Genome-wide association studies for production traits in pooled pig $F_2$ designs

I. Blaji, J. Tetens, S. Preuß, R. Wellmann, J. Bennewitz, G. Thaller

1Institute of Animal Breeding and Husbandry, Christian Albrecht University, Kiel, Germany
2Institute of Animal Breeding and Husbandry, University of Hohenheim, Stuttgart, Germany

EAAP Annual Meeting 2016, Belfast, UK
Session 67: Free communications in genetics
1st of September 2016
Outline

• Introduction

• Objectives

• Materials and Methods

• Results and Discussion
  - LD decay
  - Single cross analysis, meta-analysis and joint analysis

• Conclusions and Perspectives
Introduction

- gene mapping experiments in livestock
  - genetic architecture of quantitative traits
  - genetic markers to facilitate breeding progress
- several $F_2$ resource populations have been established and analyzed

- resolution
- precision
- power

- meioses exploited
- number of individuals included
- marker density

- LD structure
  - the length of the LD blocks can be reduced by pooling several $F_2$ crosses (Bennewitz and Wellmann, 2014)
Objectives

- three-generation experimental populations
  - Piétrain x Large White, Piétrain x (Large White x Landrace) – European breeds cross
  - Meishan x Piétrain, Wild boar x Piétrain – Asian/European breeds cross

- phenotypes: average daily gain (ADG), back fat thickness (BFT), meat to fat ratio (MFR)

- combine data from two experimental $F_2$ crosses
  - structural identification of short chromosomal regions that show evidence for trait association
Materials and Methods

• total of **2,554 animals**
  - 1,894 individuals European breeds cross P x LW/(LW x L)
  - 660 individuals Asian/European breeds cross M/W x P

• **P / F₁ / F₂** genotyped with PorcineSNP60 BeadChip (Illumina)
  - SNP chromosomal positions - current pig genome assembly (Sscrofa build 10.2)

• phenotypes were measured using similar methods and standardized techniques
  (Müller et al. 2000, Borchers 2002)
Materials and Methods

GWAS Workflow

- Individual cross
- Meta-analysis of the individual crosses
- Pooled pre-corrected data

**mixed linear model (MLM) based association analysis**

\[ y = X\beta + g + \varepsilon \]

\[ V = \frac{WW'}{N} \sigma_g^2 + I \sigma_e^2 = A\sigma_g^2 + I \sigma_e^2 \]

**Fixed effects**

<table>
<thead>
<tr>
<th>Cross/FE</th>
<th>European breeds cross</th>
<th>Asian/European breeds cross</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADG</td>
<td>stable, slaughtering period</td>
<td>sex, cross</td>
</tr>
<tr>
<td>BFT</td>
<td>sex, stable, slaughtering period, weight at slaughter</td>
<td>sex, slaughtering period, weight at slaughter, age at slaughter, cross</td>
</tr>
<tr>
<td>MFR</td>
<td>sex, stable, slaughtering period, birth weight</td>
<td>sex, cross</td>
</tr>
</tbody>
</table>
Materials and Methods

GWAS Workflow

- **Individual cross**
- **Meta-analysis of the individual crosses**
- **Pooled pre-corrected data**

**METAL version 2011, Willer et al 2010**
- sample based approach
- analytical strategy

| Input | $N_i$ - sample size for study $i$  
|       | $P_i$ - p-value for study $i$  
|       | $\Delta_i$ - direction of effect for study $i$ |

| Intermediate Statistics | $Z_i = \Phi^{-1}(P_i/2) \ast \text{sign}(\Delta_i)$  
|                        | $w_i = \sqrt{N_i}$ |

| Overall Z-score | $Z = \frac{\sum Z_iw_i}{\sqrt{\sum w_i^2}}$ |
| Overall P-value  | $P = 2 \Phi (|\cdot Z|)$ |
Materials and Methods

GWAS Workflow

- Individual cross
- Meta-analysis of the individual crosses
- Pooled pre-corrected data

- phenotypes pre-corrected in the individual crosses
- MLM: \[ y = X\beta + g + \varepsilon \]

<table>
<thead>
<tr>
<th>Cross/FE</th>
<th>European breeds cross</th>
<th>Asian/European breeds cross</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADG</td>
<td>stable, slaughtering period</td>
<td>sex</td>
</tr>
<tr>
<td>BFT</td>
<td>sex, stable, slaughtering period, weight at slaughter</td>
<td>sex, slaughtering period, weight at slaughter, age at slaughter</td>
</tr>
<tr>
<td>MFR</td>
<td>sex, stable, slaughtering period, birth weight</td>
<td>sex</td>
</tr>
</tbody>
</table>
Results and Discussion

Tab. 1: Descriptive statistics and heritabilities ($h_{SNP}^2$ and $h_{pedigree}^2$) for average daily gain (ADG), back fat thickness (BFT) and meat to fat ratio (MFR)

<table>
<thead>
<tr>
<th>Cross</th>
<th>Trait</th>
<th>N</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>max</th>
<th>$h_{SNP}^2$</th>
<th>$h_{pedigree}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>European</td>
<td>ADG[g]</td>
<td>1769</td>
<td>675.90</td>
<td>92.73</td>
<td>311.0</td>
<td>1039.0</td>
<td>0.35</td>
<td>0.47</td>
</tr>
<tr>
<td>Asian/European</td>
<td>ADG[g]</td>
<td>595</td>
<td>559.40</td>
<td>124.19</td>
<td>125.0</td>
<td>951.0</td>
<td>0.44</td>
<td>0.24</td>
</tr>
<tr>
<td>European</td>
<td>BFT[mm]</td>
<td>1766</td>
<td>27.49</td>
<td>3.84</td>
<td>16.00</td>
<td>42.30</td>
<td>0.43</td>
<td>0.43</td>
</tr>
<tr>
<td>Asian/European</td>
<td>BFT[mm]</td>
<td>595</td>
<td>19.44</td>
<td>6.93</td>
<td>3.70</td>
<td>43.30</td>
<td>0.47</td>
<td>0.56</td>
</tr>
<tr>
<td>European</td>
<td>MFR</td>
<td>1765</td>
<td>0.38</td>
<td>0.10</td>
<td>0.14</td>
<td>0.85</td>
<td>0.46</td>
<td>0.36</td>
</tr>
<tr>
<td>Asian/European</td>
<td>MFR</td>
<td>593</td>
<td>0.62</td>
<td>0.21</td>
<td>0.19</td>
<td>1.39</td>
<td>0.51</td>
<td>0.44</td>
</tr>
</tbody>
</table>
Results and Discussion

Fig. 1: LD decay over physical distance

$F_2$ design
- European breeds cross
- Asian/European breeds cross
- Joint crosses
Fig. 2: Manhattan plot of genome-wide association studies for average daily gain. The red line indicates the Bonferroni-corrected significance threshold ($P=1.1\times10^{-6}$) and the blue line indicates the threshold ($P=2.2\times10^{-5}$) for suggestive SNPs.
Fig. 3: Manhattan plot of genome-wide association studies for back fat thickness. The red line indicates the Bonferroni-corrected significance threshold \((P=1.1\times10^{-6})\) and the blue line indicates the threshold \((P=2.2\times10^{-5})\) for suggestive SNPs.
MFR – meat to fat ratio

Fig. 4: Manhattan plot of genome-wide association studies for meat to fat ratio. The red line indicates the Bonferroni-corrected significance threshold ($P=1.1\times10^{-6}$) and the blue line indicates the threshold ($P=2.2\times10^{-5}$) for suggestive SNPs.
Conclusions

- the meta-analysis was generally more powerful in detecting more precise locations and higher significance levels in the combined crosses vs. single cross
- association levels in pooled pre-corrected data were lower than in the meta-analysis
- common underlying variants that show a different frequency between the two crosses
- chromosomes showing significant evidence for trait association in the meta-analysis
  - **ADG** - SSC2; **BFT** - SSC2, SSC4, SSC7; **MFR** - SSC1, SSC2

Perspectives

- heterogeneous residual variance to be modelled in the joint analysis
- sequencing of the $P$ → imputation → Whole-genome sequence based association studies
Thank you for your attention!

Questions?

*The authors would like to thank the German Research Foundation (DFG) for funding.