Linkage disequilibrium pattern and genome-wide association mapping for meat traits in multiple porcine F2-crosses

Patrick Stratz[1], Markus Schmid[1], Robin Wellmann[1], Siegfried Preuß[1], Iulia Blaj[2], Jens Tetens[3], Georg Thaller[2] & Jörn Bennewitz[1]

[1] University of Hohenheim, Germany
[2] University of Kiel, Germany
[3] University of Göttingen, Germany

UNIVERSITY OF HOHENHEIM

11th WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION
11–16 February 2018
Aotea Centre
Auckland
New Zealand
wcgalp.com

Introduction

➢ In pig breeding numerous F2-crosses from distantly and closely related founder breeds were established in the past

➢ Availability of genotypes for a small number of microsatellite markers →QTL mapping was done using linkage analysis

➢ Advent of the standard porcine chip (62K SNP chip) →possibility to conduct genome wide association studies (GWAS) across the crosses

Introduction

➢ Stochastic simulation studies (Schmid *et al.*, under revision) to analyse the prospects of GWAS in porcine F2-crosses genotyped with 62K SNP chip

➢ High precision and power in GWAS, when the founder breeds used to establish the crosses were closely related and when sample size can be enlarged, e.g., by pooling data from several crosses (Schmid *et al.,* under revision)

Aim of the study

Analysis of

➢ LD breakdown in single and pooled F2-crosses

➢ Prospects of GWAS in multiple porcine F2-crosses

Material

2570 F2-individuals from four existing F2 crosses (Rückert and Bennewitz, 2010; Boysen *et al.*, 2010)

➢ Established from the founder breeds Wild Boar (W), Piétrain (P), Meishan (M) and a F1 Large White ✕ Landrace cross (LWL) → WxP, WxM, MxP and PxLWL

➢ Filtered 62K SNP chip genotypes and pre-corrected phenotypes for environmental effects, the effect of the cross and of *RYR1* (Fujii *et al.*, 1991) for conductivity and dressing out

Methods

Population analysis

➤ Estimation of the population differentiation index (F_{ST}) between the founder breeds

➤ Haplotype reconstruction and imputation of missing genotypes using Beagle 3.3.2 (Browning and Browning, 2008)

➤ Estimation of r^2 (≤5Mb distance) using PLINK 1.9 (Purcell *et al.*, 2007)

Methods

Statistical analysis

➢ Variance components were estimated using GCTA (Yang *et al.*, 2011)

➢ GWAS were conducted using a multi-marker model (BayesD, Wellmann and Bennewitz, 2012) and a windows based approach (WPPA, Fernando and Garrick, 2013)

UNIVERSITY OF
HOHENHEIM

11th WORLD CONGRESS
ON GENETICS
11–16 February 2018   APPLIED TO
Aotea Centre
Auckland
New Zealand   LIVESTOCK PRODUCTION
wcgalp.com

BayesD Prior distribution

SNP-effect

➢ Comes from one out of two differently scaled centralized t-distributions. $V_G = V_A + V_D$, with V_A, V_D (GCTA estimates) →scaling parameter (s²): $s^2_{small} = 0.01 * s^2_{big}$, and $E(V_G) = V_{\hat{G}}$

➢ Probability that a SNP is in LD with a QTL and therefore drawn from the larger scaled distribution was: pLD = 0.02

➢ Degrees of freedom: $df = 3$

UNIVERSITY OF HOHENHEIM

11th WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION
11–16 February 2018
Aotea Centre
Auckland
New Zealand
wcgalp.com

Window Posterior Probability of Association

SNP-effects were estimated using 100,000 Gibbs-Samples, among them 50,000 were discarded as burn-in

➢ Every 25th sample was drawn to calculate the ratio between the estimated and the expected window variance for 1Mb sliding windows

➢ The WPPA criterion (following Fernando and Garrick, 2013) was calculated as the mean of ratios above one

UNIVERSITY OF HOHENHEIM

11th | WORLD CONGRESS
ON GENETICS
11–16 February 2018
Aotea Centre
Auckland
New Zealand | APPLIED TO
LIVESTOCK PRODUCTION
wcgalp.com

Results (population analysis)

The mean F_{ST} index over the SNPs was:

➢ $F_{ST} \approx 0.07$ between Piétrain and Large White × Landrace

➢ $F_{ST} \approx 0.17$ between Piétrain and Meishan

➢ $F_{ST} \approx 0.25$ between Meishan and Large White × Landrace

UNIVERSITY OF HOHENHEIM

11th WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION
11–16 February 2018
Aotea Centre
Auckland
New Zealand
wcgalp.com

Chromosome 1

Chromosome 1

Pooled crosses

UNIVERSITY OF HOHENHEIM

11th WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION
11–16 February 2018
Aotea Centre
Auckland
New Zealand
wcgalp.com
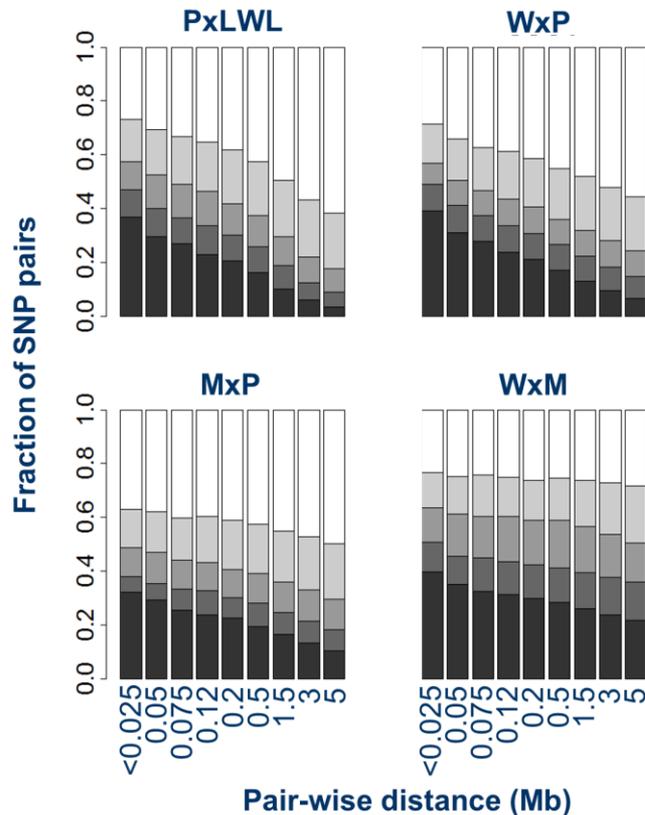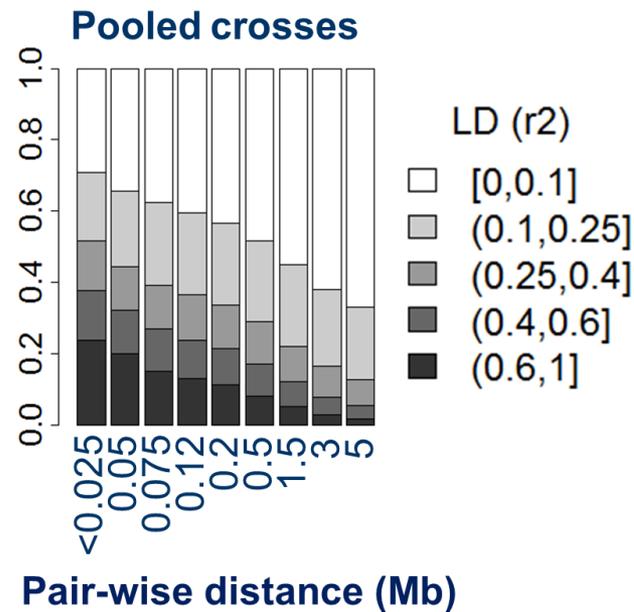
Results (statistical analysis)

Trait	CV	Variance components from GCTA		
		σ_a^2	σ_d^2	σ_g^2
Conductivity	71.37	0.17 (0.04)	0.16 (0.03)	0.33 (0.04)
Dressing out	3.21	0.30 (0.04)	0.05 (0.02)	0.35 (0.03)

CV: Coefficient of variation;
σ_a^2: Estimated additive genetic variance;
σ_d^2: Estimated dominance genetic variance;
σ_g^2: Estimated genetic variance.

UNIVERSITY OF HOHENHEIM

11th **WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION**
11–16 February 2018
Aotea Centre
Auckland
New Zealand
wcgalp.com

Results (statistical analysis)

Conductivity

Dressing out
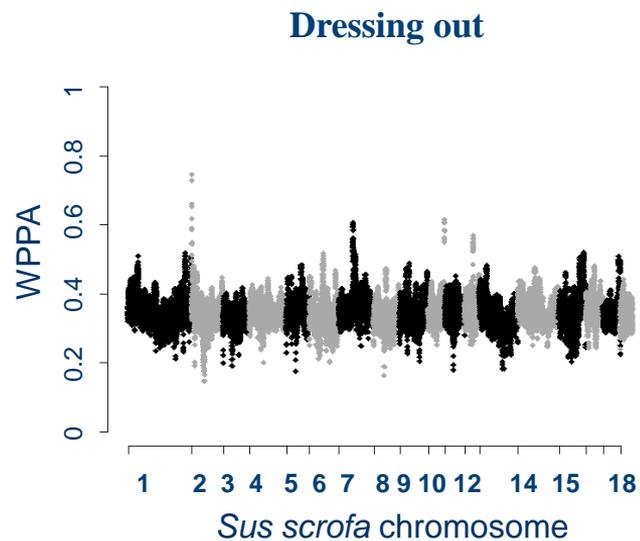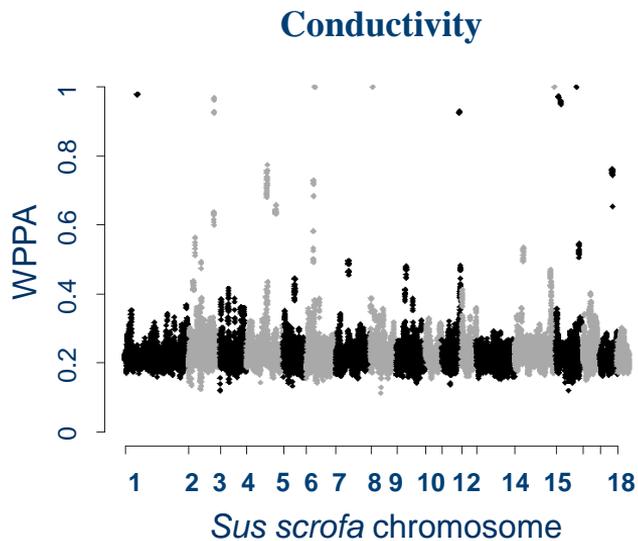
Take home message

➢ Lowest differentiation between European-type breeds →fastest breakdown of LD

➢ Greatest differentiation between the European-type and Asian-type breeds → slowest breakdown of LD

➢ Benefit for applying a GWAS in F2-crosses with slow breakdown of LD and hence long range of LD blocks is limited, even with dense marker panels

UNIVERSITY OF
HOHENHEIM

11th WORLD CONGRESS
ON GENETICS
11–16 February 2018
Aotea Centre
Auckland
New Zealand
APPLIED TO
LIVESTOCK PRODUCTION
wcgalp.com

Take home message

➢ Pooling data from different F2-crosses results in the fastest breakdown of LD and hence reduce the length of LD blocks and justify a GWAS

➢ Several trait associated regions were found in this study

→ Further investigations needed to see if SNPs in those regions are still segregating in the current Piétrain reference population
→ Preselection of target regions in the genome for subsequent fine-mapping

Acknowledgements

➤ This study was supported by the German Research Foundation

References

Bennewitz J., Edel C., Fries R., Meuwissen T.H.E. & R. Wellmann, 2017. Application of a Bayesian dominance model improves power in quantitative trait genome-wide association analysis. Genetics Selection Evolution 49:7.

Browning B.L. & Browning S.R. (2008) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American Journal of Human Genetics 84, 210-223.

Fernando R.L. & D. Garrick, 2013. Bayesian methods applied to GWAS. In: Genome-Wide Association Studies and Genomic Prediction, C. Gondro, J. van der Werf, B. Hayes (editors), Springer Science+Business Media S.A., Luxemburg, Luxemburg, p 237-274.

References

Fujii J.J. (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* 253, 448-451.

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

Schmid M., Wellmann R. & J. Bennewitz, 2016. Power and precision of mapping genes in simulated F2 crosses using whole genome sequence data. Book of Abstracts, 67th Annual Meeting of the European Federation of Animal Science, Session 27.

References

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

Yang J., Lee S.H., Goddard M.E. & Visscher P.M. (2011) GCTA: A tool for genome-wide complex trait analysis. American Journal of Human Genetics 88, 76-82.

Supplemental Material

Introduction

Piétrain is frequently used as sire line in cross breeding

➢ Selection: + muscling and leanness
                  - reduced meat quality

increased importance of meat quality:

➢ Genetic foundation of meat quality enables the mapping of genes

Introduction

➢ **In the Past:** Porcine F2-crosses established from distantly and closely related founder breeds were used for **linkage mapping**

→Majorgen of the meat quality: RYR1:g.1843C>T (Fujii *et al.*, 1991)

➢ Based on **evenly spaced SNP marker**, which occur with high frequency across the genome offer the possibility to conduct **GWAS** across the crosses

→higher power (more animals) and precision (more informative meiosis) (Schmid *et al.*, under revission)

Introduction

Single marker GWAS

➢ Multiple tests for significance of the fixed SNP effect → post-GWAS studies (Bonferroni, FDR….)

➢ Population structure is modelled by the genetic relationship matrix (GRM) or other genetic effects

➢ Lot of associations, however the variance explained by the QTL is small (incomplete LD-information, small QTL effects)

Introduction

Multi marker GWAS

➢ SNPs were treted as random and modeled simultaneously
  → approximates the population structure

➢ Adjacent SNPs can jointly explain much more QTL variance than each SNP on its own
  → SNPs in genomic window

➢ Accounting for dominance effects improves the QTL mapping power (Bennewitz *et al.*, 2017)

Introduction

GWAS

➢ Legarra et al. (2015) compared single marker mixed model association analysis and Bayes whole genome association analysis with real data structure

➢ They did not report a clear superiority of a single method, but recommended to apply more than one method on real data.

Material and Methods

Hohenheim design:

➤ F2-animals from three crosses (Meishan x Pietrain (MxP), Wildschwein x Pietrain (WxP) and Wildschwein x Meishan (WxM) (Geldermann *et al.,* 2003)

Kiel design:

➤ F2-animals from the crosses PxLW and Px(LWxLR) (Boysen *et al.,* 2010) → 8 full sib families (FS1-FS8)

Material and Methods

Quality Control:

Parent-Parent-Child-Errors = 0
MAF>0, CallFrequency >= 0,99
ClusterSeparation >= 0,4
Heterozygoty-excess < 0,15
Not on the X- or Y-chromosom
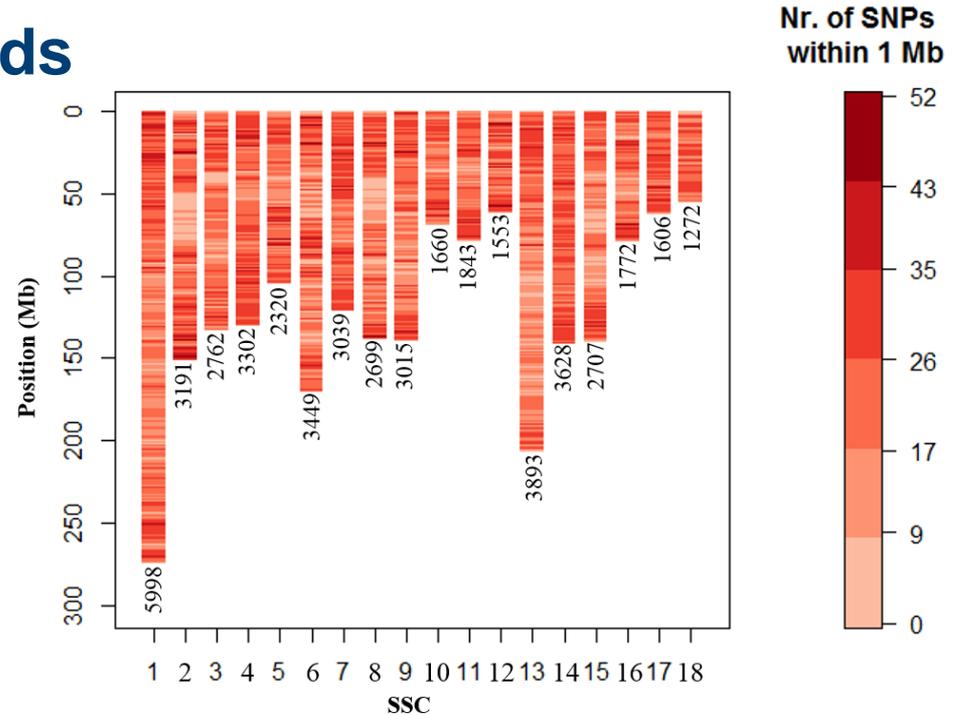
Material and Methods

Animals and Genotypes:

Filtering of the animals:

Callrate < 95%, different sex genetic/pedigree, one or both parents are in disagreement with the pedigree

- ➢ 907 F2-animals from the Hohenheim design

- ➢ 1665 F2-animals from the Kiel design

Material and Methods

The R-package genoKin of Wellmann 2016 was used to:

- ➢ built a segment-based realized relationship matrix with the function segIBD using segments comprising at least 25 SNPs and having a length of minimum 1Mb were used to
- ➢ convert the segment-based realized relationship matrix into a dissimilarity matrix using the function sim2dis

The R-package smacof of de Leeuw and Mair (2009) was used to:
- ➢ visualize the population structure with the function smacofSym using MDS

Heatmap of the marker-based relationship matrix realized for all F2-individuals
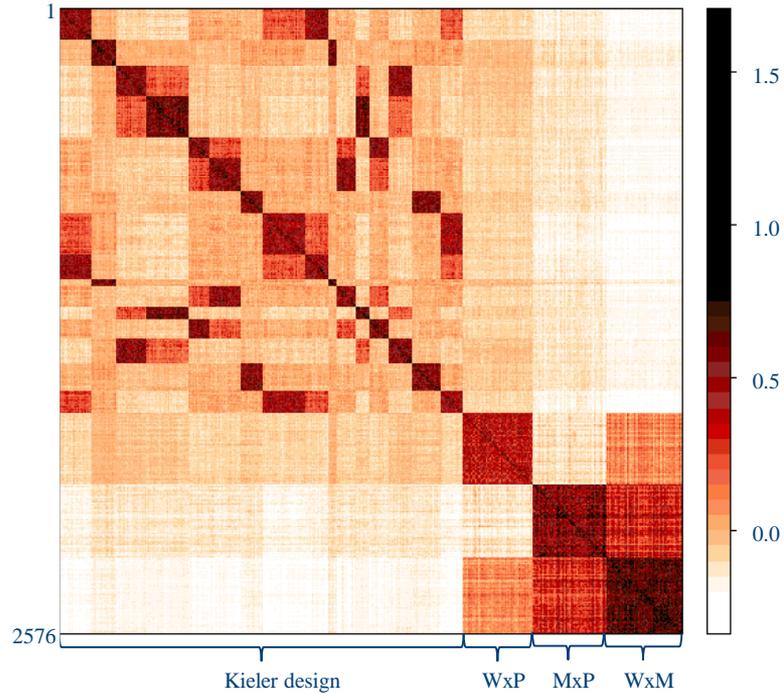
Investigations on the population structure

Multidimensional Scaling

segIBD Stress1 = 0.176
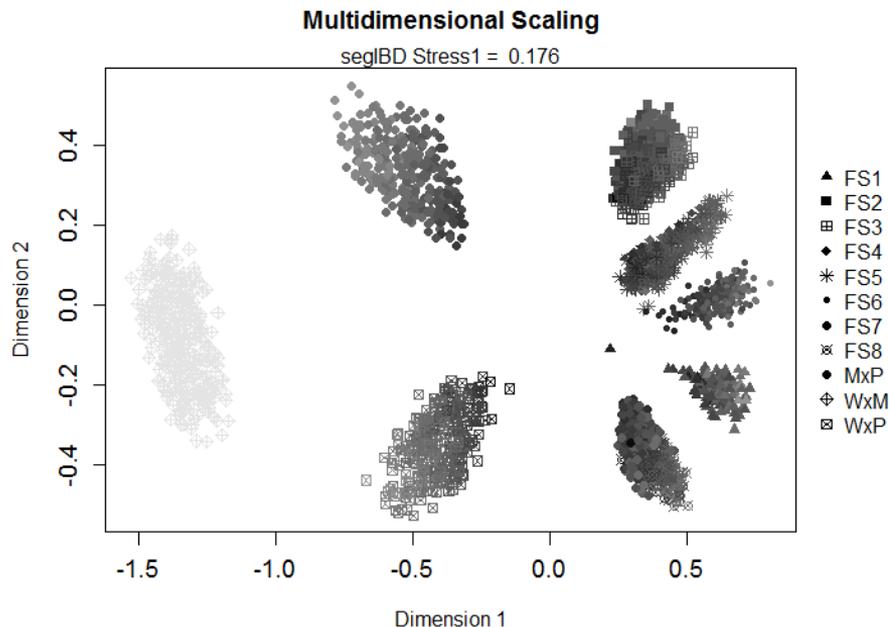
The color gradation indicates the contribution of Piétrain founders to the F2-individuals. The darker the color the higher is the contribution.

Material and Methods

pre-correction	
Hohenheim design	**Kiel design**
Age (slaughter age - 210 days)	Slaughter period
Sex	Sex
Litter	Barn
Season	
In WxP and MxP: RYR1	
cross	

Geldermann *et al., 2003*

Boysen *et al.* 2010

BayesD Prior distribution

Assumption of BayesD from Wellmann and Bennewitz (2012)

➢ Independence between the absolute value of the additive effects $|a|$ and the dominance coefficients $\frac{d}{|a|}$

➢ Small probability that d is bigger than a

→ overdominance is rare, however non negligible

GWAS for meat traits in multiple porcine F2-crosses

Conductivity

GWAS for meat traits in multiple porcine F2-crosses

GWAS for meat traits in multiple porcine F2-crosses

Conductivity

GWAS for meat traits in multiple porcine F2-crosses

Conductivity

GWAS for meat traits in multiple porcine F2-crosses

Conductivity

GWAS for meat traits in multiple porcine F2-crosses

Conductivity

GWAS for meat traits in multiple porcine F2-crosses

Conductivity

GWAS for meat traits in multiple porcine F2-crosses

Conductivity

GWAS for meat traits in multiple porcine F2-crosses

Conductivity

GWAS for meat traits in multiple porcine F2-crosses

Conductivity