W$
con24	4	86	[ATP1A2;EAL] [81.8;93.7]	3.94**	0.008	0.002	0.165	$\hat{a}_P > \hat{a}_M = \hat{a}_W$
	5	50	[SWR453;SW2] [39.0;64.4]	3.33*	0.028	0.010	0.057	$\hat{a}_M = \hat{a}_W > \hat{a}_P$
	6	99	[LIPE;TGFB1] [98.3;99.5]	6.78**	<0.001	0.012	0.491	$\hat{a}_P > \hat{a}_M > \hat{a}_W$
pH24	9	148	[SW2093;SW174] [135.6;158.1]	3.19*	0.031	0.003	0.816	$\hat{a}_M = \hat{a}_P > \hat{a}_W$
	12	2	[S0143;SW957] [0.0;32.0]	3.86*	0.009	<0.001	0.780	$\hat{a}_W = \hat{a}_P > \hat{a}_M$
	14	116	[SW210;SW2515] [84.3;151.2]	3.19*	0.021	0.024	0.065	$\hat{a}_W > \hat{a}_P > \hat{a}_M$

	17	79	[SJ063;GNAS] [69.9;86.4]	4.17*	0.429	<0.001	0.995	$\hat{a}_M = \hat{a}_P = \hat{a}_W$
con45	6	100	[LIPE;A1BG] [98.3;101.2]	5.80**	0.007	0.216	<0.001	$\hat{a}_P > \hat{a}_M > \hat{a}_W$
	6	105	[EAH;BNP1] [102.4;112.0]	5.35**	0.054	0.661	<0.001	$\hat{a}_P = \hat{a}_M > \hat{a}_W$
	14	60	[SW2083;ACTN2] [43.8;70.6]	4.07**	<0.001	0.059	0.489	$\hat{a}_W > \hat{a}_M = \hat{a}_P$

The model includes the effect of *RYR1* for all traits except drip loss.

[1] confidence interval (CI); [2] error probability for additive effects; [3] error probability for dominant effects; [4] error probability for imprinting effects; [5] $\hat{a}_P$ estimated effect of Piétrain breed, $\hat{a}_M$ estimated effect of Meishan breed, $\hat{a}_W$ estimated effect of Wild Boar breed; * above 5% chromosome-wise threshold values; ** above 1% chromosome-wise threshold values.

Table 3 Pairwise epistatic interaction results from the two-locus model (4) with test statistic (*F*-value), comparison-wise (*p*) and experiment-wise (p_e) error probability and FDR *q*-value.

Trait	QTL A		QTL B		*F*-value	*p*	p_e	*q*
	SSC	Pos	SSC	Pos				
pH24	**14**	**116**	**12**	**2**	**1.90**	**0.004**	**0.071**	**0.041**
rigor	**2**	**64**	**5**	**53**	**1.62**	**0.025**	**0.370**	**0.127**
pH45	**4**	**86**	**6**	**99**	**1.43**	**0.072**	**0.741**	**0.233**
con45	**14**	**60**	**6**	**100**	**1.38**	**0.094**	**0.829**	**0.233**
pH24	**9**	**148**	**14**	**116**	**1.34**	**0.116**	**0.892**	**0.233**
pH24	17	79	12	2	1.22	0.208	0.985	0.316
meat colour	7	47	6	100	1.20	0.222	0.989	0.316
con24	5	50	4	86	1.15	0.275	0.997	0.316
con45	14	60	6	105	1.12	0.311	0.999	0.316
meat colour	7	47	15	94	1.11	0.316	0.999	0.316
pH24	9	148	17	79	1.09	0.349	1.000	0.316
meat colour	15	94	6	100	1.06	0.379	1.000	0.316
con24	5	50	6	99	1.02	0.432	1.000	0.318
pH24	14	116	17	79	1.01	0.445	1.000	0.318
con24	4	86	6	99	0.95	0.544	1.000	0.363
drip loss	16	16	2	58	0.69	0.877	1.000	0.530
pH24	9	148	12	2	0.67	0.901	1.000	0.530

Pairwise epistatic interations with low FDR values (q<0.25) are written in bold face

Table 4 Main effect estimates of QTL involved in pairwise epistasis, results from model (4).

Trait	Cross	QTL A		QTL B		$\hat{a}_A$	$\hat{d}_A$	$\hat{i}_A$	$\hat{a}_B$	$\hat{d}_B$	$\hat{i}_B$	σ_e^2
		SSC	Pos	SSC	Pos							
pH24	MxP	14	116	12	2	-0.006	**0.034**	0.01	**-0.023**	**-0.054**	0.004	0.007
						(0.008)	(0.013)	(0.009)	(0.008)	(0.011)	(0.009)	
	WxP					0,008	0,012	**-0.014**	0.002	-0.002	-0.005	0.003
						(0,006)	(0,011)	(0.006)	(0.007)	(0.009)	(0.007)	
	WxM					**0,023**	0,003	-0.003	0.011	0.016	0.007	0.007
						(0,007)	(0,011)	(0.007)	(0.007)	(0.01)	(0.008)	
rigor	MxP	2	64	5	53	0.145	**-1.017**	-0.091	**0.423**	0.173	**-0.540**	3.206
						(0.148)	(0.260)	(0.192)	(0.173)	(0.253)	(0.157)	
	WxP					0,188	0,368	-0.142	0.101	0.387	0.097	3.039
						(0,170)	(0,380)	(0.261)	(0.160)	(0.232)	(0.166)	
	WxM					0,063	0,247	0.062	-0.189	-0.278	0.057	1.822
						(0,137)	(0,229)	(0.154)	(0.120)	(0.160)	(0.110)	
pH45	MxP	4	86	6	99	**0.055**	-0.014	0.021	**0.355**	0.067	0.053	0.092
						(0.026)	(0.038)	(0.027)	(0.026)	(0.038)	(0.028)	
	WxP					0,020	**0,101**	-0.033	**0.375**	**-0.087**	-0.030	0.067
						(0,021)	(0,032)	(0.024)	(0.021)	(0.031)	(0.024)	
	WxM					**-0,035**	0,000	-0.008	-0.017	-0.029	0.004	0.041
						(0,016)	(0,024)	(0.018)	(0.017)	(0.024)	(0.017)	

con45	MxP	14	60	6	100	-0.038 (0.337)	**-1.025** (0.475)	0.104 (0.331)	**-4.525** (0.361)	**-2.980** (0.505)	-0.412 (0.358)	15.820
	WxP					0.778 (0.401)	0.625 (0.588)	0.305 (0.438)	**-6.882** (0.404)	**-3.470** (0.597)	-0.221 (0.465)	24.926
	WxM					0.169 (0.039)	-0.049 (0.058)	0.041 (0.046)	0.011 (0.044)	0.078 (0.061)	0.004 (0.040)	0.240
pH24	MxP	9	148	14	116	0.003 (0.008)	0.004 (0.011)	0.004 (0.008)	-0.006 (0.008)	**0,030** (0,013)	0,004 (0,009)	0.008
	WxP					-0.008 (0.006)	-0.017 (0.010)	0.002 (0.005)	0.005 (0.006)	0.019 (0.011)	**-0.015** (0.006)	0.003
	WxM					-0.006 (0.011)	**-0.043** (0.020)	-0.001 (0.010)	**0.021** (0.007)	0.005 (0.012)	-0.006 (0.008)	0.007

Comparison-wise significant main effects ($p<0.05$) are written in bold face; standard errors are given in parenthesis.

Table 5 Epistatic effect estimates, results from model (4).

Trait	Cross	QTL A		QTL B		$\hat{aa}_{A,B}$	$\hat{ad}_{A,B}$	$\hat{ai}_{A,B}$	$\hat{da}_{A,B}$	$\hat{dd}_{A,B}$	$\hat{di}_{A,B}$	$\hat{ia}_{A,B}$	$\hat{id}_{A,B}$	$\hat{ii}_{A,B}$
		SSC	Pos	SSC	Pos									
pH24	MxP	14	116	12	2	.002	0.009	-0.019	-0.026	**-0.068**	-0.012	**-0.028**	-0.037	-0.003
						(.013)	(0.020)	(0.013)	(0.021)	(0.033)	(0.022)	(0.014)	(0.022)	(0.015)
	WxP					.026	**-0.033**	-0.021	-0.035	0.025	-0.006	-0.005	0.008	**0.036**
						(.014)	(0.016)	(0.011)	(0.027)	(0.030)	(0.023)	(0.015)	(0.017)	(0.014)
	WxM					-.015	0.012	-0.017	-0.016	0.042	-0.010	0.01	0.020	**-0.027**
						(.011)	(0.016)	(0.012)	(0.017)	(0.025)	(0.019)	(0.011)	(0.017)	(0.012)
rigor	MxP	2	64	5	53	-.471	0.054	-0.025	-0.209	-0.846	0.071	0.585	-0.586	0.004
						(.248)	(0.349)	(0.209)	(0.431)	(0.621)	(0.378)	(0.314)	(0.466)	(0.289)
	WxP					-.220	-0.058	-0.299	-0.161	1.372	-0.775	**-1.238**	-0.539	0.513
						(.245)	(0.351)	(0.247)	(0.551)	(0.774)	(0.539)	(0.378)	(0.527)	(0.365)
	WxM					-.028	-0.410	-0.028	<.001	-0.365	-0.036	**0.574**	0.096	0.279
						(.201)	(0.273)	(0.196)	(0.354)	(0.459)	(0.308)	(0.245)	(0.309)	(0.196)
pH45	MxP	4	86	6	99	.070	0.057	-0.077	-0.024	0.139	0.054	0.003	0.071	-0.043
						(.039)	(0.055)	(0.040)	(0.056)	(0.081)	(0.061)	(0.040)	(0.059)	(0.045)
	WxP					-.024	**0.092**	0.022	0.008	**0.175**	0.038	-0.034	0.033	-0.021
						(.030)	(0.045)	(0.035)	(0.045)	(0.068)	(0.054)	(0.032)	(0.049)	(0.039)
	WxM					.008	-0.054	0.015	-0.020	0.048	**0.069**	0.011	-0.021	-0.005
						(.023)	(0.034)	(0.024)	(0.036)	(0.051)	(0.035)	(0.028)	(0.039)	(0.025)

con45	MxP	14	60	6	100	-.112	**-1.551**	**1.229**	0.466	0.379	0.680	0.438	-0.357	**-1.458**
						(.528)	(0.747)	(0.536)	(0.746)	(1.043)	(0.740)	(0.515)	(0.717)	(0.508)
	WxP					-.141	1.312	-0.086	-0.333	-0.125	0.156	0.083	-0.017	-0.226
						(.559)	(0.836)	(0.674)	(0.836)	(1.232)	(0.959)	(0.638)	(0.923)	(0.692)
	WxM					-.016	0.073	-0.045	0.077	**-0.264**	-0.130	-0.110	-0.120	0.004
						(.059)	(0.085)	(0.058)	(0.090)	(0.125)	(0.082)	(0.074)	(0.099)	(0.064)
pH24	MxP	9	148	14	116	-.017	0.002	0.020	-0.006	0.014	-0.010	-0.007	-0.005	-0.027
						(.012)	(0.020)	(0.013)	(0.018)	(0.030)	(0.019)	(0.013)	(0.021)	(0.015)
	WxP					.014	0.014	0.002	-0.002	-0.031	**0.056**	-0.014	-0.018	0.014
						(.010)	(0.020)	(0.011)	(0.019)	(0.035)	(0.019)	(0.010)	(0.019)	(0.010)
	WxM					.022	0.001	0.017	-0.011	-0.064	-0.031	0.016	-0.037	-0.004
						(.017)	(0.028)	(0.019)	(0.031)	(0.047)	(0.031)	(0.016)	(0.024)	(0.015)

Comparison-wise significant epistatic effects ($p<0.05$) are written in bold face; standard errors are given in parenthesis.

Appendix

Table S1 Observed number of *RYR1:G.1843C>T* genotypes within the three F_2-crosses

Genotype	WxP	WxM	MxP
CC	78	335	67
CT	154	-	174
TT	83	-	75

Table S2 *RYR1:G.1843C>T* genotype estimates using model (2) for all meat quality traits

Trait	*F*-value	*p*	*RYR1* genotype	Estimate	Standard error
drip loss	0.40	0.671	-	-	-
			-	-	-
			-	-	-
pH45	237.71	<0.001	CC	0.264	0.017
			CT	-0.146	0.019
			TT	-0.466	0.025
rigor	100.59	<0.001	CC	-0.731	0.093
			CT	0.155	0.103
			TT	1.808	0.138
meat colour	157.60	<0.001	CC	4.945	0.426
			CT	-2.039	0.471
			TT	10.179	0.635
con24	291.58	<0.001	CC	-2.186	0.117
			CT	1.362	0.131
			TT	3.545	0.177
pH24	7.90	<0.001	CC	-0.009	0.004
			CT	-0.014	0.004
			TT	0.004	0.006
con45	223.08	<0.001	CC	-2.680	0.267
			CT	-0.135	0.297
			TT	8.092	0.405

Chapter two

Genome-wide association analysis for growth, muscularity and meat quality in Piétrain pigs

Patrick Stratz[a], Robin Wellmann[a], Siegfried Preuss[a], Klaus Wimmers[b,*], and Jörn Bennewitz[a]

[a] Institute of Animal Husbandry and Breeding,
University of Hohenheim,
70599 Stuttgart, Germany
[b] Leibnitz Institute for Farm Animal Biology,
Research Unit Molecular Biology,
18196 Dummerstorf, Germany

Published in **Animal Genetics** (2014) 45: 350-356
doi: 10.1111/age.12133

Genome-wide association analysis for growth, muscularity and meat quality in Piétrain pigs

P. Stratz[a], R. Wellmann[a], S. Preuss[a], K. Wimmers[b,*] and J. Bennewitz[a]

[a] Institute of Animal Husbandry and Breeding, University of Hohenheim, D-70599 Stuttgart, Germany

[b] Leibnitz Institute for Farm Animal Biology, Research Unit Molecular Biology, Wilhelm-Stahl-Allee 2, D-18196 Dummerstorf, Germany

*Corresponding author
Tel.: +49 38208 68-700
Fax: +49 38208 68-702

Email addresses:
PS: Patrick.Stratz@uni-hohenheim.de
RW: r.wellmann@uni-hohenheim.de
SP: Siegfried.Preuss@uni-hohenheim.de
KW: wimmers@fbn-dummerstorf.de
JB: j.bennewitz@uni-hohenheim.de

Summary

Improvement of growth and meat quality is one of the main objectives in sire line pig breeding programs. Mapping quantitative trait loci for these traits using experimental crosses and a linkage-based approach has been frequently performed in the past. The Piétrain breed was often involved as a founder breed to establish the experimental crosses. This breed was selected for muscularity and leanness, but shows a relatively poor meat quality. It is frequently used as a sire line breed. With the advent of genome-wide and dense SNP chips in pig genomic research it is possible to conduct genome-wide association studies also within the Piétrain breed. In this study around 500 progeny tested sires were genotyped with 60k SNPs. Data filtering showed, that around 48k SNPs were useable in this sample. These SNPs were used to conduct a genome-wide association study for growth, muscularity and meat quality traits. Because it is known, that a mutation in the RYR1 gene located on chromosome 6 shows a major effect on meat quality, this mutation was included in the models. Single-marker and multi-marker association analysis were performed. The results revealed between zero and 8 significant associations per trait with $p < 5 \times 10^{-5}$. Of special interest are SNPs located on SSC6, 10 and 15.

Keywords: single-marker analysis, BayesC, meat performance

Introduction

Improvement of growth and meat quality is one of the main objectives in sire line pig breeding programs. Meat quality can be considered as a combination of sensory properties and technological traits. Sensory properties include smell, taste and tenderness, whereas technological traits, objectively measured, include pH value, drip loss and intramuscular fat. Economically, drip loss likely plays the most important role as a measure of meat quality in Germany Factors affecting meat quality are discussed in Fischer (2007) and Huff-Lonergan and Lonergan (2007). Quantitative genetic analysis revealed a moderate heritability of traits affecting meat quality, with estimates typically in the range of 0.2-0.4 (e.g. de Vries et al. 1994). In order to identify chromosomal regions affecting genetic variance of meat quality traits, linkage analysis has been widely performed and many QTL were identified (Hu et al., 2005). Highly significant QTL were frequently reported on chromosome six. The gene *RYR1* (Fujii et al., 1991) is located in a QTL region on chromosome 6. A transition of *C* to *T* at position 1843 in the *RYR1* gene (*RYR1*:g.1843C>T) leads to a substitution of Arginine to Cysteine at the amino acid 615. The *T* allele is associated with an increase in meat content, but also with a reduced meat quality.

So far, most QTL were mapped in experimental crosses. These crosses were usually established using divergent selected breeds (Rothschild et al., 2007). However, mapping precision was in general limited, leading to large confidence intervals. This is because within the final generation of experimental crosses the linkage disequilibrium (LD) covers a long range. To overcome this limit, in some studies several crosses were pooled and analysed jointly, leading to a higher statistical power (e.g. Walling et al., 2000; Rückert and Bennewitz, 2010). Another drawback of linkage mapping in experimental crosses is that it relies on a divergent fixation of QTL in the founder breeds. Although it was frequently shown, that mapped QTL do also segregate within founder breeds, this can not a priory be assumed. Hence, if mapped QTL are to be considered in marker assisted selection schemes, it has to be tested, if they are segregating within the population where selection is to be performed.

With the advent of genome-wide and dense SNP chips in livestock species it is possible to conduct genome-wide association studies also within populations (Goddard and Hayes, 2009). This approach uses LD between SNPs and the causative mutation within a population. The LD structure spans in general a much smaller genomic region compared to the structure observed in experimental crosses. Provided that the marker density is high, it is possible to

identify short chromosomal regions or even single SNPs associated with the causative mutation, leading to a higher mapping precision compared to linkage analysis. In addition, significant and validated markers can be included straightforwardly in marker assisted or genomic selection experiments. In pigs, the Illumina PorcineSNP60 BeadChip (Ramos et al., 2009) was successful used in association studies (e.g. Fan et al., 2011, Uimari et al., 2011).

The Piétrain breed is frequently used as a terminal sire breed in crossbred pig breeding schemes. This breed was selected for muscularity and leanness, but Piétrain animals have relatively poor meat quality. The *T* allele of the *RYR1* mutation described above is observed mainly in this breed. In Germany, Piétrain herdbook associations apply sire progeny testing on stations for various growth, carcass and meat quality traits. Some of them started to implement genomic selection by genotyping these progeny tested sires with the Illumina PorcineSNP60 BeadChip and to use them as the initial reference population (Wellmann et al., 2013). The aim of the present study was to conduct a genome-wide association analysis within Piétrain to identify SNPs that show a significant association with growth, muscularity, and meat quality traits.

Materials and Methods

Genotypes and Phenotypes

Totally 794 German Piétrain boars were genotyped with the PorcineSNP60 BeadChip (Ramos et al., 2009). Genotypes from individuals were filtered with respect to call rate (removal of SNPs with a call rate less than 95%), parent progeny conflicts (removal of SNP with parent progeny conflict greater than 2%), MAF (exclusion of SNP with a minor allele frequency less than 3%) and significant deviation from the Hardy-Weinberg-equillibrium ($p<0.0001$). Animals with more than 10% missing genotypes were excluded. Filtering was done using PLINK software (Purcell et al., 2007). Sporadic missing genotypes were imputed using fastPHASE (Scheet and Stephens, 2006). This filtering resulted in 771 animals and 48062 SNPs. A number of 351 markers had an unknown physical position in the porcine genome sequence (Sus scrofa *Build 10.2* assembly). Genotypes of the *RYR1*:g.1843C>T mutation (subsequently denoted as *RYR1*) were available.

The boars did not have own performance records, but were progeny tested. The following traits were considered: daily gain (DG) recorded on station during the fattening period, carcass lean content, estimated with Bonner formula (CLC), pH1 (measured 45 min post-

mortem in cutlet), intramuscular fat content (IMF) measured between the 13th/14th rib in cutlet, and drip loss (Drip) measured between the 12th/13th rib in cutlet. All traits were recorded according to guidelines of the Zentralverband der Deutschen Schweineproduktion e.V. (ZDS, 2004). A varying number of boars with progeny records were available for the traits. For a summary statistic see Table 1. The number of full-sibs for the boars was in between 1 and 3 with an average of 2 for each trait.

It was shown that yield deviations (*YD*) are the preferred trait variable in QTL mapping with progeny data (Seidenspinner et al., 2009). We reconstructed them from the results of the routine animal evaluation scheme, which uses an animal model. We started with the calculation of the yield deviation of each progeny *k* in litter *l* of boar *i* for each trait (YD_{ilk}). The *YD* is the yield corrected for all random and fixed effects that are included in routine animal evaluation, except the breeding value. The *YD* were summarised for each boar and litter as $\overline{YD}_{il} = \frac{1}{n_{il}} \sum_{k=1}^{n_{il}} YD_{ilk}$, where $\overline{YD}_{il}$ is the average yield deviation in litter *l* of boar *i* and n_{il} is the number of littermates (number of full-sibs). The $\overline{YD}_{il}$ were subsequently summarised for each boar across litters as $YD_i = 2\frac{1}{L_i} \sum_{l=1}^{L_i} \overline{YD}_{il}$, where L_i is the number of litters of boar *i*. Due to the fact that only one half of the genes were transmitted from boar *i* to his progenies, the average was multiplied by the factor 2.

Statistical analyses

Single marker association analysis

In order to test for the significance of the *RYR1* gene, the following model was applied to the traits.

$$YD_i = \mu + RYR1 + a_i + e_i, \quad (1)$$

where YD_i is the yield deviation of boar *i*, and μ is the overall mean. *RYR1* denotes the fixed effect of the genotypes at *RYR1*, a_i is a random polygenic effect of boar *i*, with a covariance structure as $a \sim N(0, \mathrm{A}\sigma_s^2)$, where A is the numerator relationship matrix derived from the pedigree and σ_s^2 is the sire variance. The pedigree was augmented with ancestor information from previous generations. Variance components for the traits were estimated using model (1) as well, either with or without *RYR1* included.

For genome-wide association analysis the second model was applied to each SNP in turn.

$$YD_i = \mu + RYR1 + b_m x_{im} + a_i + e_i, \quad (2)$$

where x_{im} denotes the number of copies of the allele with the higher frequency of SNP *m* (x_{im} = 0, 1, or 2) and b_m is the regression coefficient for SNP *m*. The other terms are as defined above. For the trait without significant *RYR1* effect (i.e. DG, see results section), this term was omitted. Because of the varying number of offspring observations used to estimate the *YD*, the residual variance was modelled heterogeneous. The derivation of the heterogeneous residual variance is described in the Supplemental information, see File S1. All models were solved using ASReml (Gilmour et al., 2006). For each SNP *m*, the null hypothesis was that the regression coefficient b_m was equal to zero. The alternative hypothesis was that the coefficient was significant different from zero. In total, 48062 tests were conducted, resulting in point-wise error probabilities *p*. Following the suggestion of Teyssedre et al. (2012), different thresholds were applied. The stringent threshold was 5×10^{-6}. It corresponds to an approximation of 10000 independent and Bonferroni-corrected tests in their study. The less stringent threshold 5×10^{-5} was applied to find associations across traits. In order to judge how many false positives were among the significant associations we applied the false discovery rate (FDR) technique. Therefore we calculated for each association test an FDR *q*-value using the software QVALUE (Storey 2002, Storey and Tibshirani, 2003). The FDR *q*-value of the significant SNP with the highest *p*-value ($p \approx 5\times10^{-5}$, $p \approx 5\times10^{-6}$, respectively) provided an estimate of the proportion of false positives among the significant associations.

Multi marker association analysis

In single marker association analysis putative LD structure among the SNPs is ignored. Due to extensive LD pattern in livestock (Goddard and Hayes 2009), it might be that even distant SNPs show a significant association with the causative mutation. This reduces mapping precision. If all SNP are fitted simultaneously, the SNP in highest LD with the causative mutation should have the largest effect on the trait. In order to model all SNPs simultaneously, we applied the BayesC approach of Verbyla et al. (2009, 2010).

The following model was applied using all M = 48062 SNPs simultaneously.

$$YD_i = \mu + \sum_{m=1}^{M} b_m x_{im} + e_i, \quad (3)$$

The terms are as described above. For traits with a significant *RYR1* effect, the *YDs* were pre-corrected for this effect. A mixture of two *t*-distributions was assumed for the SNP effects, where few SNP effects come from the t-distribution with large scale parameter, and a second larger set of SNPs comes from the t-distribution with small scale parameter. Both t-distributions had ν=4 degrees of freedoms. The scale parameter of marker *m* is proportional to $\kappa_m = (1-\gamma_m) * \varepsilon + \gamma_m$, where $\varepsilon = 0.01$. Whether a SNP effect *m* is from the distribution with large scale parameter is determined by the indicator variable $\gamma_m \sim bernoulli(pLD)$, which can take either 0 or 1, respectively, with *pLD* being the probability that γ_m is one. For the portion of SNP that are associated with a QTL we assumed $pLD = 0.02 \approx \frac{1000}{48062}$ (Wellmann and Bennewitz, 2011). The variance of the marker effects was calculated from the data as $\sigma_b^2 = \frac{\hat{\sigma}_s^2 (1-c)}{M\bar{h}E(\kappa_m^2)} \frac{\nu-2}{\nu}$, where $\hat{\sigma}_s^2$ is the estimated sire variance of the trait (results from model 1), $\bar{h}$ is the mean heterozygosity ($\bar{h} = 0.35$, not shown elsewhere), estimated as $\bar{h} = \frac{1}{M}\sum_{m=1}^{M} 2p_m(1-p_m)$, with p_m being the allele frequency of SNP *m*. See Wellmann and Bennewitz (2012) for a detailed derivation of this expression. The parameter *c* denotes for the fraction of additive genetic variance not explained by markers and was fixed at 0.25, as suggested by Wellmann et al. (2013). The residual variance was assumed to be heterogeneous. See File S1 for the calculation of the heterogeneous residual variance. SNP effects and expected probabilities of an SNP belonging to the distribution with the large scale parameter ($E(\gamma_m \mid y)$) were estimated as the means of the respective posterior distribution. We performed 50000 Gibbs sampling iterations. The first 25000 were discarded as burn-in. The software was taken from Wellmann and Bennewitz (2012).

Results and Discussion

Results from association analysis

The Piétrain breed was frequently involved as a founder breed to generate experimental crosses for QTL analysis (e.g. Liu et al., 2007; Mohrmann et al., 2006; Boysen et al., 2011). In recent studies we conducted joint QTL analysis for growth and meat quality traits in pigs using three connected F2 crosses, where Piétrain was used as a founder breed (Rückert and

Bennewitz, 2010, Rückert et al., 2012, Stratz et al., 2013). We were able to map numerous QTL affecting muscularity, growth and meat quality. However, it was not obvious, if the QTL segregate also within the Piétrain breed. Hence, the mapped QTL could not be used for breeding purposes to improve this breed. The current study is, to our knowledge, the first that reports the results from genome-wide association analysis within the Piétrain breed.

The results of model (1) are shown in Table 2. The effect of *RYR1* was highly significant for all traits except DG. The estimated sire variance decreased when *RYR1* was included in the model. This is most obvious for pH1 and Drip. For IMF the decrease was only small. Summary statistics of the results of model (2) are shown in Table 3. For the threshold $p < 5 \times 10^{-5}$ no significant associations could be found for IMF. For the other traits the number of significant SNPs was between 1 (CLC) and 8 (DG). The FDR among the SNPs was between 0.070 for Drip and 0.451 for pH1. For $p < 5 \times 10^{-6}$ the number of significant SNPs was reduced and no significant associations could be found for CLC and pH1. All the SNPs from model (2) with $p < 5 \times 10^{-5}$, $p < 5 \times 10^{-6}$, can be found in Table S1.

The intersection between SNPs with $p < 5 \times 10^{-5}$ (results from model (2)) and BayesC posterior probability $E(\gamma_m \mid y) \geq$ 0.0625 (results from model (3)) are listed in Table 4. Additional SNPs with $E(\gamma_m \mid y) \geq$ 0.0625 are shown in Table S2. The histogram of $E(\gamma_m \mid y)$ is shown in Figure 1. The correlation between the *F*-values and $E(\gamma_m \mid y)$ was calculated and was 0.147, 0.275, 0.226, 0.237 and 0.254 for IMF, DG, pH1, CLC and Drip, respectively. The most significant SNPs were not the ones having the highest $E(\gamma_m \mid y)$ (Table S1-S2) and vice versa. However, many SNP with $p < 5 \times 10^{-5}$ in model (2) also showed an elevated BayesC posterior probability (Table 4). In the following section the results listed in Table 4 are discussed.

For DG one significant SNP was found on SSC15 (Table 4). BayesC marker effects for SNPs located on SSC15 are shown in Figure 2. For Drip significant SNPs were found on SSC6 and 10 (Table 4). The SNPs on SSC6 and SSC10 are shown in Figure 3. For pH1 SNPs were found on SSC6. Hence, it seems that also within the Piétrain, multiple genes segregating on SSC6 affecting meat quality, in addition to *RYR1*. In the BayesC-results, two markers showed a substantial effect although the phenotypes were adjusted for the *RYR1* effect (Figure 4).

Genomic selection using low-density SNP arrays is currently implemented in this Piétrain population (Wellmann et al. 2013). The most interesting SNPs found in this study are candidate SNPs that could be considered to be included in the low density array. The results showed that a selection of SNPs for a low density array based solely on single marker association analysis might be suboptimal, because these SNPs did not show always largest effects in the BayesC-analysis. The results of both models differ in some cases, because in BayesC the LD structure between the markers is taken into account, whereas in single marker association analysis only the LD between the marker and the causative mutation is used. In general, some significant SNPs showed also substantial BayesC-marker effects, which supported the presence of a causative mutation close to these SNPs.

Apart from the size of the gene effect of the particular traits, important factors determining the power to detect true associations is the number of genotyped individuals and the reliability of the observations used. In our study around 500 individuals with reliable progeny yield deviations were used. The results showed that the power of this study is limited. This can be seen by the relative low number of significant SNPs and the high FDR among the significant SNPs. Hence, confirmation of significant SNPs in an independent sample is highly needed in order to pinpoint true associations.

Conclusion

SNPs affecting growth, muscularity, and meat quality traits within Piétrain pigs were identified. Of special interest are regions harbouring SNPs on SSC6, 10 and 15. The *RYR1* gene did not explain all the genetic variance for meat quality attributable to the chromosomal region on SSC6.

Acknowledgements

The project was funded by the Ministerium für Ländlichen Raum und Verbraucherschutz Baden-Württemberg, Germany, Schweinezuchtverband Baden-Württemberg and Schweinezuchtverband Schleswig-Holstein, both in Germany. PS received funding from the Carl-Zeiss-Stiftung, Stuttgart, Germany.

References

Boysen, T. J., J. Tetens, G. Thaller (2011). Evidence for additional functional genetic variation within the porcine IGF2 gene affecting body composition traits in an experimental Piétrain × Large White/Landrace cross. *Animal* **5**: 672–677.

De Vries, A. G., P. G. van der Wal, T. Long, G. Eikelenboom, J. W. M. Merks (1994). Genetic parameters of pork quality and production traits in Yorkshire populations. *Livestock Production Science* **40**: 277–89.

Fan, B., S. K. Onteru, Z. Q. Du, D. J. Garrick, K. J. Stalder, M. F. Rothschild (2011). Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *PLoS ONE* **6**(2): e14726. doi:10.1371/journal.pone.0014726.

Fischer, K. (2007). Drip loss in pork: influencing factors and relation to further meat quality traits. *Journal of Animal Breeding* and *Genetics* **124**(Suppl. 1): 12–18.

Fujii, J., K. Otsu, F. Zorzato, S. DeLeon, V. K. Khanna, J. E. Weiler, P. J. O'Brien, D. H. McLennan (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253**: 448–51.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, R. Thompson (2006). ASReml user guide release 2.0. VSN International Ltd., Hemel Hempstead, U.K.

Goddard, M. E., and B.J. Hayes (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10**: 381–391.

Hu, Z. L., S. Dracheva, W. Jang, D. Maglott, J. Bastiaansen, M. F. Rothschild, J. M. Reecy (2005). A QTL resource and comparison tool for pigs: PigQTLDB. *Mammalian Genome* **16**: 792–800.

Huff-Lonergan, E., and S. M. Lonergan (2007). New frontiers in understanding drip loss in pork: recent insights on the role of post-mortem muscle biochemistry. *Journal of Animal Breeding and Genetics* **124**(Suppl 1): 19–26.

Liu, G., D. G. J. Jennen, E. Tholen, H. Juengst, T. Kleinwächter, M. Hölker, D. Tesfaye, G. Ün, H. J.Schreinemachers, E. Murani, S. Ponsuksili, J. J. Kim, K. Schellander, K. Wimmers (2007). A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Piétrain resource population. *Animal Genetics*, **38**: 241–252.

Mohrmann, M., R. Roehe, P. W. Knap, H. Looft, G. S. Plastow, E. Kalm (2006). Quantitative trait loci associated with AutoFOM grading characteristics, carcass cuts and chemical body composition during growth of Sus scrofa. *Animal Genetics* **37**: 435–43.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P.

Sklar, P. I. W. de Bakker, M .J. Daly, P. C. Sham (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *The American Journal of Human Genetics* **81**: 559-575.

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. V. Tassell, J. F.Taylor, R. T. Wiedmann, L. B. Schook, 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.

Rothschild, M. F., Z. L. Hu, Z. Jiang (2007). Advances in QTl mapping in pigs. *International Journal of Biological Sciences* **3**:192-197

Rückert, C., and Bennewitz J. (2010). Joint analysis of three connected F_2-crosses on pigs. *Genetics Selection Evolution* **42**: 40.

Rückert, C., P. Stratz, S. Preuss, and J. Bennewitz (2012). Mapping QTL for growth and muscling traits in three connected porcine F_2 crosses. *Archives of Animal Breeding* **55**: 577-589.

Scheet, P., and M. Stephens (2006). A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. *The American Journal of Human Genetics* **78**: 629 – 644.

Seidenspinner, T., J. Bennewitz, F. Reinhardt, and G. Thaller (2009). Need for sharp phenotypes in QTL detection for calving traits in dairy cattle. *Journal of Animal Breeding and Genetics* **126**: 455 – 462.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* **64**: 479-498.

Storey, J. D., and R. Tibshirani (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*. 100: 9440–9445.

Stratz, P., C. Baes, C. Rückert, S. Preuss, and J. Bennewitz (2013). A two-step approach to map quantitative trait loci for meat quality in connected porcine F2 crosses considering main and epistatic effects. *Animal Genetics*. **44**: 14-23.

Teyssedre, S., M. C. Dupuis, G. Guerin, L. Schibler, J. M. Denoix, J. M. Elsen, and A.Ricard (2012). Genome-wide association studies for osteochondrosis in French Trotter horses. *Journal of Animal Science.* **90**: 45–53.

Uimari, P., A. Sironen, and M. L. Sevón-Aimonen (2011). Whole-genome SNP association analysis of reproduction traits in the Finnish Landrace pig breed. *Genetics Selection Evolution* **43**: 42-49.

Verbyla, K. L., B. J. Hayes, P. J. Bowman, and M. E. Goddard (2009). Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genetics Research Cambridge* **91**: 307–311.

Verbyla, K. L., P. J. Bowman, B. J. Hayes and M. E. Goddard (2010). Sensitivity of genomic selection to using different prior distributions. *BMC Proceedings* **4** (Suppl. 1): S5.

Walling, G. A., P. M. Visscher, L. Andersson, M. F. Rothschild, L. Wang, G. Moser, M. A. M. Groenen, J. P. Bidanel, S. Cepica, A. L. Archibald, H. Geldermann, D. J. De Koning, D. Milan, and C. S. Haley (2000). Combined analyses of data from quantitative trait loci mapping studies: Chromosome 4 effects on porcine growth and fatness. *Genetics* **155:** 1369-1378.

Wellmann, R., and J. Bennewitz (2011). The contribution of dominance to the understanding of quantitative genetic variation. *Genetics Research* **93***:* 139–154.

Wellmann, R., and J. Bennewitz (2012). Bayesian models with dominance effects for genomic evaluation of quantitative traits. *Genetics Research*. **94**: 21-37.

Wellmann, R., S. Preuß, E. Tholen, J. Heinkel, K. Wimmers, and J. Bennewitz (2013). Genomic selection using low density marker panels with application to a sire line in pigs. *Genetics Selection Evolution.* **45**. 28.

ZDS, Zentral Verband der Deutschen Schweineproduktion e. V. (2004). Richtlinie fuer die Stationspruefung auf Mastleistung, Schlachtkoerperwert und Fleischbeschaffenheit beim Schwein. Bonn, Germany.

Table 1 Descriptive statistics for the traits, number of sires (N), minimum, mean, and maximum number of progeny.

Trait	Trait abbreviation	N	Min-Mean-Max number of progeny
Daily gain, recorded on station during fattening period	DG	571	5-10-113
Carcass lean content, estimated with Bonner formula	CLC	571	5-10-114
pH measured 45min post- mortem	pH1	584	5-10-113
Intramuscular fat content measured between the 13th/14th rib in cutlet	IMF	353	4-7-41
Drip loss measured between the 12th/13th rib in cutlet	Drip	497	4-8-64

Table 2 Test statistic of the *RYR1*:g.1843C>T genotype, estimated sire variance ($\hat{\sigma}_s^2$) without and with *RYR1* included in the model, results from model (1).

Trait	*F*-value	*p*-value	$\hat{\sigma}_s^2$ without *RYR1*	$\hat{\sigma}_s^2$ with *RYR1*
DG	1.31	0.255	1649.450(287.000)	-
CLC	40.62	<0.001	0.505(0.091)	0.497(0.090)
pH1	180.17	<0.001	0.058(0.007)	0.013(0.004)
IMF	15.43	<0.001	0.010(0.003)	0.009(0.003)
Drip	111.22	<0.001	2.511(0.349)	1.375(0.287)

Standard errors are given in parentheses.

Table 3 Number of significant SNPs (n SNPs) for the traits and the different threshold levels and FDR *q*-values (FDR) of the significant SNP with the highest error probabilities ($p \approx 5\times10^{-5}$, $p \approx 5\times10^{-6}$, respectively) are shown. Results from model (2).

Trait	$p < 5\times10^{-5}$		$p < 5\times10^{-6}$	
	n SNPs	FDR	n SNPs	FDR
DG	8	0.255	1	0.164
CLC	1	0.385	0	-
pH1	2	0.451	0	-
IMF	0	-	0	-
Drip	2	0.070	2	0.070

For all traits except DG the effect of *RYR1* was included in the model.

Table 4 Chromosomal regions harbouring SNPs with $p \approx 5\times10^{-5}$ (model (2)) and with BayesC posterior probabilities $E(\gamma_m \mid y) \geq 0.0625$ (model (3)), their positions in the genome (chromosome, position), $E(\gamma_m \mid y)$, and *F*-value.

Trait	Marker	SSC	Position (Mb)	$E(\gamma_m \mid y)$	*F*-value
DG	ASGA0069460	15	97.1009	0.114	17.38
Drip	ALGA0035551	6	29.478	0.083	23.85
Drip	ALGA0059133	10	58.503	0.339	22.65
pH1	ALGA0116528	6	83.970	0.099	19.07
pH1	MARC0098796	6	83.970	0.070	19.07

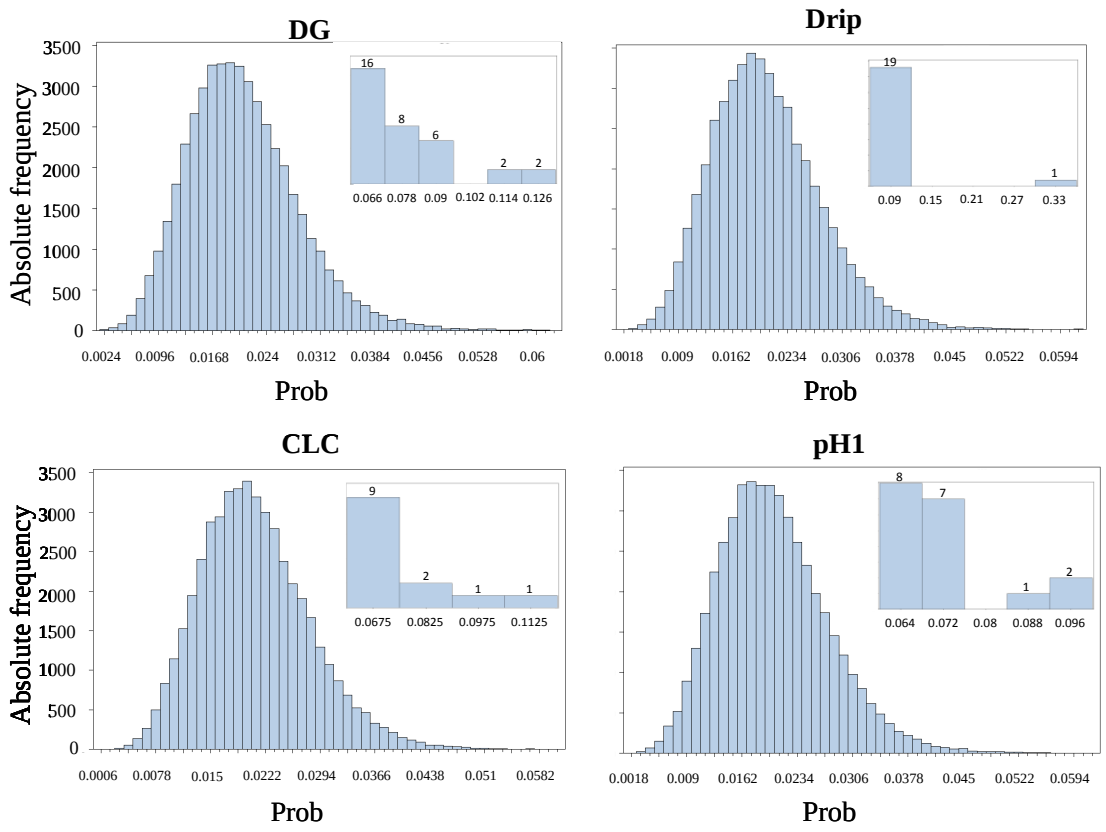


Figure 1 Distribution of the BayesC posterior probabilities for an SNP belonging to the distribution with the larger scale parameter ($E(\gamma_m \mid y)$). The x-axis denotes the $E(\gamma_m \mid y)$, the y-axis the absolute frequency of the markers. The absolute frequency of markers with $E(\gamma_m \mid y) < 0.0625$ is shown in the main window, whereas in the upper right corner only markers above that limit are shown, which are of most interest. Their absolute frequency is written above the bars.

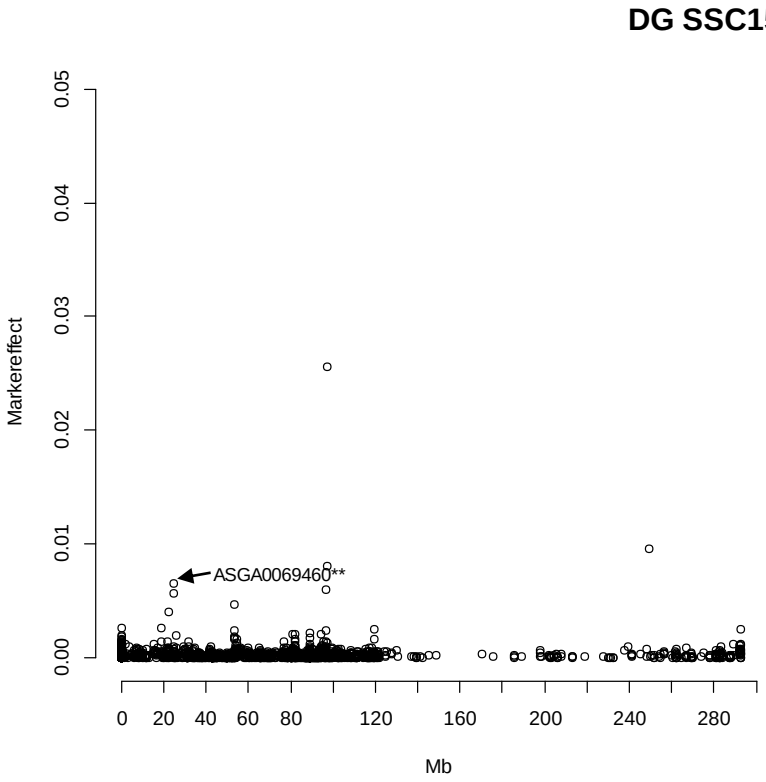


Figure 2 BayesC marker effects for DG on SSC15. The x-axis denotes chromosomal position in Mb and the y-axis the marker effect in units of sire standard deviation. The significant SNP with $p < 5 \times 10^{-5}$ (results from model 2) is marked with two asterisks.

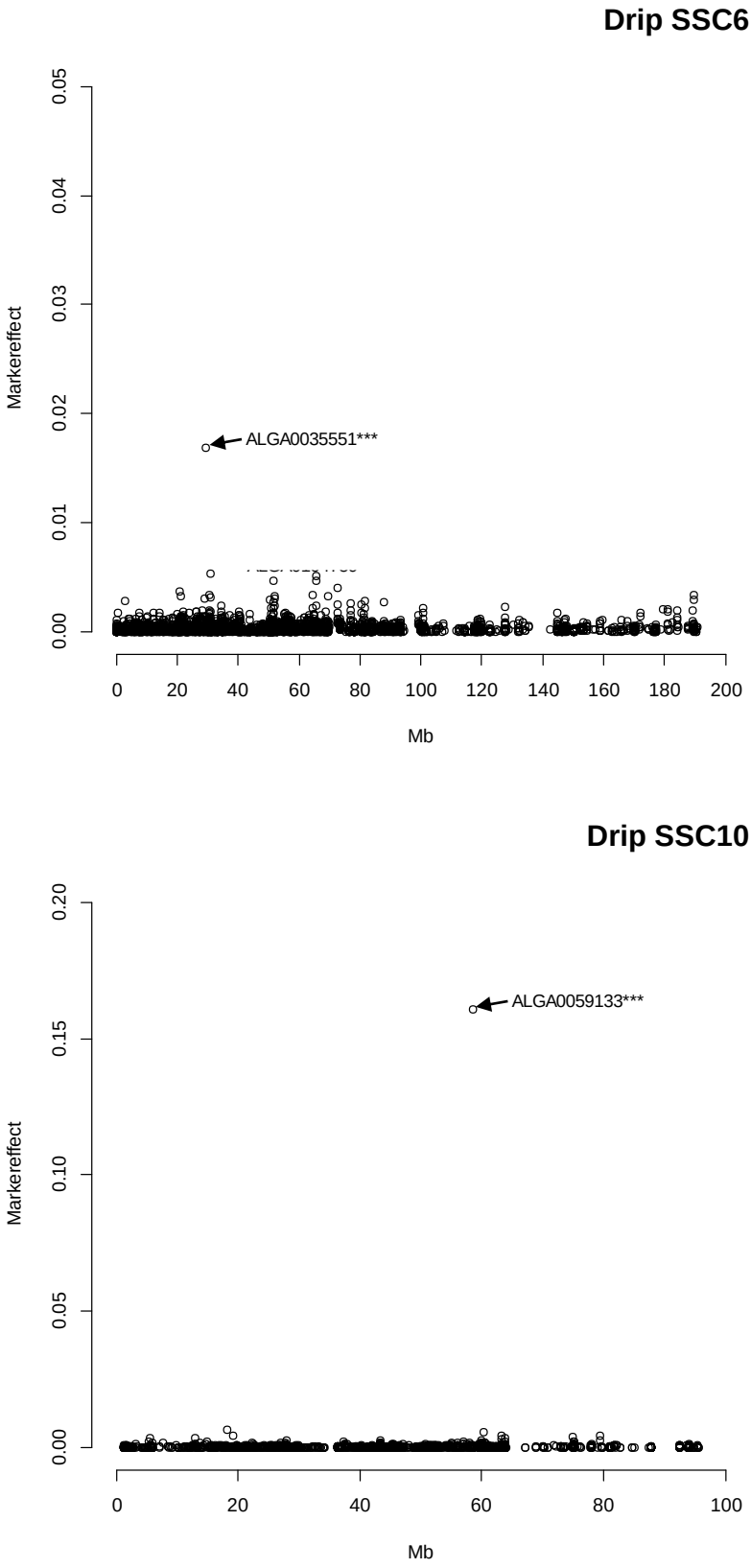


Figure 3 BayesC marker effects for Drip on SSC6 and SSC10. The x-axis denotes chromosomal position in Mb and the y-axis the marker effect in units of sire standard

deviation. Significant SNPs with $p < 5 \times 10^{-6}$ (results from model 2) are marked with three asterisks.

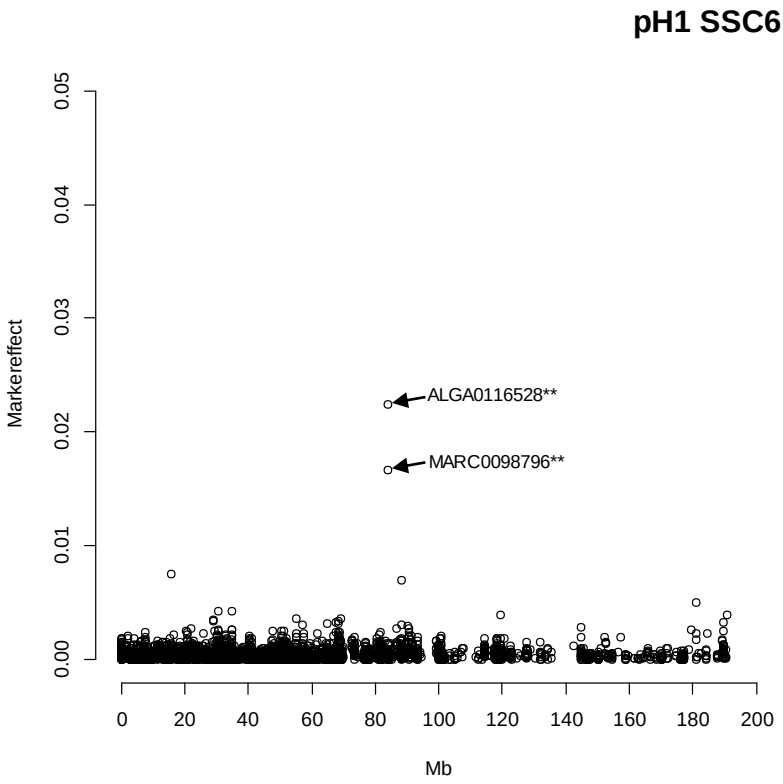


Figure 4 BayesC marker effects for pH1 on SSC6. The x-axis denotes chromosomal position in Mb and the y-axis the marker effect in units of sire standard deviation. Significant SNPs with $p < 5 \times 10^{-5}$ (results from model 2) (results from model 2) are marked with two asterisks.

Appendix

In the first part of the appendix, the calculation of the residual variances for model (1) and (2) is described. We can write

$$YD_{ilk} = BV_{ilk} + u_{il} + e_{ilk}, \quad (1)$$

whereby YD_{ilk} is the yield of an offspring of boar *i*, corrected for all random and fixed effects that are included in routine animal evaluation, except the breeding value, u_{il} is the environmental effect of the *l*-th litter, e_{ilk} is the residual, and BV_{ilk} is the breeding value of the offspring of boar *i*. The breeding value can be written as

$$BV_{ilk} = \frac{s_i + d_{il}}{2} + m_{ilk}, \quad (2)$$

where s_i is the breeding value of the sire, d_{il} is the breeding value of the sow that produced the *l*-th litter of boar *i*, and m_{ilk} is the mendelian sampling effect. Mating partners of boar *i* are assumed to be randomly chosen. Inserting (2) in (1) results after some algebra

$$Var(YD_i|s_i) = \frac{V_A + 4V_U}{L_i} + \frac{2V_A + 4V_E}{NeO_i},$$

where L_i is the number of litters of boar *i*, and the term

$$NeO_i = \frac{L_i}{\frac{1}{L_i}\sum_{l=1}^{L_i}\frac{1}{n_{il}}}$$

is called the effective number of offspring of boar *i*, n_{il} is the number of offspring in litter *l*, V_A is the additive variance, $V_U = \text{var}(u_{il})$ is the variance of the environment effect of the litter, and $V_E = \text{var}(e_{ilk})$ is the variance of the residual effect. The effective number of offspring differs from the actual number of offspring if the litters contain unequal numbers of offspring with records. Since $E(YD_i|s_i) = s_i$, from the law of total variance it follows that

$$Var(YD_i) = Var(s_i) + \frac{V_A + 4V_U}{L_i} + \frac{2V_A + 4V_E}{NeO_i}$$

The first term is the variance of the breeding value of the boar, and the sum of the other terms is the variance of the error by using the representation

$$YD_i = \mu + s_i + e_i.$$

An approximation of this model was done, because the proposed heterogeneous variance model was not available in ASReml (Gilmour et al., 2006). Therefore, for simplicity, we

assumed that the same number of full sibs was tested from each litter. Hence, $\lambda = \frac{NeO_i}{L_i}$ is a constant. It follows that the variance of the error is

$$Var(e_i) = \frac{\alpha}{NeO_i},$$

where $\alpha = \lambda V_A + 4\lambda V_U + 2V_A + 4V_E$ is an unknown parameter of the model. This heterogeneous residual variance was used in the single marker association analysis (model 1 and 2). For the multi marker association analysis (BayesC, model 3 in main text), we used the following representation of the breeding value of sire *i*:

$$s_i = \sum_{m=1}^{M} b_m x_{im} + p_i,$$

where the left summand is the sum of the *m* random marker effects, and p_i is a polygenic term. From the representation

$$YD_i = \mu + s_i + e_i$$

we have

$$Var(E(YD_i|x_i)) = \sigma_b^2 \sum_{m=1}^{M} x_{im}^2 + cV_A,$$

where the left summand is the variance explained by the markers and the right summand is the part of the additive variance not explained by the markers and x_i is the genotype vector of boar *i*. Since only the genotypes of the sires are fixed (and not the genotypes of the ungenotyped randomly chosen mating partners) we have

$$Var(YD_i|x_i) = Var(YD_i|s_i)$$

which was calculated above. From the law of total variance it follows that

$$Var(YD_i) = \sigma_b^2 \sum_{m=1}^{M} x_{im}^2 + cV_A + \frac{\alpha}{NeO_i},$$

where

$$\sigma_e^2 = cV_A + \frac{\alpha}{NeO_i}$$

is the variance of the residual. In order to allow for the use of existing software for BayesC, the parameter α was estimated with ASReml (model 1 in main text) and BayesC was applied by assuming that

$$\sigma_e^2 \propto cV_A + \frac{\hat{\alpha}}{NeO_i}.$$

References

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson (2006). ASReml user guide release 2.0. VSN International Ltd., Hemel Hempstead, U.K.

Table S1 Chromosomal regions harbouring SNPs with $\rho < 5 \times 10^{-5}$ (model (2)), their position in the genome (chromosome, position), $E(\gamma_m \mid y)$, *F*-value, and FDR *q*-value (FDR).

Trait	Marker	SSC	Position (Mb)	$E(\gamma_m \mid y)$	*F*-value	FDR
DG	**ASGA0095267**	**6**	**57.3788**	**0.050**	**22.04**	**0.164**
DG	ALGA0123547	6	58.4457	0.050	19.57	0.255
DG	ASGA0010149	2	47.9801	0.038	17.86	0.255
DG	ALGA0032754	5	67.6703	0.049	17.43	0.255
DG	ASGA0069460	15	97.1009	0.114	17.38	0.255
DG	ALGA0032748	5	67.6138	0.056	17.14	0.255
DG	ASGA0010155	2	47.9806	0.022	17.11	0.255
DG	MARC0026068	6	57.3788	0.039	17.01	0.255
CLC	ALGA0005165	1	106.7391	0.041	18.76	0.385
pH1	ALGA0116528	6	83.9701	0.099	19.07	0.451
pH1	MARC0098796	6	83.9701	0.070	19.07	0.451
Drip	**ALGA0035551**	**6**	**29.4777**	**0.083**	**23.85**	**0.070**
Drip	**ALGA0059133**	**10**	**58.5032**	**0.339**	**22.65**	**0.070**

Significant SNP with the highest error probabilities ($\rho < 5 \times 10^{-6}$) are written in boldface.

Table S2 Chromosomal regions harbouring SNPs with BayesC posterior probabilities $E(\gamma_m \mid y) \geq 0.0625$ (model (3)), their position in the genome (chromosome, position), $E(\gamma_m \mid y)$, and *F*-value.

Trait	Marker	SSC	Position (Mb)	$E(\gamma_m \mid y)$	*F*-value
CLC	INRA0002055	1	58.070	0.070	10.55
CLC	H3GA0002760	1	135.380	0.064	11.5
CLC	ASGA0007824	1	260.374	0.073	12.22
CLC	MARC0038518	5	0.277	0.079	7.57
CLC	ALGA0121463	5	12.416	0.096	16.54

CLC	ASGA0084324	5	12.416	0.117	16.54
CLC	MARC0073291	9	64.732	0.069	11.26
CLC	MARC0080678	9	160.464	0.064	10.47
CLC	MARC0089734	10	82.220	0.065	10.99
CLC	MARC0084792	13	31.901	0.064	12.89
CLC	ALGA0121217	13	56.086	0.072	15.09
CLC	MARC0084251	13	123.197	0.070	13.25
CLC	ALGA0074803	14	0.001	0.087	8.73
DG	CASI0009218	1	16.692	0.077	8.1
DG	ASGA0094502	1	42.770	0.065	11.1
DG	ALGA0002190	1	44.095	0.072	3.35
DG	H3GA0005056	1	259.425	0.068	8.63
DG	ALGA0010815	1	263.439	0.090	10.96
DG	ALGA0121048	2	23.082	0.073	7.04
DG	ASGA0010202	2	47.988	0.065	11.49
DG	ASGA0089130	3	20.951	0.080	10.18
DG	ASGA0022952	4	121.829	0.064	9.83
DG	ALGA0029567	4	124.784	0.067	9.73
DG	ALGA0032718	5	67.410	0.088	15.56
DG	ASGA0026268	5	68.816	0.095	13.38
DG	MARC0070754	5	72.799	0.064	6.43
DG	MARC0035525	6	5.678	0.078	4.47
DG	DIAS0000949	6	146.154	0.088	9.76
DG	MARC0029221	6	171.866	0.073	11.37
DG	ALGA0038836	7	28.496	0.067	6.16
DG	MARC0111713	9	66.473	0.069	6.5
DG	MARC0056124	10	10.715	0.066	8.74
DG	ASGA0050290	11	28.024	0.089	8.99
DG	MARC0026806	13	0.001	0.070	6.22
DG	ALGA0072425	13	82.495	0.066	6.63
DG	H3GA0037486	13	83.007	0.126	15.21
DG	H3GA0037490	13	83.092	0.070	8.83
DG	MARC0023388	13	94.195	0.080	9.76

DG	H3GA0037662	13	103.003	0.090	9.69
DG	DIAS0001351	14	130.609	0.120	10.26
DG	MARC0050687	14	130.612	0.121	10.26
DG	ALGA0085130	15	22.468	0.063	8.25
DG	ASGA0069460	15	24.630	0.075	11.65
DG	ALGA0087665	15	97.072	0.069	15.19
DG	ALGA0087667	15	97.101	0.114	17.38
DG	DIAS0000678	15	249.573	0.082	12.57
DG	ALGA0089402	16	24.542	0.063	5
Drip	ASGA0001096	1	16.753	0.080	8.35
Drip	ALGA0000100	1	16.781	0.070	9.99
Drip	ALGA0001915	1	40.790	0.120	1.08
Drip	MARC0055037	1	228.698	0.077	12.74
Drip	ASGA0008250	1	270.499	0.087	14.6
Drip	M1GA0004871	3	84.261	0.079	12.03
Drip	ALGA0030881	5	30.618	0.066	6.02
Drip	ALGA0034199	5	85.544	0.073	11.17
Drip	ALGA0034219	5	86.139	0.068	12.94
Drip	ALGA0035551	6	29.478	0.083	23.85
Drip	ALGA0104759	6	30.856	0.071	13.41
Drip	ALGA0108400	6	65.737	0.066	6.47
Drip	CASI0008871	7	129.202	0.068	10.23
Drip	ASGA0042087	9	25.421	0.063	3.5
Drip	H3GA0029101	10	18.115	0.074	5.66
Drip	ALGA0059133	10	58.503	0.339	22.65
Drip	ALGA0101388	10	74.999	0.064	8.02
Drip	ASGA0050169	11	25.350	0.086	9.34
Drip	MARC0064755	12	44.403	0.091	12.76
Drip	ASGA0103690	16	17.875	0.095	11.93
IM	ALGA0056816	10	19.097	0.068	8.96
IM	ASGA0103107	10	52.039	0.070	14.82
pH1	ALGA0002190	1	44.095	0.068	7.33
pH1	H3GA0006315	2	33.624	0.094	13.01

pH1	ALGA0016913	2	101.888	0.067	6.5
pH1	ALGA0123033	2	161.039	0.090	8.21
pH1	MARC0077006	2	161.039	0.067	6.91
pH1	ASGA0100278	3	2.457	0.069	7.07
pH1	ASGA0025477	5	54.232	0.065	16.28
pH1	ASGA0025481	5	55.778	0.063	8.12
pH1	ALGA0116528	6	83.970	0.099	19.07
pH1	MARC0098796	6	83.970	0.070	19.07
pH1	MARC0049988	7	124.216	0.063	7.3
pH1	ALGA0059205	10	59.932	0.075	14.15
pH1	ALGA0059207	10	59.940	0.075	13.1
pH1	ASGA0048247	10	59.970	0.072	9.23
pH1	ASGA0049437	11	4.868	0.063	3.43
pH1	ASGA0055015	12	47.544	0.063	3.75
pH1	ALGA0089171	16	17.410	0.068	8.62
pH1	ASGA0077407	17	74.326	0.064	11.89

Chapter three

Functional analysis of selected candidate genes to characterize meat quality traits in a segregating pig population

Patrick Stratz, Maria Maushammer, Siegfried Preuss,
and Jörn Bennewitz

Institute of Animal Husbandry and Breeding,
University of Hohenheim,
70599 Stuttgart, Germany

Unpublished Data

Functional analysis of selected candidate genes to characterize meat quality traits in a segregating pig population

P. Stratz, M. Maushammer, S. Preuss and J. Bennewitz

Institute of Animal Husbandry and Breeding, University of Hohenheim, D-70599 Stuttgart, Germany

Email addresses:
PS: Patrick.Stratz@uni-hohenheim.de
MM: maushammer.maria@web.de
SP: Siegfried.Preuss@uni-hohenheim.de
JB: j.bennewitz@uni-hohenheim.de

Summary

The aim of the study was the functional investigation of genes located in the vicinity of significant SNPs found in genome-wide association studies (GWAS; Stratz et al., 2014). A literature search was conducted to suggest one putative candidate gene, called *MYOD1*. Traits for growth, muscularity and meat quality were recorded from 20 boars out of a segregating Piétrain population at the LSZ in Boxberg. Muscle and blood tissue were collected from those animals after slaughtering. Muscle tissue was homogenized and RNA was isolated to record the gene expression of *MYOD1*. DNA was isolated from blood and the SNP ASGA0010149:g. 47980126G>A was genotyped. Significant Correlations (FDR *q-value*<0.15) between the expression of *MYOD1* and back fat length and daily gain were found. Associations between the traits recorded in Boxberg, respectively the expression of *MYOD1* and the SNP ASGA0010149:g. 47980126G>A were tested. The gene substitution effect was significant (FDR *q-value*<0.15) for the traits back fat thickness measured with the FOM-instrument and back fat thickness at the middle of the back (thinnest part), but not for the expression of *MYOD1*. Due to the low statistical power the results have to be considered critical. Increasing the number of animals might increase the power to detect true associations.

Keywords: functional analysis, *MYOD1*, meat quality, segregating Piétrain population

Introduction

Heritability estimates of 0.2-0.4 indicate considerable genetic variance for meat quality traits (e.g.: De Vries et al., 1994; Borchers et al., 2007). The availability of whole genomic sequence data in pigs was leading to a progress in genetics and genomics. A high density in marker data enables a genome-wide search for mutations influencing complex traits. A first step toward this goal is made by GWAS (Goddard and Hayes, 2009), which enable the search for candidate genes in the vicinity of significant SNPs. A further step toward the determination of putative candidate genes is to perform tissue and development specific expressions studies (Rockman and Kruglyak, 2006). Expressions are moderate to high heritable traits (Schliekelman 2008). Considering the expression as phenotype, SNPs can be detected influencing the transcription of a gene and hence the trait they determine. The aim of the study was to search for candidate genes located next to significant SNPs (Stratz et al., 2014) and to investigate them functionally.

Materials and Methods

Based on GWAS (Stratz et al., 2014) candidate genes were searched in the vicinity of significant SNPs (*P-value*<0.0001) using the NCBI map viewer (Dombrowski and Maglott, 2002). The SNP ASGA0010149:g. 47980126G>A is located on Sus Scrofa Chromosom (SSC) 2 at position 47980126 and characterized through a nucleotide exchange of guanin to adenine. It is located in the intron of the gene *SERGEF* (secretion regulating guanine nucleotide exchange factor). In the study of Stratz et al. (2014) ASGA0010149:g. 47980126G>A was significant for the trait DG. Several genes located close to the SNP are listed in Table 1 but only one putative candidate gene called *MYOD1*, which was already investigated in two other studies, was suggested for functional studies. Lee et al. (2012) and Urbański and Kurył (2004) found that polymorphisms in the gene *MYOD1* are influencing muscularity and meat quality traits. STRING 9.1 (Franceschini et al., 2013) was used to construct networks for *MYOD1*. 20 Piétrain boars, kept at the Landesanstalt für Schweinezüchtung (LSZ) in Boxberg, were slaughtered with a final weight between 102kg and 110.3 kg and aged between 166 and 218 days. Traits for growth, muscularity and meat quality were recorded according to guidelines of the Zentralverband der Deutschen Schweineproduktion e.V. (ZDS, 2004). The traits are listed in Table S1. For a summary statistic see Table S2. Tissue samples including muscle (*M. semimembranosus*) and blood tissue were taken from the boars in the slaughterhouse at four different days. High-quality RNA was isolated out of muscle tissue using the RNeasy Fibrous Tissue Midi Kit (QIAgen).

The RNA was transcribed to cDNA using Affinity Script QPCR cDNA Synthesis Kit (Agilent Technologies). For gene expression analysis, primers for the gene *MYOD1* were designed with the program Primer3Plus (Untergasser et al., 2012). Information on the primers used for real-time qPCR for *MYOD1* and the housekeeping gene *RPL32* is listed in Table S1. Real-time qPCR amplification was carried out using the BrilliantII SYBR Green QRT-PCR Kit (Agilent Technologies) and the CFX96 Real-Time PCR Detection System (Bio-Rad). For each sample, two technical replicates were measured. *RPL32* was used as reference gene for the normalization calculation of relative quantification of expression (Muráni et al., 2007). For a summary statistic of the normalized expression of *MYOD1* see Table 2. The normalized expressions of *MYOD1* were $\log_2$-transformed due to their distribution (Ponsuksili et al., 2010; Steibel et al., 2011).The normalized and $\log_2$-transformed expressions as well as the traits recorded at the LSZ in Boxberg were pre-corrected for the fixed effect of slaughter day. The Pearson correlation coefficient was computed for the relation between the normalized $\log_2$-transformed and pre-corrected expressions of *MYOD1* and the pre-corrected traits recorded in Boxberg.

High-quality DNA was isolated out of blood using the Maxwell® 16 instrument (Promega). A PCR-based restriction fragment length polymorphism (RFLP) analysis was applied. Primers were designed with the program Primer3Plus (Untergasser et al., 2012). MspI was used as restriction endonuclease to genotype the SNP ASGA0010149:g. 47980126G>A. Information on the primers used for PCR based RFLP is given in Table S2. Additional information on the restriction fragment length is given in Table S3. Afterwards the polymorphism was tested for Hardy-Weinberg.

An association analysis was conducted by the regression of pre-corrected traits, respectively the normalized $\log_2$-transformed and pre-corrected expressions of *MYOD1*, on the number of A-alleles at the SNP_*MYOD1*.

Results and Discussion

In Figure 1 the results from the network analysis are shown. *MYOD1* is coding for the myogenic differentiation factor 1. As myogenic factor it has an effect in muscle differentiation, activates muscle specific promotors and induces the differentiation of fibroblasts to myoblasts (Parker et al., 2003). The gene product has an effect on the myocyte enhancer factor 2A (*MEF2A*). The myocyte enhancer factor 2A is a transcription activator. It is involved in the activation of numerous growth factors and stress induced genes (Black and Olson, 1998). Similarly the myogenic differentiation factor 1 activates the myocyte enhancer

factor 2C. This is a transcription factor, controlling the cardiale morphogenesis und myogenesis (Juszczuk-Kubiak et al., 2012). The myogenic differentiation factor 1 binds to and activates *MYF6*. The gene product of *MYF6*, the myogenic factor 6, induces a differentiation of the fibroblasts to myoblasts (Parker et al., 2003). Furthermore it has an effect on the expression of *MYH3* and *MYH4*, coding for myosin heavy chain 3 and 4, whereby myosin heavy chain 3 is expressed in the embryonic stadium (Tajsharghi et al., 2008). Myosin is a motor protein and an essential component of thick filaments of the striated musculature (Tajsharghi et al., 2008).

The correlations between the log_2-transformed pre-corrected normalized expressions and the pre-corrected traits recorded at the LSZ in Boxberg are listed in Table 3. To correct for multiple testing the false discovery rate (FDR) was conducted using the QVALUE software (Storey, 2002). For two correlations an FDR *q-value*<0.15 was calculated. The correlations between the expressions of *MYOD1* and the traits BFTL and DG were significant (*P-value*<0.05).

The SNP_*MYOD1* is in Hardy-Weinberg equilibrium (*P-value*<0.05) and was therefore suitable for association studies. Regressing the pre-corrected traits recorded in Boxberg on the number of A-allels at the SNP_*MYOD1* resulted into significant gene substitution effects (*P-value*<0.05, FDR *q-value*<0.15) for the traits backFOM and BFTB (Table 4). Regressing the log_2-transformed pre-corrected normalized expression on the number of A-alleles at the SNP_*MYOD1* resulted into a non significant gene substitution effect (*P-value*<0.815). Hence no association between the genotype at the SNP_*MYOD1* and the expression of the gene *MYOD1* exists.

The results of this study have to be examined critically. As if already mentioned in the study of Schliekelman (2008) the sample size is a factor which determines the power to detect eQTL. The low sample size may be responsible for the discrepancy between the results of the GWAS conducted by Stratz et al. (2014) and the results of this association study. For the future the number of samples should be higher. Traits or expressions could also be summarized to units to reduce the number of tests without any significant loss of information (Ponsuksili et al., 2010; Schliekelman 2008).

The stage of development is also playing an essential role in expression studies. Zhao et al. (2011) used two different pig breeds (Landrace and Lantang) to investigate the expression of genes being involved in the muscle development during different stages from embryo to the

adult animal. The authors found an up - and down regulation in the expression of the gene *MYOD1* in prenatal stages.

Conclusion

The highest correlation was found for the expressions of *MYOD1* and the traits BFTL and DG. Slightly significant gene substitution effects were found for the traits backFOM and BFTB. However no association between the expression of the gene *MYOD1* and the genotype at the SNP_*MYOD1* exists. For further expression studies considering growth, muscularity or meat quality traits it is suggested to increase the sample size and to take muscle tissues from different developmental stages.

Acknowledgements

The authors thank the employees at the LSZ Boxberg for collecting tissue samples and for providing the trait data. PS received funding from the Carl-Zeiss-Stiftung, Stuttgart, Germany.

References

Black, B. L., and E. N. Olson (1998). Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annual review of cell and developmental biology*, **14**: 167–96.

Borchers, N., G. Otto, E. Kalm (2007). Genetic relationship of drip loss to further meat quality traits in purebred Piétrains. *Archiv für Tierzucht, Dummerstorf*. **50**: 84–91.

De Vries, A. G., P. G. van der Wal, T. Long, G. Eikelenboom, J. W. M. Merks (1994). Genetic parameters of pork quality and production traits in Yorkshire populations. *Livestock Production Science* **40**: 277–289.

Dombrowski, S. M., and D. Maglott (2002). Using the Map Viewer to Explore Genomes. In: J. McEntyre, J. Ostell, editors, The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 20. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21089/.

Franceschini, A., D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, et al. (2013). STRING V9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acid Research* **7**: 1.

Goddard, M. E., and B. J. Hayes (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Revisiom Genetics* **10**: 381–391.

Juszczuk-Kubiak, E., R. R. Starzyński, T. Sakowski, K. Wicińska, K. Flisikowski (2012). Effects of new polymorphisms in the bovine myocyte enhancer factor 2D (MEF2D) gene on the expression rates of the longissimus dorsi muscle. *Molecular Biology Reports* **39**: 8387–8393.

Lee, E. A., J. M. Kim, K. S. Lim, Y. C. Ryu, W. M. Jeon, K.C. Hong (2012). Effects of variation in porcine MYOD1 gene on muscle fiber characteristics, lean meat production, and meat quality traits. *Meat Science*. **92:** 36–43.

Muráni, E., M. Murániová, S. Ponsuksili, K. Schellander, K. Wimmers (2007). Identification of genes differentially expressed during prenatal development of skeletal muscle in two pig breeds differing in muscularity. *BMC Developmental Biology* **7:** 109–125.

Parker, M. H., P. Seale, M. Rudnicki (2003). Looking back to the embryo: defining transcriptional networks in adult myogenesis. *Nature reviews. Genetics*. **4**: 497–507.

Ponsuksili, S., E. Murani, M. Schwerin, K. Schellander, K. Wimmers (2010). Identification of expression QTL (eQTL) of genes expressed in porcine M. longissimus dorsi and associated with meat quality traits. *BMC Genomics*. **11**: 572-586.

Rockman, M. V., and L. Kruglyak (2006). Genetics of global gene expression. *Nature reviews. Genetics*. **7**: 862–872.

Schliekelman, P. (2008). Statistical power of expression quantitative trait loci for mapping of complex trait loci in natural populations. *Genetics.* **178**: 2201–16.

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. Cardoso, D. B. Edwards, C. W. Ernst (2011). Genome-wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs. *PloS One* **6**: e16766.

Storey, J.D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* **64**: 479-498.

Stratz, P., R. Wellmann, S. Preuss, K. Wimmers, Bennewitz J. (2014). Genome-wide association analysis for growth, muscularity and meat quality in Piétrain pigs. *Animal Genetics* DOI: 10.1111/age.12133.

Tajsharghi, H., E. Kimber, A. K. Kroksmark, R. Jerre, M. Tulinius, A. Oldfors (2008). Embryonic myosin heavy-chain mutations cause distal arthrogryposis and developmental myosin myopathy that persists postnatally. *Archives of neurology*. **65**: 1083–1090.

Untergasser, A., Cutcutache I., Koressaar T., Ye J., Faircloth B. C., Remm M., and Rozen S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic acids research*. **40**: e115.

Urbański, P., and Kurył J. (2004). New SNPs in the coding and 5' flanking regions of porcine MYOD1 (MYF3) and MYF5 genes. *Journal of Applied Genetics*. **45**: 325–329.

ZDS, Zentral Verband der Deutschen Schweineproduktion e. V. (2004). Richtlinie fuer die Stationspruefung auf Mastleistung, Schlachtkoerperwert und Fleischbeschaffenheit beim Schwein. Bonn, Germany.

Zhao, X., D. Mo, A. Li, W. Gong, S. Xiao, Y. Zhang, L. Qin, Y. Niu, Y. Guo, X. Liu, P. Cong, Z. He, C. Wang, J. Li, Y. Chen (2011). Comparative analyses by sequencing of transcriptomes during skeletal muscle development between pig breeds differing in muscle growth rate and fatness. *PloS One*. **6**: e19774.

Table 1 Overview of the SNP_*MYOD1* (GWAS, Stratz et al., 2014), its position and nucleotide exchange in the genome, positional genes in the vicinity and literature.

	polymorphism-ID	SSC	positional genes	literature
SNP_*MYOD1*	ASGA0010149:g. 47980126G>A	2	***MYOD1*** *SERGEF* *KCNC1* *TPH1*	Lee et al. (2012), Urbański and Kurył (2004)

Putative candidate genes are written in bold face.

Table 2 Summary statistic of the normalized expression of *MYOD1* (n=20)

trait	min	max	mean	median	std	CV (%)
Expression of *MYOD1*	0.22	2.74	0.90	0.55	0.69	1.26

Table 3 Pearson correlation coefficients (r^2) calculated between the gene expression of *MYOD1* and the traits, with comparison wise error probability and FDR *q-value* (*P-value* and *q-value*)

Merkmal	r^2	*P-value*	*q-value*
RSPL	**-0.53**	**0.02**	**0.14**
DG	**-0.53**	**0.02**	**0.14**
Bacon FOM	-0.29	0.23	0.95
LM FOM	0.23	0.35	0.95
pH2S	-0.22	0.37	0.95
pH1K	0.15	0.53	0.98
pH1S	0.08	0.73	0,98
BFTB	-0.06	0.80	0,98
con2S	0.03	0.89	0,98
BFTW	0.03	0.89	0,98
con2K	-0.03	0.90	0,98
Muscle FOM	-0.01	0.98	0,98
pH2K	0.01	0.98	0,98

Significant correlations (*q-value*<0.15) are written in bold face.

Table 4 Association analysis results for the traits recorded in Boxberg. The gene substitution effect ($\hat{\alpha}$), standard deviation (STD), *t-value*, *P-value* and FDR *q-value*

trait	$\hat{\alpha}$	STD	t-value	*P-value*	*q-value*
Bacon FOM	**-1.49**	**0.50**	**-2.98**	**0.01**	**0.13**
BFTB	**-0.20**	**0.08**	**-2.57**	**0.02**	**0.14**
pH2S	0.05	0.05	1.13	0.28	0.82
LM FOM	-0.20	0.19	-1.10	0.29	0.82
DG	-24.77	25.17	-0.98	0.34	0.82
con2K	0.04	0.04	0.96	0.35	0.82
BFTW	-0.11	0.13	-0.80	0.43	0.87
Muscle FOM	0.80	1.81	0.44	0.67	0.98
BFTL	-0.03	0.07	-0.37	0.72	0.98
con2S	0.04	0.48	0.09	0.93	0.98
pH2K	-0,00	0,12	-0.03	0.97	0.98
pH1K	-0,00	0,10	-0.03	0.98	0.98
pH1S	-0,00	0,10	-0.03	0.98	0.98

Significant gene substitution effects (*q-value*<0.15) are written in bold face.

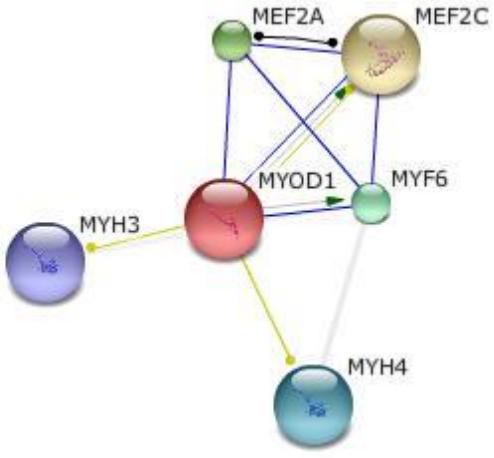


Figure 1 Protein-protein interaction networks for the gene products of *MYOD1*, blue lines represent bindings, yellow lines represent influences on the expression and black lines represent reactions. Green arrows symbolize an activating effect.

Appendix

Table S1 Traits recorded at the LSZ in Boxberg

trait	recording of the data
final weight	weght in kg at the day before slaughter
slaugher age	age upon completion of the part of examination.
pH1K	pH_1-cutlet: measured35-45min post mortem in the cutlet by 4cm deep insertion of the electrode between the 13th and 14th rip
pH1S	pH_1-ham: measured35-45min post mortem in the ham, 4-6cm above and 2-3 cm laterally to the pelvic bone
pH2K	pH_2-cutlet: measured 24h post mortem in the cutlet section
pH2S	pH_2-ham: measured 24h post mortem in ham, 4-6cm above and 2-3 cm laterally to the pelvic bone
con2K	conductivity 24h post mortem measured in the cutlet
con2S	conductivity 24h post mortem measured in the ham
LM FOM	Lean meat content estimated with the FOM-instrument
BFTW	Back fat thickness at the withers (thickest part)
BFTB	Back fat thickness at the middle of the back (thinest part)
BFTL	Back fat thickness at the loin (thinest part)
Muscle FOM	Muscle thickness measured with the FOM-instrument; tip of the FOM-probe has to penetrate between the 2nd and 3d last rip of the half-carcass 7cm laterally to the center line
Bacon FOM	Back fat thickness measured with the FOM-instrument; tip of the FOM-probe has to penetrate between the 2nd and 3d last rip of the half-carcass 7cm laterally to the center line
DG	average daily gain (in g) in the part of examination: gain in the part of examination (kg)/days in the part of examination x 1000

Table S2 Summary statistic of the dependent variables (n=20)

trait	min	max	mean	median	std	CV (%)
final weight	102	110.3	105.86	106	2.52	2.38
slaugher age	166	218	188.96	189.5	12.56	6.65
pH1K	5.78	6.9	6.31	6.36	0.28	4.44
pH1S	5.78	6.98	6.25	6.31	0.34	5.44
pH2K	5.28	5.86	5.61	5.61	0.13	2.32
pH2S	5.39	5.86	5.64	5.62	0.11	1.95
con2K	1.8	6.7	3	2.5	1.21	40.33
con2S	1.2	3.8	1.74	1.55	0.61	35.06
LM FOM	60.7	66.9	63.62	63.5	1.58	2.48
BFTW	1.58	2.91	2.29	2.34	0.37	16.16
BFTB	0.67	1.5	1.17	1.2	0.24	20.51
BFTL	0.22	1.01	0.5	0.49	0.21	42.00
Muscle FOM	54.1	71.5	62.01	61.45	5.05	8.14
Bacon FOM	6,7	12.6	8.8	8.45	1.6	18.18
DG	617	853	739.04	763.5	68.69	9.29

Table S3 Primer used for real-time qPCR with primer length, annealing temperature, GC-content (%), amplicon lengh and sequence

primer	primer-length	annealing temperature	GC-content	amplicon length	sequence
*MYOD1*_F	20	60,1°C	55	248	TTCGAGACTCTCAAGCGCTG
*MYOD1*_R	20	60,0°C	50		ATCATGCCGTCGGAACAGTT
*RPL32*_F	20	55°C	-	165	AGCCCAAGATCGTCAAAAAG
*RPL32*_R	20		-		TGTTGCTCCCATAACCAATG

Table S4 Primer for PCR based RFLP with primer length, annealing temperature, GC-content (%), amplicon length and sequence

primer	primer-length	annealing temperature	GC-content	amplicon length	sequence
SNP_*MYOD1*_F	20	58.3°C	50.00		TAACCCTCTGAGCCACAACT
SNP_*MYOD1*_R*	22	58.2°C	45.45	210	GCTAGTCAGATTCGTTTCCACT

*fluorescence marked primer

Table S5 Restriction enzyme with recognition site, temperature optimum for enzyme digestion and the resulting fragment length in bp after digestion

	amplicon-length	RFLP-length	enzyme	recognition site	temperature optimum
SNP_*MYOD1*	210	109	MspI	CCGG	37°C

Chapter four

Investigations on the pattern of linkage disequilibrium and selection signatures in the genomes of German Piétrain pigs

Patrick Stratz[1], Klaus Wimmers[2], Theo H. E. Meuwissen[3], and J. Bennewitz[1]

[1] Institute of Animal Husbandry and Breeding,
University of Hohenheim,
70599 Stuttgart, Germany
[b] Leibnitz Institute for Farm Animal Biology,
Research Unit Molecular Biology,
18196 Dummerstorf, Germany
[3] Department of Animal and Aquacultural Sciences,
Norwegian University of Life Sciences,
1432 Ås, Norway

Accepted for Publication in:
Journal of Animal Breeding and Genetics

Investigations on the pattern of linkage disequilibrium and selection signatures in the genomes of German Piétrain pigs

P. Stratz[1], K. Wimmers[2], T.H.E. Meuwissen[3], and J. Bennewitz[1]

[1] Institute of Animal Husbandry and Breeding, University of Hohenheim, Stuttgart, BW, Germany

[2] Leibnitz Institute for Farm Animal Biology, Research Unit Molecular Biology, Dummerstorf, MV, Germany

[3] Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Box, Norway

Correspondence

P. Stratz, Institute of Animal Husbandry and Breeding, University of Hohenheim, Garbenstrasse 17, D-70599 Stuttgart, BW, Germany. Tel: +49 711 459 22473; Fax: +49 711 459 23101; E-mail: Patrick.Stratz@uni-hohenheim.de

Summary

The aim of the present study was to study the population structure, to characterise the LD structure and to define core regions based on low recombination rates among SNP-pairs in the genome of Piétrain pigs using data from the PorcineSNP60 BeadChip. This breed is a European sire line and was strongly selected for lean meat content during the last decades. The data were used to map signatures of selection using the *REHH* test. In the first step, selection signatures were searched genome-wide using only core haplotypes having a frequency above 0.25. In the second step the results from the selection signature analysis were matched with the results from the recently conducted genome-wide association study for economical relevant traits in order to investigate putative overlaps of chromosomal regions. A small subdivision of the population with regard to the geographical origin of the individuals was observed. The extent of LD was determined genome-wide using r^2-values for SNP pairs with a distance 5Mb and was on average 0.34. This comparable low r^2 value indicates a high genetic diversity in the Piétrain population. Six *REHH* values having a *P-value*<0.001 were genome-wide detected. These were located on SSC1, 2, 6 and 17. Three positional candidate genes with potential biological roles were suggested, called *LOC100626459*, *LOC100626014* and *MIR1*. The results imply that for genome-wide analysis especially in this population a higher marker density and higher sample sizes are required. For a number of 9 SNPs, which were successfully annotated to core regions, the *REHH* test was applied. However no selection signatures were found for those regions (*P-value*<0.1).

Keywords: sire line; selective sweeps; single nucleotide polymorphism

Introduction

Improvement of growth, muscularity and meat quality was one of the main objectives in sire line pig breeding programs in Germany during the last three decades. Roughly since 1970, the breed Piétrain is frequently used as a sire line breed. The origin of the breed, which dates back to around 1920 in the village Piétrain in the province Brabant in Belgium, is not completely known. It is assumed that Piétrain descends from the French breed Bayeux, the English breeds Berkshire and eventually Yorkshire. The herd book was established in the 1950s. In the end of 1960, after the export of Piétrain to Germany, the Belgium Landrace and later sporadic Hampshires and Large White breeds were used in the breeding program.

Mapping genes or quantitative trait loci (QTL) has received considerable attention in pig breeding during the last decades. With regard to this, predominantly two genetically divergent founder breeds were crossed and developed to F2 outbred crosses (Rothschild et al., 2007). Due to its extreme muscularity and leanness the Piétrain breed was used in many studies to establish an F2-cross (Mohrmann et al., 2006; Liu et al., 2007; Rückert & Bennewitz, 2010). In these studies, the individuals were mainly genotyped with microsatellite markers and the F2-individuals were performance tested for the traits of interest. QTL mapping relied on linkage between marker and QTL, and variation within the founder breeds was not utilised. An alternative mapping approach is association mapping. It relies on linkage disequilibrium (LD) between marker and QTL and, in contrast to linkage analysis, utilises also historical meiosis (e.g. Goddard & Hayes, 2009). If association mapping is to be performed across the whole genome, genome-wide and dense marker maps have to be used. In pig breeding, the availability of the Illumina PorcineSNP60 BeadChip (Ramos et al., 2009) with around 62K SNPs offered new possibilities in genome-wide evaluations of the genome. In the German Piétrain breed it was used recently by Wellmann et al. (2013) for genomic selection and by Stratz et al. (2014) for association analysis. In the study of Stratz et al. (2014), around 500 progeny-tested sires from three different breeding organizations were genotyped with the PorcineSNP60 BeadChip. After data filtering around 48k SNPs were used to conduct genome-wide association studies for growth, muscularity and meat quality traits. However a thorough characterisation of the LD pattern using this SNP chip in the German Piétrains has not been done yet. Having knowledge about this pattern would help to design mapping and genomic selection experiments more efficiently.

The PorcineSNP60 BeadChip can also be used to find chromosomal regions with signatures of recent positive selection. Several methods are available to detect selective sweeps (e.g. Qanbari et al., 2010a; Akey, 2009; Nielsen, 2005). In livestock genetics the Extended Haplotype Homozygosity statistics (EHH, Sabeti et al., 2002) was applied by Qanbari et al. (2010a). It relies on the assumption that a mutation which is subject to positive selection will accumulate faster in frequency which results in longer LD blocks around the mutation. However this statistic is highly sensitive to relatedness between the individuals of the population. Unequally related individuals may cause a similar departure from the neutral model than selective sweeps and may introduce a bias if it is not corrected for (Schmid et al., 2005). Giving attention to the population structure may prevent potentially flawed results.

The aim of the present study was to study the population structure, to characterise the LD structure and to define core regions based on low recombination rates among SNP-pairs in the German Piétrain population using data from the PorcineSNP60 BeadChip. It can be speculated that sites in the genome at which we know selection has occurred (Stratz et al.2014) coincide with selective sweeps (Qanbari et al., 2010a). SNPs that were associated with economical relevant traits in recently conducted genome-wide association analysis were tried to get annotated to core regions to conduct a targeted search for selection signatures. The EHH test statistic was applied genome-wide and in targeted regions.

Materials and methods

Data preparation and haplotype reconstruction

A total of 849 German Piétrain boars were genotyped with the PorcineSNP60 BeadChip (Ramos et al., 2009). A proportion of 68%, 23% and 9% of the boars were originated from breeding organizations located in the states Baden-Württemberg, Nordrhein-Westfalen and Schleswig-Holstein, respectively. The same filtering criteria as in Stratz et al. (2014) were used to compare the results of both studies. Genotypes from individuals were filtered with respect to call rate (removal of SNPs with a call rate less than 95%), parent - progeny conflicts (removal of SNPs with parent - progeny conflict greater than 2%), MAF (exclusion of SNP with a minor allele frequency less than 3%) and significant deviation from the Hardy-Weinberg-equillibrium (*P-value*<0.0001). Animals with more than 10% missing genotypes were excluded. Filtering was done using PLINK (Purcell et al., 2007). Sporadic missing genotypes were imputed using fastPHASE (Scheet & Stephens, 2006). Markers which had an unknown physical position in the porcine genome sequence (Sus scrofa Build 10.2 assembly)

were excluded. Haplotypes were reconstructed using default parameters in fastPHASE (Scheet & Stephens, 2006).

Population structure

The structure of the population was visualized using two different methods, i.e. unscaled principle component (PC) analysis and multidimensional scaling analysis (MDS). For both purposes, the design matrix of the dimension (number of animals x number of SNPs), representing the number of animals and SNPs, was used. PCA was applied by inserting the design matrix in the function prcomp() of the R-package graphics. The identity-by-state matrix (IBS) was calculated form the design matrix (see Patterson et al., 2006). IBS-relationships were converted into genetic dissimilarities by subtracting them from one. For MDS the resulting IBS dissimilarity matrix was inserted in the function smacofsym() of the R-package smacof. This method visualizes genetic dissimilarities between objects. The boars were arranged in a 2-dimensional Euclidean space such that the distances between the boars correspond to their IBS dissimilarity as good as possible. This approximation is done by the objective function called stress (de Leeuw & Mair, 2009).

Extent of LD and core region partitioning

The most frequently used coefficients r^2 and D' which measure the extent of LD have different properties (Mueller, 2004). The LD coefficient r^2 is useful in the context of association studies. Calculating the confidence bounds on D' is of interest to assess the probability for historical recombination and to define regions as cores. Reconstructed haplotypes were inserted into Haploview v.4.1 (Barrett et al., 2005) to estimate r^2 and D' (Hill & Robertson, 1986) values. This was done for marker pairs being <5MB apart over the autosomes. Additionally r^2 was estimated for adjacent markers.

The core region consists of a core haplotype set and is characterized by a low historical recombination rate among SNPs. Core regions were identified using the block identification algorithm of Gabriel et al. (2002) which is implemented in Sweep v.1.1 (Sabeti et al., 2002). The algorithm relies on confidence bounds on D', rather than point estimates, because point estimates on D' tend to be upwards fluctuated when using a small number of samples or rare alleles (Gabriel et al., 2002). SNP-pairs were defined to be in strong LD if the upper 95% confidence bound on D' is between 0.7 and 0.98. This leads to different sized regions of markers being in strong LD. Regions were defined as core regions, if a very small proportion

(5%) of comparisons among SNP pairs had an upper confidence bound on D' smaller than 0.9. Gabriel et al. (2002) found that information from as few as three markers is sufficient to identify regions as cores.

It was assumed, that regions harbouring significant SNPs in GWAS for economically important traits may point to selection signatures. Therefore we tried to annotate significant SNPs which were found in the study of Stratz et al. (2014) to core regions.

Application of the EHH test

To determine regions under recent selection, the Extended Haplotype Homozygosity (*EHH*; Sabeti et al., 2002) test was performed. *EHH* is defined as the probability that two randomly chosen gametes carrying a particular core haplotype *t* are identical by descend for the entire interval from the core region to a distance x (Sabeti et al., 2002). It detects the transmission of an extended haplotype *i* without recombination. *EHH* of a tested core haplotype *t* is mathematically defined as:

$$EHH_t = \frac{\sum_{i=1}^{s} \binom{e_{ti}}{2}}{\binom{c_t}{2}}, \quad (1)$$

where c_t is the number of samples which carry a particular core haplotype *t* within a core region, e_{ti} is the number of samples having core haplotype *t* and carrying a particular extension *i*, and *s* is the number of particular extended haplotypes. Local recombination rates were not considered in the *EHH* test statistic. Therefore the Relative Extended Haplotype Homozygosity (*REHH*; Sabeti et al., 2002) test was applied. It is computed by $EHH_t / \overline{EHH}$ where $\overline{EHH}$ is calculated as:

$$\overline{EHH} = \frac{\sum_{j=1, j\neq t}^{n} \left[\sum_{i=1}^{s} \binom{e_{ji}}{2} \right]}{\sum_{j=1, j\neq t}^{n} \binom{c_j}{2}}, \quad (2)$$

where n is the total number of core haplotypes in a particular core region, and c_j is the number of samples having a core haplotype j (with j≠t). In contrast to equation (1), where the *EHH* decay for a particular core haplotype *t* was considered, equation (2) gives a measure for the *EHH* decay on all core haplotypes in the same core region combined except the core

haplotype *t*. Therefore the terms in the denominator and in the numerator were summed up over the number of particular core haplotypes in one core region.

To determine the empirical significance of *REHH* values, core haplotypes were placed in bins based on their frequency. To achieve normality, *REHH* values were log-transformed and their mean and standard deviation was calculated. Afterwards they were compared to all other equally frequent extended haplotypes within one bin and *P-values* were calculated. Core haplotypes with extreme *REHH* in the distribution were considered significant, as suggested by Sabeti et al. (2002). The *REHH* test was conducted genome-wide, taking core regions harbouring haplotypes above a certain frequency into account. Furthermore the *REHH* test was conducted for haplotypes, where significant SNPs, found in the study of Stratz et al. (2014), were successfully annotated to core regions. To judge how many false positives were among the significant associations, the false discovery rate (*FDR*) technique was applied. For each association test an *FDR q-value* was calculated, using the software QVALUE (Storey 2002; Storey & Tibshirani 2003).

Mapping of selective sweeps to genome annotations

The position of SNPs within each core region was aligned to the porcine genome sequence (Sus scrofa Build 10.2 assembly). The NCBI map viewer (Dombrowski & Maglott 2002) was used to identify positional genes in a 1MB region upstream and downstream of the cores. To confirm their role as candidates, only genes being involved in metabolisms which determine traits included in the breeding goal were searched.

Results

Data description

A descriptive summary of the genome-wide marker distribution is presented (Table S1). Totally 47549 markers passed the filtering criteria and were included in the final analysis. This subset covers 2983 Mb of the genome with 84.6 228.9 mean distance (kb) between adjacent markers, whereof the largest gap between SNPs (4.989 Mb) was located on SSC8. The average observed heterozygosity and mean MAF were estimated as 0.36 ± 0.13 and 0.27 ± 0.14, respectively.

Population structure

The first two PCs (Figure1, left side) accounted for 2.2%, respectively 1.7% of the variance in the data. On the right side (Figure 1), boars were arranged such that the distance between them corresponds to their genetic dissimilarity as much as possible. In MDS the stress value was 0.16. It can be shown, that independent of the method which was used for studying the structure of the population, a small substructure with regard to the origin of the boars is observed. Boars from breeding organizations located in the states Nordrhein-Westfalen and Schleswig-Holstein are genetically more similar than boars from Baden-Württemberg.

LD and core region structure

All possible SNP pairs with a distance 5Mb produced 3284064 pairwise LD values on the 18 porcine autosomes. The average D' (r^2) between SNP pairs was 0.43 (0.34). Average D' (r^2) at various distances were computed by grouping all SNP combinations by their pairwise distance in classes of 50kb of length starting at 0 to 5MB. In Figure S1 the decline is shown for D' over distance. Figure 2 displays an overview of the decline of r^2 over distance. For all inter-marker distances the average D' was higher than the average r^2. Additionally the r^2 decay for all adjacent marker pairs is shown in Figure S2 over the autosomes and in Figure S3 for every autosome, respectively. The average r^2 for adjacent SNPs estimated over the autosomes was 0.55. On a chromosome level, the mean distance between adjacent markers is in between 51,7 kb on SSC4 and 165.1 kb on SSC15. The estimated mean r^2 between adjacent markers was in between 0.19 on SSC12 and 0.94 on SSC14 (Table S1).

A descriptive summary of the core region distribution in this dataset is presented in Table 1. A total of 5700 core regions spanning 954.010 Mb (32%) of the entire genome and containing 22854 core haplotypes were detected; most of them on SSC1. Mean core region length was estimated as 167.4 275.9 kb respectively with a maximum of 5052 kb on SSC6. The proportion of total core region lengths relative to autosome length was lowest for SSC10, carrying 230 core regions spanning 19.438 Mb of the chromosome, and highest for SSC1, carrying 820 core regions, spanning 196.484 Mb of the chromosome. 23092 SNPs (49%) participated in forming the core regions. Mean number of core region building SNPs was estimated as 4.46 3.03 with a maximum of 19 SNPs on chromosomes 1, 4-9, 11, 13-15.

Autosomal screen to identify targets of strong recent selection

8966 core haplotypes having a frequency above 0.25 were arranged in 15 core haplotype frequency bins (0.25-0.30, 0.30-0.35, …, 0.95-1.00) and *P-values* of the normalized *REHH* values were calculated. The distribution of –log10(*P-value*) is shown as box plots within each bin (Figure 3). The test statistic of 8966 core haplotypes, with a frequency above 0.25, over the chromosomes is shown (Figure S4). The threshold for outliners displayed in both Figures was set to 0.01 and 0.001. A number of 918 normalized *REHH* values were detected having a *P-value*<0.1 whereof 69 are showing a *P-value*<0.01 and 6 are showing a *P-value*<0.001 (Table 2). Normalized *REHH* values having *a P-value*<0.001 were found on SSC1, 2, 6 and 17 (Table 2). It is evident, that the outliers appear in medium frequency bins (0.30-0.45).

It was possible to identify positional genes which meets the requirements of candidates, because they are involved in metabolisms which determine traits included in the breeding goal. The positional candidate genes are listed in Table 2. The genes *LOC100626459* and *LOC100626014* (calcium/calmodulin-dependent protein kinase type II subunit alpha-like) located on SSC2 are involved in Ca^{2+} signalling. The Ca^{2+}-metabolism is important for meat quality, growth and muscling traits (Fujii et al., 1991). *CaMK2* is an enzyme which is composed of four different chains: alpha, beta, gamma and delta and belongs to the serine/threonine protein kinase family and the Ca^{2+}/calmodulin-dependent protein kinase subfamily. CaMk2A is coding for the alpha chain of this enzyme. MacDonnell et al. (2009) showed that in cardiac myocytes of neonatal rats the activation of cytoplasmic CaMKII inhibits antiapoptotic and hypertrophic target genes. Another gene which was found on SSC17 is *MIR1* (microRNA mir-1). MicroRNAs represent a group of small noncoding RNAs with regulatory functions. Hong et al. (2012) investigated polymorphisms in the porcine microRNA *mir-1* in three pig breeds, called Berkshire, Landrace and Yorkshire. The authors found an association for polymorphisms on muscle fiber type composition and mir-1 expression.

REHH test in candidate regions

It was not always possible to annotate significant SNPs, which were found in Stratz et al. (2014), to core regions. The list of SNPs and their associated trait names, which could be arranged in core regions, is summarized (Table S2). For daily gain, carcass lean content, pH1, and drip loss, it was possible to locate SNPs in core regions. The second most frequent haplotypes and their test statistic downstream and upstream of the core are listed in Table S2.

For the SNP ASGA0094502, which was found to be significant for daily gain on SSC1 in Stratz et al. (2014), the test statistic was only available for the downstream region. This is attributable to the lower LD upstream of the core (not shown). For three core haplotypes, being extended in one direction from the core, a *P-value*<0.1 for *REHH* was obtained (marked in bold, Table S2). The FDR among the significant *REHH* tests was 0.8.

Discussion

Population structure

The results of both methods (Figure 1) showed that there is only a small tendency for population structure of individuals with respect to geographical origin. Hence, no specific modelling of population structure beyond taking the relationship of the individuals into account is needed in applying GWAS (Stratz et al., 2014) and genomic selection (Wellmann et al., 2013). In general the genetic distance between individuals was large. This indicates that the individuals are less related. Therefore correction for population structure is not required before searching for selective sweeps.

Extent of LD

Care needs to be taken when comparing levels of LD over studies, because of different sample sizes, LD measures, marker densities and sets of markers (Pritchard & Przeworski, 2001). Therefore only studies reporting LD for the PorcineSNP60 BeadChip data were used for comparison with the results in our study. The LD coefficient r^2 is the most relevant measurement for association mapping, because of the simple relationship with the sample size required to detect an association between a trait and the marker loci. Therefore the emphasis of this section was on the comparison of the LD coefficient r^2 estimated in this studies with the r^2 coefficients estimated in other studies. For Finish Landrace and Yorkshire average r^2-values for marker pairs being <5MB apart of 0.43 and 0.46, respectively, were reported by Uimari and Tapio (2011). Veroneze et al. (2013) found an average r2 of 0.39 for a purebred Piétrain sire line, which was lowest, compared to the other lines representing synthetics and pure lines from Netherlands.

Compared to those studies the value of r^2 found in this study (0.34) is low. This points to a high genetic diversity within this breed. This is also in agreement with the estimated genetic distances between the individuals in our study and with the large effective population size of this breed (Ne>200, BLE 2010). The past effective population size from SNP data was not estimated, because of the introgression of several breeds in the Piétrain breed since its origin.

The high effective population size, as a result of the introgression of several breeds, may lead to new alleles with low-frequency which could not be separated from the ones that already exists as low-frequency alleles in the population. To avoid the confounding effect of population demographic history only alleles having a certain minor allele-frequency (MAF>3%) were used to define core regions.

Badke et al. (2012) used four US pig breeds and estimated adjacent r^2-values between 0.36 and 0.46. Compared to the study of Badke et al. (2012) the value of r^2 found in this study (0.55) is high. Although having a relative high r^2 between adjacent marker pairs, Stratz et al. (2014) observed a low number of significant SNPs for growth, muscularity and meat quality traits. There are two possible explanations for this. Either, only alleles having a low impact in trait determination are segregating in the population or that on chromosome with low r^2 the experimental power is not high enough to detect true associations. Reducing the distance between adjacent markers for chromosomes with low r^2 would then be helpful to increase r^2, and hence, increasing the experimental power in GWAS.

Pattern of LD

The average estimated core length was 167.4kb. The size of the core regions is comparable to the length in the study of Qanbari et al. (2010b), who used 40854 SNP covering the whole bovine genome. For the pig genome, Amaral et al. (2008) detected larger core regions. Veroneze et al. (2013) detected larger core regions in a purebred Piétrain sire line from the Netherlands. Both authors showed that LD is mostly organized in cores up to 400kb. The number of core regions, the small core region size as well as the low number of core-building SNPs in this study are in agreement with the low level of r^2 for all possible marker pairs. Considering the whole genome, the proportion of total number of SNPs forming core regions on number of SNPs used is higher compared to the proportion of total core region length on chromosome length.

Only few significant selection signatures were found (*REHH P-value*<0.001). There are several possible explanations for this. The power of the study design might not be high enough to detect existing signatures. Alternatively, selection has acted on the whole genome rather than on single chromosomal regions which might be attributable to the complexity of the genetic architecture of the traits under selection. A large number of loci with very small effects on the traits were slowly accumulated in frequency. This is a possible explanation why

selection signatures were found in core haplotypes of medium frequency, although it seems to be more likely that a higher core haplotype frequency is an indicator of positive selection. A third possibility is that intensive selection was carried out only for the last few generations in the population and the time to generate extensive haplotype homozygosity around the target genes was too short.

Significant SNPs found by Stratz et al. (2014) were matched to core regions. This was only partially possible and strong evidence for recent selection was not found in those regions. Maybe selective sweeps found in our study might belong to other traits which were not considered in the study of Stratz et al. (2014), or are under natural selection. This in-accordance was also found by Qanbari et al. (2010a) in a dairy cattle population and by Kemper et al. (2014) in eight domestic Bos taurus cattle belonging to dairy and beef breeds. Kemper et al. (2014) studied sites in the genome at which they know selection has occurred to see if a signature of selection has been left behind. The authors studied three types of loci with different effects on the traits. Two statistics that indicate selection signatures within a breed were applied. For qualitative traits they were able to show clear signatures of selection within breeds because selection pressure has applied to one single locus. However weak evidence for selection signatures was found at regions in the genome associated with complex traits under selection.

Conclusions

The population structure indicates that there is no specific modelling of the geographic origin of the animals in GWAS or genomic selection necessary. The low r^2 for all possible marker pairs estimated in this study is in accordance with the breeding history of the population and indicates a high diversity in the German Piétrain pig population. Only few significant selection signatures could be found genome-wide. Results from recently conducted GWAS for economical relevant traits did not match with regions of increased extended haplotype homozygosity.

Acknowledgements

P. Stratz received funding from the Carl-Zeiss-Stiftung, Stuttgart, Germany. His stay at the Norwegian University of Life Science was supported by the German Academic Exchange Service (DAAD). The manuscript has benefited from the critical and helpful comments of three anonymous reviewers.

References

Akey, J.M. (2009). Constructing genomic maps of positive selection in humans: Where do we go from here? *Genome Research* **19**: 711–722.

Amaral, A. J., Megens, H.J., Crooijmans, R.P.M.A., Heuven, H.C.M., Groenen, M.A.M. (2008). Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* **179**: 569-579.

Badke, Y. M., R. O. Bates, C. W. Ernst, C. Schwab, and J. P. Steibel (2012). Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics* **13**: 24.

Barrett J.C., Fry B., Maller J., Daly M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265. BLE, Bundesanstalt fuer Landwirtschaft und Ernaehrung (2010). Erhebung von Populationsdaten tiergenetischer Ressourcen in Deutschland: Schweine. http://download.ble.de/06BE001.pdf Bonn, Germany.

BLE, Bundesanstalt fuer Landwirtschaft und Ernaehrung (2010) Erhebung von Populationsdaten Tiergenetischer Ressourcen in Deutschland: Schweine. http://download.ble.de/06BE001.pdf, Bonn, Germany.

Dombrowski, S. M., D. Maglott (2002). Using the Map Viewer to Explore Genomes. In: McEntyre, J., Ostell, J. (ed.), *The NCBI Handbook [Internet]*. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 20. Pp.1-9. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21089/.

Fujii, J., K. Otsu, F. Zorzato, S. DeLeon, V. K. Khanna, J. E. Weiler, P. J. O'Brien, D. H. McLennan (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253**: 448–451.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S. N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E. S. Lander, M. J. Daly, D. Altshuler (2002). The Structure of Haplotype Blocks in the Human Genome. *Science* **23**: 2225-2229.

Goddard, M. E., and B. J. Hayes, (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10**: 381–391.

Hill, W.G., Robertson, A. (1986) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, **38**, 226-231.

Hong, J.S., Noh, S.H., Lee, J.S., Kim, J.M., Hong, K.C., Lee, Y.S. (2012) Effects of polymorphisms in the porcine microRNA miR-1 locus on muscle fiber type composition and miR-1 expression. *Gene*, **506**,211–216.

Kemper, K. E., Saxton, S. J., Bolormaa, S., Hayes B. J., Goddard M. E. (2014) Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics*, **15**, 246.

de Leeuw J., Mair P. (2009) Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, **31**, 1–30.

Liu, G., D. G. J. Jennen, E. Tholen, H. Jüngst, T. Kleinwächter, M. Hölker, D. Tesfaye, C. Ün, H. J. Schreinemachers, E. Murani, S. Ponsuksili, J. J. Kim, K. Schellander, K. Wimmers (2007). A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc–Piétrain resource population. *Animal Genetics* **38**: 241–252.

MacDonnell, S. M., J. Weisser-Thomas, H. Kubo, M. Hanscome, Q. Liu, N. Jaleel, R. Berretta, X. Chen, J. H. Brown, A. K. Sabri, J. D. Molkentin, S. R. Houser (2009). CaMKII negatively regu-lates calcineurin-NFAT signaling in cardiac myocytes. *Circulation Research* **105**: 316–325.

Mohrmann, M., R. Roehe, P. W. Knap, H. Looft, G. S. Plastow, E. Kalm (2006). Quantitative trait loci associated with AutoFOM grading characteristics, carcass cuts and chemical body composition during growth of Sus scrofa. *Animal Genetics* **37**: 435–43.

Mueller, J.C. (2004) Linkage disequilibrium for different scales and applications. *Briefings in Bioinformatics*, **5**, 355-364.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, **39**: 197–218.

Patterson N., Price A., Reich D. (2006)Population structure and eigenanalysis. PLoS Genet., 2, e190.Pritchard, J. K., and M. Przeworski (2001). Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics* **69**, 1–14.

Pritchard J.K., Przeworski M. (2001) Linkage disequilibrium in humans: models and data. *The American Journal of Human Genetics*, **69**, 1–14.

Purcell S., Neale B., Todd-Brown, Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., de Bakker P.I., Daly M.J., Sham P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**, 559–575.

Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi, H. Simianer, (2010a). A Genome-Wide Scan for Signatures of Recent Selection in Holstein Cattle. *Animal Genetics* **44**: 377-389.

Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi, H. Simianer (2010b). The pattern of linkage disequilibrium in German Holstein cattle. *Animal*

Genetics **41**: 346-356.

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. V. Tassell, J. F. Taylor, R. T. Wiedmann, L. B. Schook, 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.

Rothschild, M. F., Z. L. Hu, Z. Jiang (2007). Advances in QTl mapping in pigs. *International Journal of Biological Sciences* **3**: 192-197.

Rückert, C., and J. Bennewitz (2010). Joint QTL analysis of three connected F2-crosses in pigs. *Genetics Selection Evolution* **42**: 40.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson., G. J. McDonald, H. C. Ackerman, S. J. Campbell, D. Altshuler, R. Cooper, D. Kwiatkowski, R. Ward, E. S. Lander (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832–837.

Scheet, P., and M. Stephens (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics* **78**: 629–644.

Schmid, K. J., Ramos-Onsins, S., Ringys-Beckstein, H., Weisshaar, B., Mitchell-Olds, T. (2005) A multilocus sequence survey in Arabidopsis thaliana reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics*, **169**, 1601–1615.

Storey J.D. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 479–98.

Storey J.D., Tibshirani R. (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, **100**, 9440–5.

Stratz, P., R. Wellmann, S. Preuss, K. Wimmers, J. Bennewitz (2014). Genome-wide association analysis for growth, muscularity and meat quality in Piétrain pigs. *Animal Genetics*, **45**, 350–356.

Uimari, P., and M. Tapio (2011). Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *Journal of Animal Science*

89: 609-614.

Veroneze R., Lopes P.S., Guimarães S.E., Silva F.F., Lopes M.S., Harlizius B., Knol E.F. (2013) Linkage disequilibrium and haplotype block structure in six commercial pig lines. *Journal of Animal Science*, **91**, 3493-501.

Wellmann, R., S. Preuß, E. Tholen, J. Heinkel, K. Wimmers, J. Bennewitz (2013). Genomic selection using low density marker panels with application to a sire line in pigs. *Genetics Selection Evolution* **45**: 28.

Table 1: Summary of core region (CR) distribution across the autosomes in the Piétrain pig genome.

Chr	No. CR (n)	Mean CR length (kb) ± SD	Coverage CR length[a] (Mb)	Max CR length (kb)	CR length/Chr length[b]	CR SNP[c](n)	Mean CR SNP (n) ± SD	Max CR SNPs (n)	CR SNPs/SNP[d]
1	820	239.6 ± 329.9	196.484	2103.0	0.52	4009	5.2 ± 3.8	19	0.71
2	424	123.8 ± 170.7	52.486	1960.0	0.32	1559	4.1 ± 2.4	17	0.51
3	363	119.1 ± 186.7	43.244	1446.9	0.28	1246	3.9 ± 2.1	16	0.47
4	434	146.9 ± 187.5	63.767	1285.1	0.46	1793	4.6 ± 3.4	19	0.57
5	330	114.3 ± 199.0	37.722	1929.1	0.34	1244	4.2 ± 2.3	19	0.54
6	253	147.0 ± 380.8	37.181	5052.0	0.19	985	4.3 ± 2.7	19	0.29
7	420	188.8 ± 283.4	79.305	2532.8	0.57	1845	4.8 ± 3.2	19	0.62
8	296	166.1 ± 317.7	49.158	3023.0	0.29	1210	4.5 ± 3.1	19	0.59
9	306	89.4 ± 146.8	27.360	1269.6	0.16	1071	3.9 ± 2.1	19	0.36
10	230	84.5 ± 129.6	19.438	1081.7	0.20	733	3.7 ± 1.7	13	0.42
11	182	170.5 ± 268.1	31.039	1954.5	0.35	644	4.0 ± 2.6	19	0.37
12	218	113.6 ± 226.0	24.757	2023.0	0.29	681	3.6 ± 1.6	13	0.43
13	284	233.4 ± 287.2	66.290	1604.0	0.25	1412	5.3 ± 3.9	19	0.38
14	237	361.3 ± 515.9	85.635	2796.7	0.52	1467	6.5 ± 5.0	19	0.42
15	218	201.4 ± 249.7	43.911	1508.5	0.15	913	4.6 ± 3.4	19	0.36
16	251	101.6 ± 240.4	25.507	1938.8	0.20	789	3.6 ± 1.4	11	0.45
17	257	173.1 ± 237.4	44.480	1351.9	0.49	901	3.9 ± 1.8	13	0.57
18	177	148.3 ± 176.2	26.244	786.2	0.30	590	3.8 ± 1.8	13	0.54
Total	5700	167.4 ± 275.9	954.010	5052.0	0.32	23092	4.46 ± 3.03	19	0.49

[a] Total length covered by core regions; [b] The proportion of total core region lengths on autosome length; [c] Number of SNPs forming core regions; [d] The proportion of total number of SNPs forming core regions on number of SNPs used.

Table 2 Summary statistics for core haplotypes with frequency above 0.25 and showing the lowest *P-values* in the *REHH* test.

SSC	Core Position (Mb)	Hap freq (%)	*EHH*	*REHH*	*REHH P-value*	positional candidate genes[a]
2	158.375-158.375	25	0.21	19.43	0.000009	*LOC100626459, LOC100626014*
2	101.578-101.611	32	0.06	8.73	0.000459	-
1	104.906-104.942	34	0.82	8.58	0.000505	-
1	104.983-105.005	37	0.68	7.59	0.000731	-
6	55.378-55.397	47	0.39	7.34	0.000786	-
17	78.982-78.995	42	0.87	7.06	0.00082	*MIR1*

[a] Positional candidate genes with potential biological roles that could be found in a distance of 1Mb upstream and downstream of the core regions.

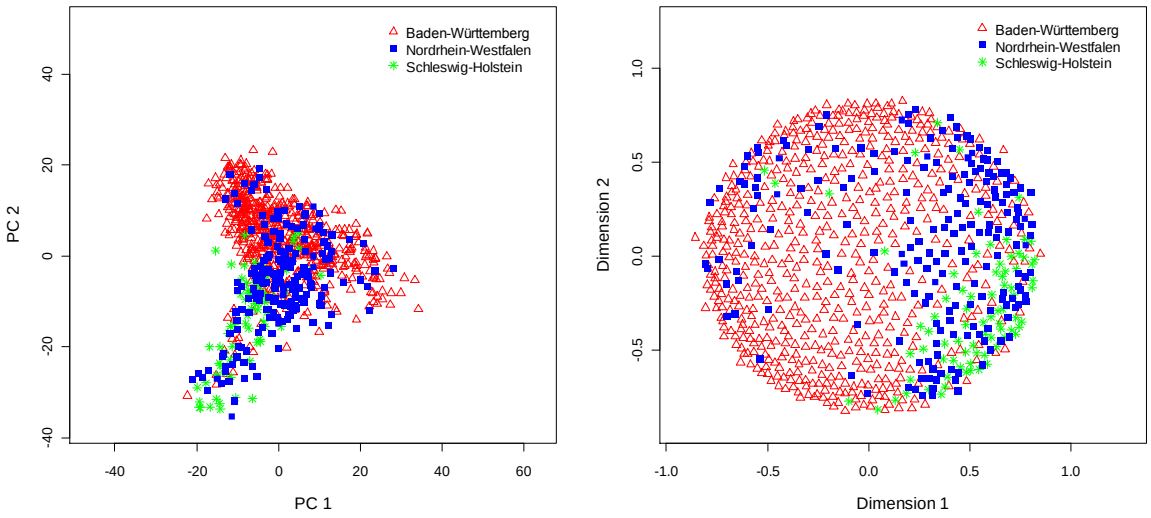


Figure 1 Clustering the Piétrain population based on SNP genotype data. Boars belonging to the 3 breeding organisations (Baden-Württemberg, Nordrhein-Westfalen, and Schleswig Holstein) are shown with different symbols. On the left side the first two principle components are shown, whereas on the right side the results of multidimensional scaling are shown in a 2-dimensional Euclidean space.

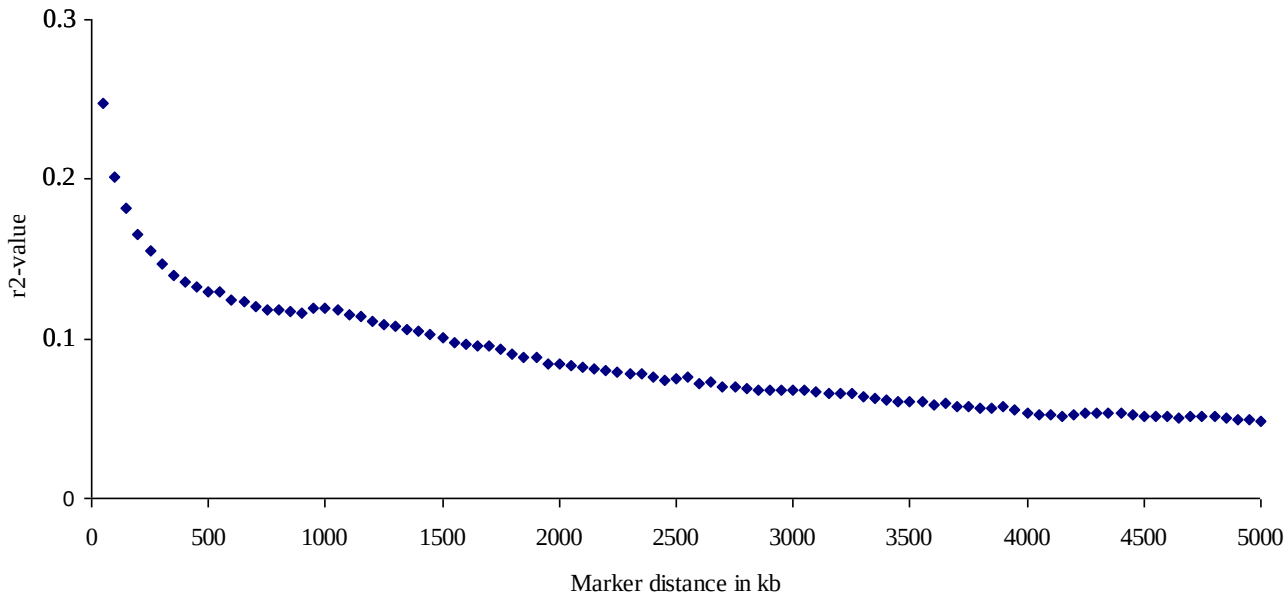


Figure 2 Decay of average LD over distance. The x-axis denotes marker distance in kb and the y-axis the LD for all possible marker pairs separated by less than 5 MB using r^2.

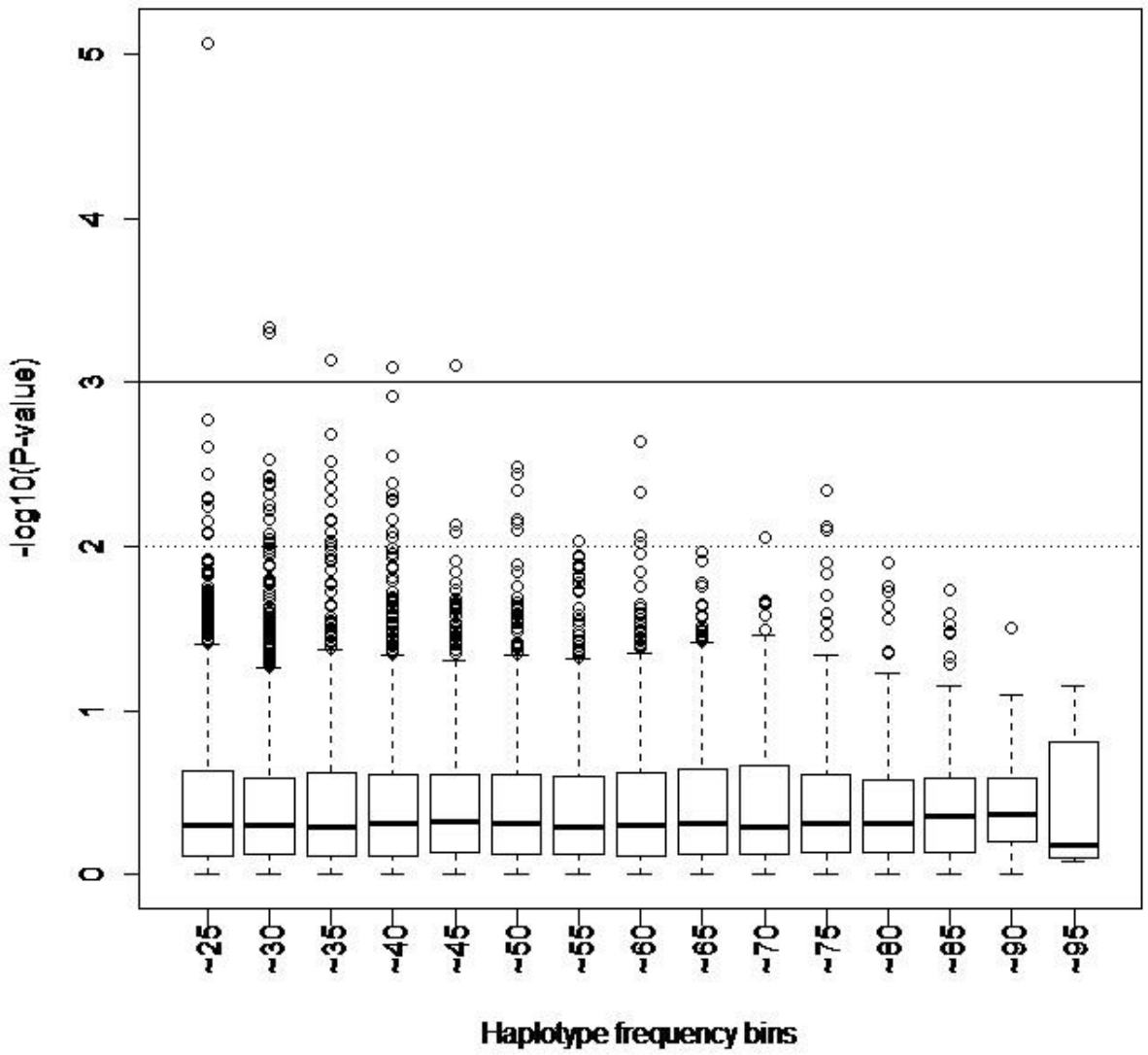


Figure 3: Box plots of the distribution of logarithmic *REHH P-value*s in core haplotype frequency bins of 5% difference. The x-axis denotes core haplotype frequency bin and the y-axis –log10(*P-value*). Only core haplotypes with a frequency above 0.25 are shown. Two significance levels are indicated, *P-value*=0.01 (dashed line) and *P-value*=0.001 (continuous line).

Appendix

Table S1 Summary of marker distribution across the autosomes in the Piétrain pig genome.

Chr	SNP (n)	Chr length (Mb)	Mean dist. (kb) adjacent SNPs	Max gap (Mb) adjacent SNPs	Mean r^2 adjacent SNPs
1	5658	376.1	69.0	4.302	0.87
2	3071	162.6	83.7	3.319	0.55
3	2661	152.1	89.0	3.530	0.43
4	3157	140.1	51.7	3.099	0.76
5	2323	112.4	66.3	3.220	0.56
6	3437	190.9	127.6	4.697	0.32
7	2980	138.8	54.7	1.834	0.71
8	2062	169.5	128.2	4.989	0.55
9	2935	172.7	80.7	4.949	0.49
10	1726	95.6	97.4	4.683	0.26
11	1764	89.2	75.7	3.145	0.59
12	1586	84.6	99.8	2.505	0.19
13	3727	264.3	123.0	3.192	0.64
14	3509	166.2	62.7	3.192	0.94
15	2546	292.4	165.1	4.430	0.58
16	1742	129.0	82.1	3.785	0.33
17	1582	91.3	72.3	1.390	0.54
18	1083	87.7	117.9	4.510	0.51
Total	47549	2983	84.6	4.989	0.55

Table S2 Summary statistics for the *REHH* test for selection signatures in candidate regions.

Results from GWAS[a]			Core position (Mb)	Hap freq (%)	*EHH*	*REHH*	*REHH P-value*
Trait[b]	Marker	SSC					
CLC	ASGA0084324	5	12.416	H1:62	0.09/0.13	0.89/0.72	0.30/0.43
	ALGA0121463		-12.416	H2:37	0.10/0.18	1.11/1.39	0.38/0.58
CLC	MARC0084792	13	31.615-	H1:30	0.30/0.35	0.69/0.45	0.71/0.89
			32.038	H2:19	0.54/0.79	1.60/1.41	0.32/0.40
DG	ASGA0094502	1	42.770-	H1:25	-/0.51	-/0.85	-/0.65
			42.770	H2:24	-/0.46	-/0.72	-/0.75
DG	ASGA0010202	2	47.986-	H1:41	0.01/0.17	0.27/0.28	0.96/0.95
			48.383	H2:31	0.02/0.58	0.90/2.55	0.56/**0.08**
DG	ALGA0032718	5	67.410-	H1:64	0.00/0.17	0.20/0.42	0.96/0.73
			67.410	H2:28	0.01/0.40	3.44/2.37	**0.04**/0.13
DG	ALGA0087665	15	97.072-	H1:63	0.31/0.40	0.60/0.71	0.53/0.43
			97.101	H2:21	0.31/0.43	0.89/1.03	0.64/0.55
Drip	ASGA0103690	16	17.875-	H1:32	0.50/0.77	0.84/2.65	0.60/**0.07**
			17.892	H2:30	0.27/0.21	0.40/0.35	0.92/0.94
pH1	ASGA0025473	5	53.889-	H1:29	0.80/1.00	0.96/137	0.58/0.37
			53.889	H2:27	0.50/0.54	0.57/0.65	0.89/0.90

[a] significant SNPs (Stratz el al., in press) annotated to core haplotypes; [b] carcass lean content (CLC), daily gain (DG), drip loss (Drip), and pH1 (measured 45 min post-mortem in cutlet); *EHH*, *REHH* and *REHH P-value* presented for upstream, downstream regions of the second most frequent haplotypes in the cores, respectively.

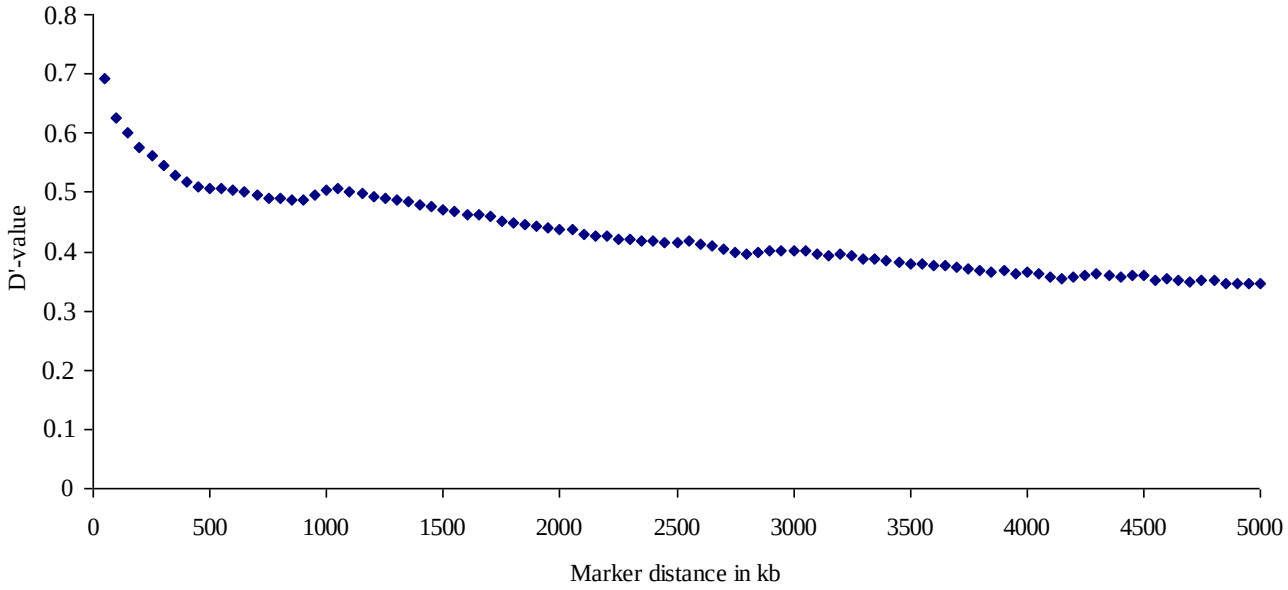


Figure S1 Decay of average LD over distance. The x-axis denotes marker distance in kb and the y-axis the LD for all possible marker pairs separated by less than 5 MB using D'.

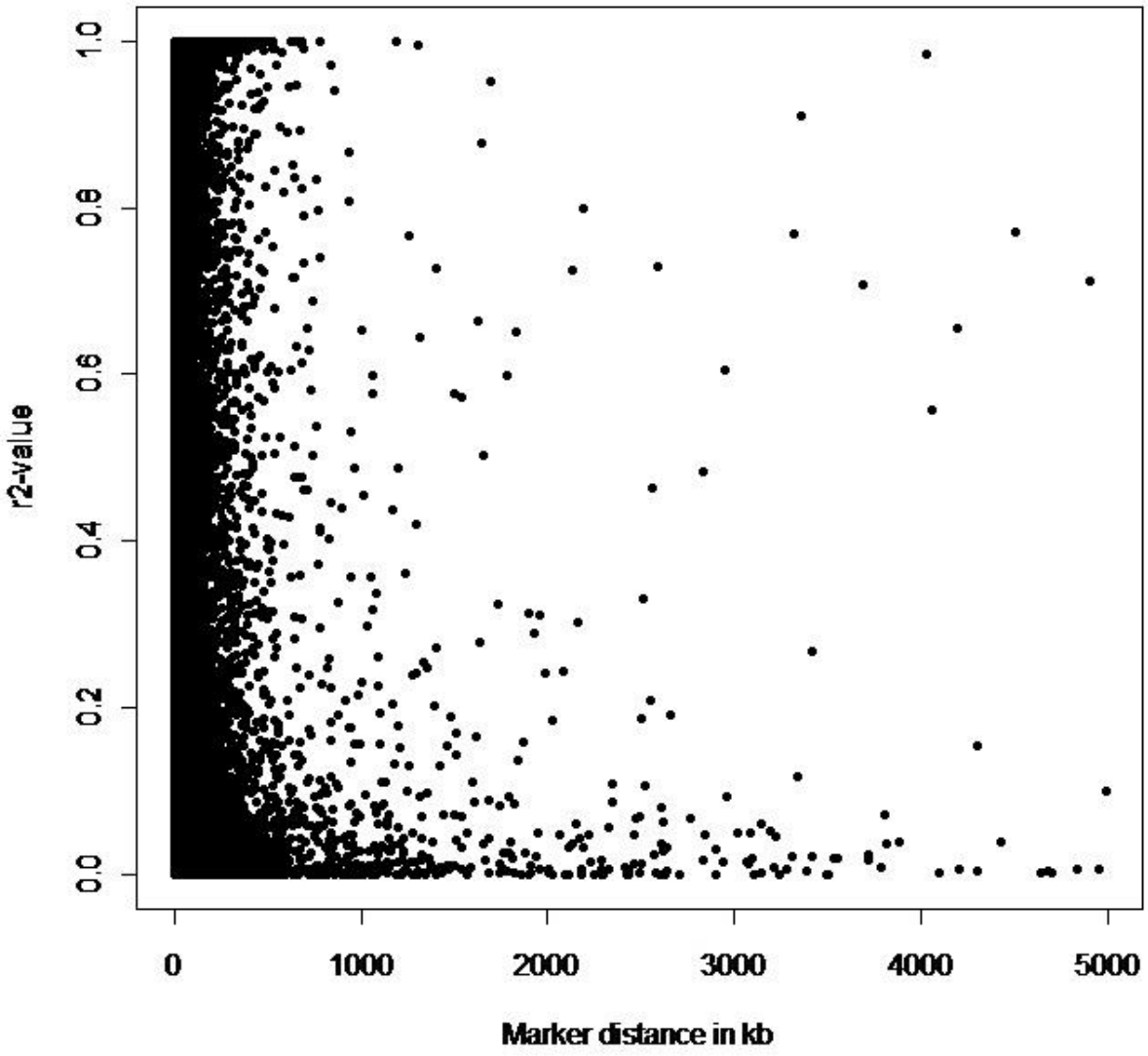


Figure S2 Pairwise LD values for adjacent markers for all autosome. The x-axis denotes marker distance in kb and the y-axis the r^2 for adjacent marker pairs.

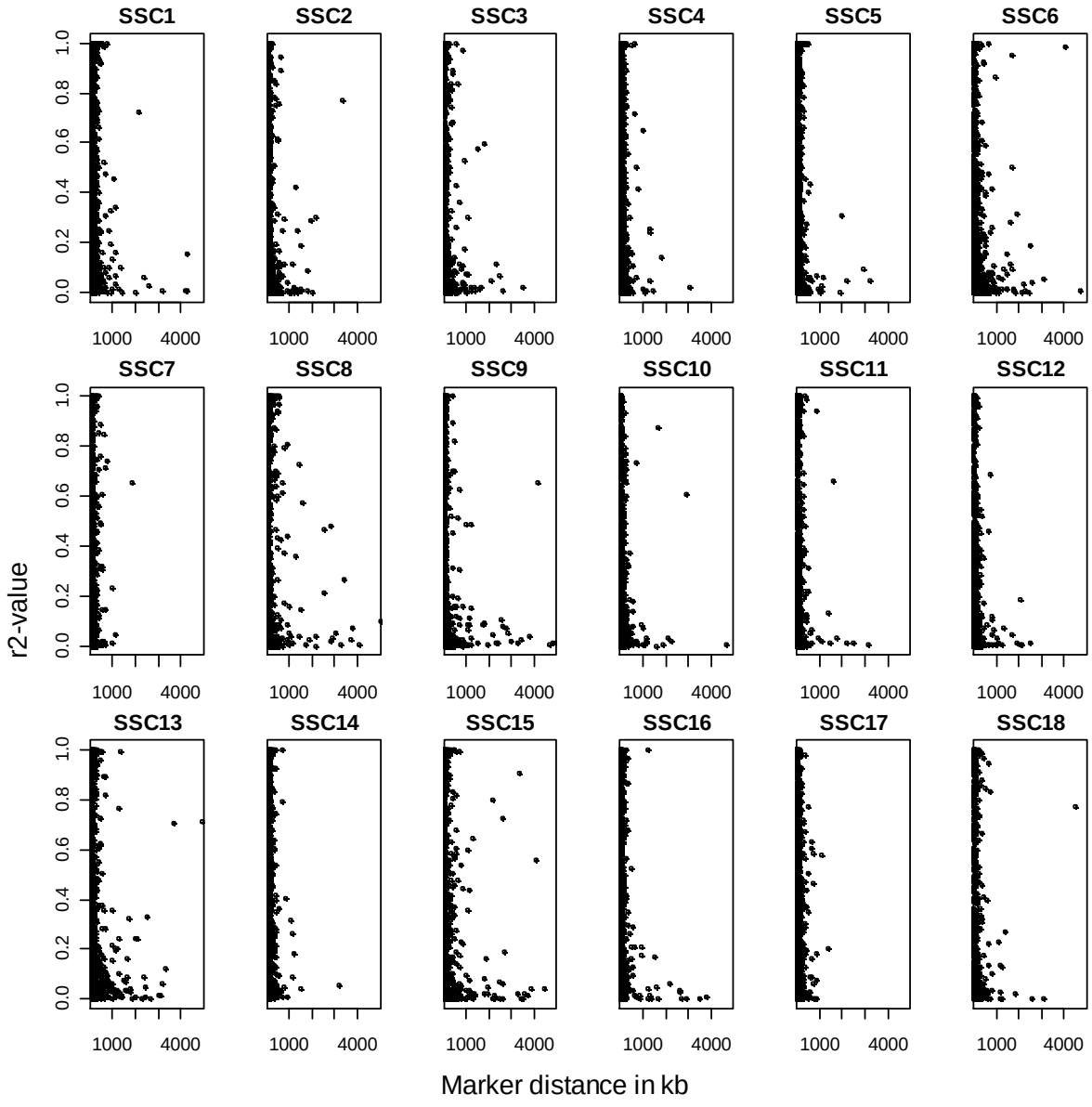


Figure S3 Pairwise LD values for adjacent markers for every autosome. The x-axis denotes marker distance in kb and the y-axis the r^2 for adjacent marker pairs.

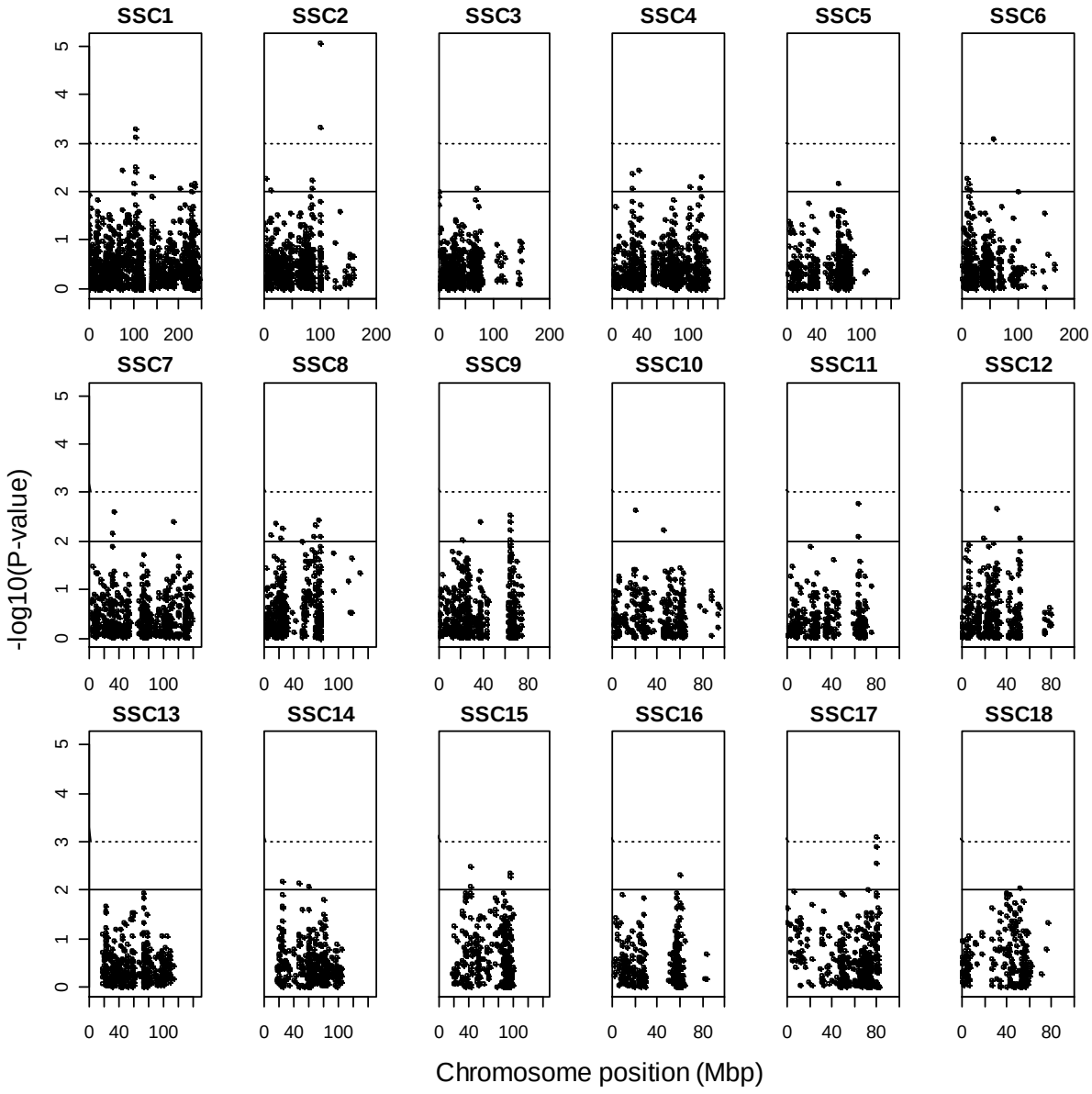


Figure S4 Autosomal maps of *REHH P-values* for core haplotypes with a frequency above 0.25. The x-axis denotes chromosome position (Mbp) and the y-axis –log10(*P-value*). For each autosome, two significance levels are indicated, *P-value*=0.01 (dashed line) and *P-value*=0.001 (continuous line).

Chapter five

Strategies to implement genomic selection in pig breeding using very low marker density

Patrick Stratz, Robin Wellmann,
and Jörn Bennewitz

Institute of Animal Husbandry and Breeding,
University of Hohenheim,
70599 Stuttgart, Germany

Accepted for Publication in:
10th World Congress of Genetics Applied to Livestock Production

Strategies to implement genomic selection in pig breeding using very low marker density

P. Stratz, R. Wellmann and J. Bennewitz

Institute of Animal Husbandry and Breeding, University of Hohenheim, D-70599 Stuttgart, Germany

Email addresses:
PS: Patrick.Stratz@uni-hohenheim.de
RW: r.wellmann@uni-hohenheim.de
JB: j.bennewitz@uni-hohenheim.de

Abstract

To reduce the costs of genomic selection, selection candidates can be genotyped with an SNP panel of reduced density. Two strategies for the selection of SNPs to be considered in a very low density SNP panel (384 SNPs) are investigated using 895 progeny tested and genotyped German Piétrain boars. In the first strategy equal spaced SNPs were selected, which were used to impute the high density genotypes. In the second strategy SNPs were selected based on results of association analysis. Accuracies of DGV for the two strategies were obtained from cross validation. A regression approach to correct for the upward bias of the cross validation accuracy of the DGV was used. The first strategy resulted in more accurate DGV. This implies that imputation is beneficial even if only 384 SNPs are genotyped for the selection candidates.

Keywords: Piétrain sire line, Unbiased DGV accuracies, Very low-density SNP panel

Introduction

In genomic selection a large number of single nucleotide polymorphisms (SNPs) spread across the genome are used for breeding value estimation in a reference population and for subsequent selection of candidates based on genomically enhanced breeding values (Meuwissen et al., 2001; Goddard and Hayes, 2009). Because of already small generation intervals in pigs there is not so much scope for further reduction of the generation interval, as it is in dairy cattle. Nevertheless genomic selection is relevant also in pig breeding schemes (Lillehammer et al., 2011). Some breeding organisations have started to implement genomic selection by genotyping progeny tested sires with the Illumina PorcineSNP60 BeadChip (Ramos et al., 2009) and to use them as the initial reference population (Wellmann et al., 2013). From an economical point of view, most critical are the high cost of genotyping.

One way to reduce routine costs is to genotype selection candidates with an SNP panel of reduced density. A very low-density panel (VLD panel, 384 SNPs) was suggested to be sufficient for imputation. The strategy is to select SNPs based on equidistant location, high MAF and low correlation of genotypes (equal-spaced, ES). This is described in Wellmann et al. (2013). Missing genotypes can then be imputed using genotyping information from the individuals in the reference population. Genomic breeding values can then be estimated for the selection candidates in the same way as if they were genotyped for the full set of SNPs. To avoid an error accumulation, selection candidates which are selected for breeding the next generation have to be subsequently genotyped for the full set of markers.

Another strategy to select SNP for the VLD panel is to choose SNP being significant in GWAS or having high average BayesA effects (largest-effect, LE). An imputation step and a subsequent typing of the selected individuals afterwards are not needed.

The aim of the present study was to apply different strategies to implement genomic selection in the sire line Piétrain population using VLD panels. The SNPs were selected to be included in the VLD panel using index-similar procedures. Methods for genotype imputation and for the estimation of the accuracy of genomic breeding values were applied. The methods were validated using correlation between direct genomic values (DGV) and estimated breeding values (EBV), and approximate accuracies of DGV. A special emphasis was laid on obtaining bias reduced cross validation accuracies of DGV.

Materials and Methods

Genotypes and conventional EBV for 14 traits, (growth (2), carcass (9) and meat quality (3)) of 895 German Piétrain boars were available from breeding organizations. Boars were genotyped with the PorcineSNP60 Bead Chip (Ramos et al., 2009). A total of 48062 markers remained in the data set after quality control and were used for the analysis. Haplotypes were reconstructed using default parameters in fastPHASE (Scheet and Stephens, 2006). These were inserted into Haploview v4.1 (Barrett et al., 2005) to estimate r^2 (Hill and Robertson, 1986) values for SNP pairs with a distance ≤ 5Mb. Genotyped boars were split into a training and a validation set. The validation set for imputation and genomic selection consisted of 100 boars, which were the youngest animal with high-density (HD) genotyped sires. The remaining 795 boars were included in the training set.

From the set of markers in the validation set two VLD panels were built. SNPs were selected based on scores. For the first panel the score was calculated as the product of a function of *P-values* and the estimated contribution to the additive genetic variance (largest-effect, LE). The significance of the SNP was determined by GWAS (Stratz et al., 2014). Within the function, markers with small *P-values* were scored higher than markers with larger *P-values*. BayesA was used to estimate the contribution to the additive genetic variance. Herein the effect of the marker was defined as its average contribution to the additive variance of traits. The traits were standardized to have the same additive variance.

For the second subset, SNPs were selected based on equidistant location, high MAF, and low correlation of genotypes (equal-spaced, ES). Thereby even maker being at similar genomic positions but being not in LD were included in the VLD panel. These VLD SNPs were used to impute the 48062 SNPs using LD and linkage information. SNP selection and imputation was done as described in Wellmann et al. (2013). The imputation error rate and accuracies were calculated for the VLD-ES strategy. SNPs not included in the VLD panel were masked in the validation set and imputed using the training set. The imputation error rate was computed as the proportion of masked SNP genotypes that were not correctly imputed. The imputation accuracy for an individual was computed as the squared correlation between its true and imputed genotypes.

DGV were estimated with GBLUP (Meuwissen et al., 2001) using deregressed EBVs and different marker subsets. Boars were progeny tested with varying number of offspring. Therefore

GBLUP was extended to account for heterogeneous error variances as described in Garrick et al. (2009).

The BLUP EBV of individuals of the training and a validation sets were estimated in a single evaluation, which results in a correlation of the EBV and the prediction errors. The error correlation leads to an overestimation of the DGV accuracies if it is not accounted for (Amer and Banos, 2010). This is especially a problem for EBV with low accuracies due to a limited number of offspring. To obtain bias reduced accuracies of DGV, the multiple regression approach of Wellmann et al. (2013) was applied for the EBV as follows.

$$r_{DGV,\,EBV_t} = a_0 + a_1 r_t^{Val} + a_2 r_t^{Train} + e_t \quad (1)$$

This equation predicts the expected correlation between DGV and EBV. The intercept a_0 and the regression coefficients a_1 and a_2 are fixed effects, and the errors e_t are normally distributed. r_t^{Val} and r_t^{Train} are the accuracies of the EBV in the validation set, respectively the accuracy in the training set. Note that for $r_t^{Val} \to 1$, the EBV approximates the TBV, so the contribution of the prediction error to the correlation approaches zero. Thus, for $r_t^{Val} = 1$, the expected correlation between EBV and DGV equals the expected accuracy of DGV for a randomly chosen trait with r_t^{Train} specified. This can be estimated as

$$\hat{r}_{DGV,\,TBV_{rand}} = \hat{a}_0 + \hat{a}_1 1 + \hat{a}_2 r_t^{Train}$$

For simplicity we assumed that possible dependency of the error e_t on r_t^{Val} is negligible. Therefore, the accuracy of DGV for trait *t* was estimated as

$$\hat{r}_{DGV,\,TBV_t} = \hat{a}_0 + \hat{a}_1 1 + \hat{a}_2 r_t^{Train} + \hat{e}_t \quad (2)$$

Genomic predictions, $r_{DGV,\,EBV}$ and $r_{DGV,\,TBV}$ were estimated for three SNP sets, i.e. the full set (HD), the VLD-ES set, and VLD-LE set.

Results and Discussion

All possible SNP pairs with a distance $\leq$5Mb produced 3284064 pairwise LD values on the 18 porcine autosomes. The average r^2 between the SNP was 0.34 and points to a high genetic diversity within this breed. The mean imputation error rate (accuracy) for masked markers was 0.133 (0.79).

In Figure 1 the results of eq. (1) are visualized. The solid line shows the regression function as estimated in eq. (1). The dotted line shows the function when the pedigree information between sires were omitted, i.e. the EBV was determined solely by progeny records (see Wellmann et al., 2013 for details). This figure shows that the lower the accuracy of the EBV in the validation set the greater the correlation between DGV and EBV if the full pedigree information is used. This is very likely due to the error correlation. The regression approach (eq. 1+2) accounts for this upward bias and produces bias corrected DGV accuracies. In Table 1, the correlations and accuracies for the HD strategy, the VLD-ES and the VLD-LE strategy are shown. The mean $r_{DGV,\ EBV}$(0.62) and $r_{DGV,\ TBV}$(0.42) were highest for the HD strategy. $r_{DGV,\ EBV}$ and $r_{DGV,\ TBV}$ was 0.59 and 0.37 for the VLD-ES strategy. For VLD-LE strategy $r_{DGV,\ EBV}$ and $r_{DGV,TBV}$ was 0.49 and 0.19, respectively. Averaged over all traits the results of the scenarios with regard to the $r_{DGV,\ EBV}$ is HD>VLD-ES>>VLD-LE.

Although the accuracies of DGV in the HD strategy were highest, this method was not implemented in the routine genomic selection application, because of the high genotyping costs. The VLD-LE strategy seemed to be promising, because an imputation step and a subsequent typing of selection candidates afterwards are not needed. Therefore the VLD-LE strategy seems to be easier to organize and to implement. However this strategy is less flexible and also the accuracies of the DGVs were substantially lower compared to the VLD-ES strategy.

Conclusion

We showed that with the VLD-ES strategy more accurate DGV were obtained than with the VLD-LE strategy, even if only 384 SNPs are to be genotyped. The regression method for obtaining unbiased estimates of the accuracy of DGV is a promising approach which has to be evaluated in detail using stochastic simulations. Based on the results of this study and of Wellmann et al. (2013) the VLD-ES strategy was implemented in the routine genomic selection application.

Literature Cited

Amer, P. R., and G. Banos (2010). Implications of avoiding overlap between training and testing data sets when evaluating genomic predictions of genetic merit. *Journal of Dairy Science* **93**: 3320-3330.

Garrick, D. J., J. F. Taylor, R. L. Fernando (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetic Selection Evolution* **41**: 44.

Goddard, M. E., and B. J. Hayes (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10**: 381-391.

Lillehammer, M., T. H. E. Meuwissen, A. K. Sonesson (2011). A comparison of dairy cattle breeding designs that use genomic selection. *Journal of Dairy Science* **94**: 493-500.

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

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. V. Tassell, J. F.Taylor, R. T. Wiedmann, L. B. Schook, 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.

Stratz, P., R. Wellmann, S. Preuss, K. Wimmers, J. Bennewitz (2013). Genome-wide association analysis for growth, muscularity and meat quality in Piétrain pigs. *Animal Genetics*: DOI:10.1111/age.12133.

Wellmann R., S. Preuß, E. Tholen, J. Heinkel, K. Wimmers, J. Bennewitz (2013). Genomic selection using low density marker panels with application to a sire line in pigs. *Genetics Selection Evolution* **45**: 28.

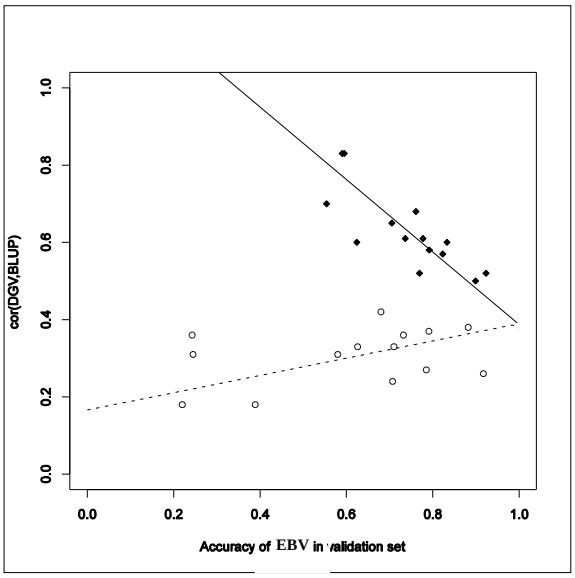
Acknowledgements

The project was funded by the Ministerium für ländliche Räume und Verbraucherschutz in Baden-Württemberg, Germany, and by the Schweinezuchtverband Baden-Württemberg and Schleswig-Holstein, both Germany.

Table 1 Correlations between DGV and EBV and accuracies of DGV. Results from the HD-strategy (48062 marker), VLD-ES strategy and VLD-LE strategy (384 marker)

Trait group	$r_{DGV, EBV}$			$r_{DGV, TBV}$ (eq. 1+2)		
	HD	VLD-ES	VLD-LE	HD	VLD-ES	VLD-LE
Growth	0.51	0.52	0.35	0.41	0.41	0.21
Muscularity	0.60	0.56	0.43	0.42	0.36	0.17
Meat quality	0.76	0.70	0.68	0.43	0.34	0.20
mean	0.62	0.59	0.49	0.42	0.37	0.19

Figure 1 Correlations between DGV and validation set EBV



The regression lines show how the correlation between DGV and EBV depends on the accuracies of the EBV in the validation set. The solid line corresponds to the situation in which the standard complete pedigree was used to estimate the EBV. The dotted line shows the function when the pedigree information between sires were omitted, i.e. the EBV was determined solely by progeny records.

General Discussion

In **chapter one** and **two**, QTL mapping for economical relevant traits was performed between populations, using three connected porcine F_2-crosses, and within a segregating German Piétrain population. In **chapter three** first attempts were made to study the association between the expression of the candidate gene *MYOD1* and economical relevant traits. In **chapter four** the genome of the Piétrain population was investigated in respect of the pattern of linkage disequilibrium and of selection signatures. In **chapter five**, different strategies to implement genomic selection in the sire line Piétrain population using very low density panels were investigated.

Methods for QTL mapping

In linkage studies (Mohrmann et al., 2006; Liu et al., 2007; Rückert and Bennewitz, 2010), the F_2 individuals were mainly genotyped with microsatellite markers and performance tested for the traits of interest. In **chapter one**, QTL mapping relied on linkage between marker and QTL, and variation within the founder breeds was not utilised. In a simulation study of Bennewitz and Wellmann (submitted) it was shown that the identification of causal genes using pooled F_2 families with maximum marker density is only suitable if the alleles are also segregating in the founder breeds. This enables to use meiosis which occurred in the pooled crosses in combination with historical meiosis. It is being considered to use sequence data from the F_2 individuals from the design of Geldermann et al. (2003) to conduct genome-wide association studies in the F_2 individuals. If the alleles are also segregating in the founder breeds it should be possible to increase the mapping precision and possibly find the causal mutation. Putative interaction effects between the mapped causal mutation as well as between the mutation and the *RYR1:g.1843C>T* genotype remain to be investigated.

Apart from the already discussed factors determining the experimental power to detect true associations, the choice of the relationship matrix is essential to prevent population stratification effects. In **chapter two** the numerator relationship matrix (NRM) was used in the single-marker association analysis. It might be of interest to use the genetic relationship matrix (GRM) instead of the NRM for modelling the sample structure. GRM models the genome-wide sample structure based on marker data. There are two possibilities for

modelling. In MLMi (mixed-linear-model included), the SNP is fitted in the model, as a fixed effect for which association is tested for and as a random effect as part of the GRM (Sawcer et al., 2011). However it was shown, that double fitting of the markers can lead to a loss in power (Lippert et al., 2011; Listgarten et al., 2013). Therefore in MLMe (MLM excluded) the SNPs for which association is tested for are excluded from the GRM (Yang et al., 2013).
BayesC (Verbyla et al., 2009; 2010) was used for the multi-marker association analysis in **chapter two.** In BayesC modeling of population stratification effects represents no problem because this is done by markers belonging to the small distribution. However until now there is no formal testing statistic for markers belonging to the large distribution available, but two ways were suggested. The easiest way is to show the BayesC effects or posterior probabilites. This was done in this study. An alternative is presented by Peters et al (2012), who simultaneously tested for associations of windows with phenotype using BayesC analyses.

Methods for candidate gene search for association studies

In **chapter three** the gene *MYOD1* was suggested as a candidate gene because it is influencing muscularity and meat quality traits (Lee et al., 2012; Urbański and Kurył, 2004), which was in agreement with trait correlated expressions found in the segregating German Piétrain population. However no association was found between the expression of *MYOD1* and the genotypes at the SNP ASGA0010149:g. 47980126G>A.
Different strategies are available to find genes which are involved in the determination of the traits. Ponuksilli et al. (2008; 2010) used microarray data of F_2 animals to search for genes having trait correlated expressions. The authors took the trait correlated expression of genes as phenotype to determine genomic loci that control variation in gene expression, so called eQTL. Wysocki et al. (2013) used microarray data from White Leghorn lines, being divergent selected for the propensity to feather pecking, to identify differential expressed genes. To study the role of functional candidate genes, gene ontology (GO) terms can be assigned to the transcripts using e.g. Blast2GO (Conesa et al., 2005). To allow for the differentiation of splice variants and the identification of new transcripts next generation sequencing is suggested to be used in future.

Methods to detect Selection Signatures

For the detection of selection signatures two types of methods are available. The first one is based on haplotypes and the second one on frequency. One haplotype-based method is the Extended Haplotype Homozygosity test (*EHH*, Sabeti et al., 2002). An extension of the *EHH*-test is the integrated haplotype score (*iHS*, Voight et al., 2006). Therefore the *EHH* is summed over all sites away from a core SNP, and compared between haplotypes carrying the ancestral and derived allele at the SNP locus. Applying the *EHH* test statistic in **chapter four** was not successful to find selective sweeps. It is suggested to conduct haplotype-based methods using higher marker density to find regions under recent selection.

More recent studies concentrated on detecting pattern of positive selection between breeds or lines based on allele frequency differences. To determine a genome-wide pattern of positive selection between breeds or lines, the F_{st} at each locus can be calculated according to Wright (1931). A sliding window approach was used by Mancini et al. (2013) to determine regions with high (low) F_{st}, where divergent (balanced) selection has taken place between breeds or lines.

Strategies for genomic selection in pig breeding

Two methods for the selection of SNPs to be included on a very low-density panel (VLD panel, 384 SNPs) were discussed in **chapter five.** Thereby the equal spaced (ES) strategy with imputation of missing genotypes outperformed the largest effect (LE) strategy because of the more accurately estimated DGV. In this study SNPs for the VLD-ES method were chosen based on equidistant location, high MAF and low correlation of genotypes. For the VLD-LE method SNP were chosen based on the results from **chapter two**. Having information on the whole genomic sequence from the animals in combination with the presented methods (**chapter one, two, three and four**) might be helpful to dissect down to the causal mutation. Maybe including the causal mutations on the low density SNP chip might lead to more accurate DGV in the VLD-LE strategy.

References

Conesa, A., S. Götz, J. M. Garcia-Gomez, J. Terol, M. Talon, M. Robles (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674-3676.

Lee, E. A., J. M. Kim, K. S. Lim, Y. C. Ryu, W. M. Jeon, K.C. Hong (2012). Effects of variation in porcine MYOD1 gene on muscle fiber characteristics, lean meat production, and meat quality traits. *Meat Science* **92:** 36–43.

Lippert, C., J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson, D. Heckerman (2011). FaST linear mixed models for genome-wide association studies. *Nature Methods* **8**: 833–835.

Listgarten, J. C. Lippert, C. M. Kadie, R. I. Davidson, E. Eskin, D. Heckerman (2012). Improved linear mixed models for genome-wide association studies. *Nature Methods* **9**: 525–526.

Liu, G., D. G. J. Jennen, E. Tholen, H. Juengst, T. Kleinwächter, M. Hölker, D. Tesfaye, G. Ün, H. J. Schreinemachers, E. Murani, S. Ponsuksili, J. J. Kim, K. Schellander, K. Wimmers (2007). A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Piétrain resource population. *Animal Genetics* **38**: 241–252.

Mohrmann, M., R. Roehe, P. W. Knap, H. Looft, G. S. Plastow, E. Kalm (2006). Quantitative trait loci associated with AutoFOM grading characteristics, carcass cuts and chemical body composition during growth of Sus scrofa. *Animal Genetics* **37**: 435–43.

Mancini, G., M. Gargani, G. Chillemi, E. L. Nicolazzi, P. A. Marsan, A. Valentini, L. Pariset (2014). Signatures of selection in five Italian cattle breeds detected by a 54K SNP panel. *Molecular Biology Reports* **41**: 957-965.

Peters, S.O., K. Kizilkaya, D. J. Garrick, R. L. Fernando, J. M. Reecy, R. L. Weaber, G. A. Silver, M. G. Thomas (2012). Bayesian genome wide association analyses of growth and yearling ultrasound measures of carcass traits in Brangus heifers. *Journal of Animal Science* **90**: 3398-3409.

Ponsuksili, S., E. Jonas, E. Murani, C. Phatsara, T. Srikanchai, C. Walz, M. Schwerin, K. Schellander, K. Wimmers (2008). Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle. *BMC Genomics* **9:** 367.

Ponsuksili, S., E. Murani, M. Schwerin, K. Schellander, K. Wimmers (2010). Identification of expression QTL (eQTL) of genes expressed in porcine M. longissimus dorsi and associated with meat quality traits. *BMC Genomics* **11**: 572.

Rückert, C., and J. Bennewitz (2010). Joint analysis of three connected F_2-crosses on pigs.

Genetics Selection Evolution **42**: 40.

Rückert, C., P. Stratz, S. Preuss, J. Bennewitz (2012). Mapping QTL for growth and muscling traits in three connected porcine F_2 crosses. *Archives of Animal Breeding* **55**: 577.

Sawcer, S., G. Hellenthal, M. Pirinen, C. C. Spencer, N. A. Patsopoulos, L. Moutsianas, A. Dilthey, Z. Su, C. Freeman, S. E. Hunt (2010). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**: 214–219.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson, G. J. McDonald, H. C. Ackerman, S. J. Campbell, D. Altshuler, R. Cooper, D. Kwiatkowski, R. Ward, E. S. Lander (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832–837.

Seidenspinner, T., J. Bennewitz, F. Reinhardt, G. Thaller (2009). Need for sharp phenotypes in QTL detection for calving traits in dairy cattle. *Journal of Animal Breeding and Genetics* **126**: 455 – 462.

Stratz, P., C. Baes, C. Rückert, S. Preuss, J. Bennewitz (2013). A two-step approach to map quantitative trait loci for meat quality in connected porcine F2 crosses considering main and epistatic effects. *Animal Genetics* **44**: 14-23.

Urbański, P., and Kurył J. (2004). New SNPs in the coding and 5' flanking regions of porcine MYOD1 (MYF3) and MYF5 genes. *Journal of Applied Genetics* **45**: 325–329.

Verbyla, K. L., B .J. Hayes, P. J. Bowman M. E. Goddard (2009). Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genetics Research Cambridge* **91**: 307–311.

Verbyla, K.L., P. J. Bowman, B.J. Hayes M. E. Goddard (2010). Sensitivity of genomic selection to using different prior distributions. *BMC Proceedings* **4**: S5.

Wellmann, R., S. Preuß, E. Tholen, J. Heinkel, K. Wimmers, J. Bennewitz (2013). Genomic selection using low density marker panels with application to a sire line in pigs. *Genetics Selection Evolution* **45**: 28.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.

Wysocki, M., S. Preuss, P. Stratz, J. Bennewitz (2013). Investigating gene expression differences in two chicken groups with variable propensity to feather pecking. *Animal Genetics* **44**: 773–777.

Yang, J., N. A. Zaitlen, M. E. Goddard, P. M. Visscher, A. L. Price (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics* **46**: 100-106.

Liste der in die Dissertation eingebundenen Publikationen

Stratz, P., C. Baes, C. Rückert, S. Preuss, J. Bennewitz (2012). A two-step approach to map quantitative trait loci for meat quality in connected porcine F2 crosses considering main and epistatic effects. *Animal Genetics* **44**: 14–23.

Stratz, P., R. Wellmann, S. Preuss, K. Wimmers, J. Bennewitz (2014). Genome-wide association analysis for growth, muscularity and meat quality in Piétrain pigs. *Animal Genetics* **45**: 350-356.

Stratz, P., K. Wimmers, T. H. E. Meuwissen, J. Bennewitz. Investigations on the pattern of linkage disequilibrium and selection signatures in the genomes of strongly selected sire line European pigs. *Journal of Animal Breeding Genetics*, in press.

Stratz, P., M. Maushammer, S. Preuss, J. Bennewitz. Functional analysis of selected candidate genes to characterize meat quality traits in a segregating pig population. Unpublished Data.

Stratz, P., R. Wellmann, and J. Bennewitz. Strategies to implement genomic selection in pig breeding using very low marker density. *10^{th} WCGALP*, accepted.

Gesamtliste aller wissenschaftlicher Veröffentlichungen

Rückert, C., P. Stratz, S. Preuss, J. Bennewitz (2011). Mapping QTL for metabolic and cytological fatness traits of connected F2 crosses in pigs. *Journal of Animal Science* **90**: 399-409.

Rückert, C., P. Stratz, S. Preuss, J. Bennewitz (2012). Mapping QTL for growth and muscling traits in three connected porcine F_2 crosses. *Archives of Animal Breeding* **55**: 577-589.

Stratz, P., C. Baes, C. Rückert, S. Preuss, J. Bennewitz (2012). A two-step approach to map quantitative trait loci for meat quality in connected porcine F2 crosses considering main and epistatic effects. *Animal Genetics* **44**: 14–23.

Wysocki, M., S. Preuss, P**.** Stratz, J. Bennewitz, (2013). Investigating gene expression differences in two chicken groups with variable propensity to feather pecking. *Animal Genetics* **44**: 773–777.

Stratz, P., R. Wellmann, S. Preuss, K. Wimmers, J. Bennewitz (2014). Genome-wide association analysis for growth, muscularity and meat quality in Piétrain pigs. *Animal Genetics* **45**: 350-356.

Bennewitz, J., S. Bögelein, P. Stratz, M. Rodehutscord, H. P. Piepho, J. B. Kjaer, W. Bessei. Genetic parameters for feather pecking and aggressive behavior in a large F2-cross of laying hens using generalized linear mixed models. *Poultry Science*, **93**: 810-817.

Stratz, P., K. Wimmers, T. H. E. Meuwissen, J. Bennewitz. Investigations on the pattern of linkage disequilibrium and selection signatures in the genomes of strongly selected sire line European pigs. *Journal of Animal and Breeding Genetics*, in press.

Stratz, P., R. Wellmann, S. Preuss, K. Wimmers, J. Bennewitz (2012). GWAS for mapping SNPs affecting meat quality related traits in pigs. *DGFZ/GFT Tagung*, Halle/Saale.

Stratz, P., C. Baes, C. Rückert, S. Preuss J. Bennewitz (2011). Analysing three connected porcine F2-crosses for meat quality traits. *DGFZ/GFT Tagung*, Freising-Weihenstephan.

Stratz, P., T. H. E. Meuwissen, J. Bennewitz. Recent selection signatures in the Piétrain pig genome. *64th Annual Meeting EAAP*, Nantes, France.

Stratz, P., M. Maushammer, S. Preuss, J. Bennewitz. Functional analysis of selected candidate genes to characterize meat quality traits in a segregating pig population. Unpublished Data.

Stratz, P., R. Wellmann, J. Bennewitz. Strategies to implement genomic selection in pig breeding using very low marker density. *10th WCGALP*, accepted.

Danksagung

Zu allererst möchte ich mich bei Herrn Prof. Dr. Jörn Bennewitz für die Überlassung des Themas, die Betreuung und Unterstützung während der Promotion bedanken.

Des Weiteren möchte ich mich für die Übernahme des Koreferates bei Herrn Prof. Dr. K. Wimmers bedanken.

Allen Freunden und Kollegen, die mich während der Promotion begleitet haben, ein Dankeschön. Ein Dank auch an die Mitarbeiter der LSZ Boxberg, die uns bei der Muskel- und Blutgewinnung unterstützt und die Daten aus der Leistungsprüfung zur Verfügung gestellt haben

Ein großes Dankeschön gilt meiner Familie, die immer ein offenes Ohr für mich hatte.

Ein ganz besonders herzliches Dankeschön gilt meiner Freundin, die mir in schwierigen wie auch in guten Zeiten zur Seite stand und mich tatkräftig unterstützt hat.

Ebenfalls möchte ich mich beim DAAD (Deutscher Akademischer Austauschdienst) für die finanzielle Unterstützung während des Aufenthaltes an der Norwegischen Universität der Biowissenschaften bedanken. Besonders bedankende möchte ich mich dabei bei Herrn Prof. Dr. T. H. E. Meuwissen, der mich während des Aufenthaltes betreut hat

Zu allerletzt möchte ich mich bei der Carl-Zeiss-Stiftung für die finanzielle Unterstützung bedanken.

Curriculum Vitae

Geburtsdatum	15.02.1986	
Geburtsort	Villingen	
Staatsangehörigkeit	deutsch	
Eltern	Karlheinz Stratz, Ingenieur	
	Isolde Stratz, Bürokauffrau	
Schulische Ausbildung	1992-1996	Friedrichschule Furtwangen
	1996-2005	Otto-Hahn-Gymnasium Furtwangen
		Kernfächer: Physik und Biologie
		Schulabschluss: Abitur
Zivildienst	2005/06	Zivildienst an der Rehaklinik Katharinenhöhe
Studium	10.2006-09.2010	Agrarbiologiestudium, an der Universität Hohenheim
		Schwerpunkt: Agrarbiotechnologie
		Abschluss: Diplom in Agrarbiologie
Beschäftigung	09./10.2008	Praktikum auf der landwirtschaftlichen Versuchsstation 401 in Kleinhohenheim
	04./07.2012	Praktikum bei der RLP AgroScience GmbH-ALPlanta-Institut für Planzen-forschung
Promotionsstudium	11.2010-04.2014	an der Universität Hohenheim
		Betreuer: Prof. Dr. J. Bennewitz

Ort und Datum Unterschrift

Eidesstattliche Versicherung

gemäß § 8 Absatz 2 der Promotionsordnung der Universität Hohenheim zum Dr.sc.agr.

1. Bei der eingereichten Dissertation zum Thema
   "Genome-wide mapping and functional analysis of genes determining the meat quality in pigs"
   handelt es sich um meine eigenständig erbrachte Leistung.
2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.
3. Ich habe nicht die Hilfe einer kommerziellen Promotionsvermittlung oder -beratung in Anspruch genommen.
4. Die Bedeutung der eidesstattlichen Versicherung und der strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Die Richtigkeit der vorstehenden Erklärung bestätige ich. Ich versichere an Eides Statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

_____

Ort und Datum Unterschrift