Haplotype structure and recent selection signatures in the Piétrain pig genome

Patrick Stratz¹, Jörn Bennewitz¹ and Theo Meuwissen²

¹Institute of Animal Husbandry and Animal Breeding, University of Hohenheim
²Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences
Introduction

Bottom-up approaches:

- No need of well defined phenotypes.
- Identification of core regions, being under recent selection pressure, even in the absence of information as to which trait(s) they regulate.
- Main principles: neutral loci across the genome will be similarly affected by genetic drift, demography, and evolutionary history of populations. In contrast loci under selection reveal “outlier” patterns of variation.
Objectives

• To map putative regions of positive selection in German Piétrain pig breeding populations based on haplotypes (so called selective sweeps).

• To compare them with results of genome-wide association studies for growth, muscularity and meat quality (Stratz et al. 2013).

• To annotate candidate genes to core regions.
Materials and Methods

Data preparation:

- **Animals:** Totally 895 German Piétrain boars out of a segregating population, reflecting a representative sample of the currently relevant breeding population (Wellmann et al. 2013).

- **Genotypes:** Illumina 60K Chip: 47,549 SNP marker

Haplotype reconstruction:

- Default parameters in fastPHASE (Scheet and Stephens 2006)
Materials and Methods

Reconstructed haplotypes were used for

Haplo-block partitioning:

• Based on an algorithm implemented in Sweep v.1.1 (Sabeti et al. 2002)

  1. “Strong LD” if upper 95% confidence bound of $D'$ is between 0.7 and 0.98.
  2. At least three and a maximum of 20 SNPs (Gabriel et al. 2002).
Materials and Methods

‘Extended Haplotype Homozygosity’ test

• Assumption:
  Positive selection pressure increases allele frequency faster and leaves therefore longer extended haplotypes in the vicinity of the causal mutation (genetic hitchhiking) as expected under neutrality.

• Definition of EHHₜ:
  Probability that two randomly chosen gametes carrying a particular haplotype t are identical by descent for the entire interval from the core region to a distance x (Sabeti et al. 2002). It detects the transmission of an extended haplotype without recombination.
Materials and Methods

‘Extended Haplotype Homozygosity’ test

- The EHH of a tested core haplotype is mathematically calculated as:

$$EHH_t = \frac{\sum_{i=1}^{s} \left( \begin{array}{c} e_{ti} \\ 2 \end{array} \right)}{\left( \begin{array}{c} c_t \\ 2 \end{array} \right)}$$

- $$c_t$$: Number of samples of a particular core haplotype $$t$$ within a core region.
- $$e_{ti}$$: Number of samples of a particular extended haplotype $$i$$.
- $$s$$: Number of unique extended haplotypes.
Materials and Methods

‘Extended Haplotype Homozygosity’ test

• To correct for local recombination rates, the ‘Relative EHH’ (REHH; Sabeti et al. 2002) was calculated as follows:

\[ \text{REHH}_t = \frac{\text{EHH}_t}{\overline{\text{EHH}}} \]

• It compares the EHH decay of a particular ‘core haplotype’ t to the decay of EHH on all other ‘core haplotypes’ in the same ‘core region’ combined except the core haplotype we are looking at.
Materials und Methods

‘Extended Haplotype Homozygosity’ test

• To determine the empirical significance of REHH values, core haplotypes were placed in 6 bins based on their frequency (>0.7).

• To achieve normality, REHH values were log-transformed.

• Within each bin the p-value of the normalized REHH value was calculated for all haplotypes in the genome.

• REHH values above a predefined threshold (p<0.02) were considered significant.

• Annotation of genes, previously suggested as being under selection, to those core regions (candidate genes).
Results and Discussion
Pattern of haplotype blocks

- 22854 core haplotypes within 5700 core regions spanning 954 010 kb (32%) of the entire genome were detected.
- Mean core region length: $167.4 \pm 275.9$ kb; maximum: 5052 kb on SSC6.
- Overall 23092 SNPs (49%) participated in forming core regions.
- 922 core haplotypes had a frequency >0.7.
Results and Discussion

Pattern of haplotype blocks

Introduction | Objectives | Materials / Methods | Results / Discussion | Conclusion
Results and Discussion

EHH test over the genome

- Box plot of the distribution of logarithmic P-values in core haplotype frequency bins of 5% difference.
- Core haplotypes with p-values lower than 0.1 (78) and 0.02 (11, q=0.92) are above the dashed and continuous threshold lines.

Haplotype frequency bins
Results and Discussion

EHH test over the genome

- One selective sweep was found on SSC1 and SSC6
- Three selective sweeps were found on SSC18 and six on SSC8
- Only core regions harbouring candidate genes are shown below:

<table>
<thead>
<tr>
<th>Chr</th>
<th>Position</th>
<th>Core length (kb)</th>
<th>Hap freq</th>
<th>EHH</th>
<th>REHH</th>
<th>REHH p-Value</th>
<th>Candidate genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>128 652 415-128 724 560</td>
<td>72.15</td>
<td>0.75</td>
<td>0.51</td>
<td>3.25</td>
<td>0.00461</td>
<td>PPP3CA</td>
</tr>
<tr>
<td>18</td>
<td>31 531 915-31 769 165</td>
<td>237.25</td>
<td>0.76</td>
<td>0.38</td>
<td>2.53</td>
<td>0.01253</td>
<td>CAV1/CA V2</td>
</tr>
<tr>
<td>8</td>
<td>114 555 082-117 361 292</td>
<td>2806.21</td>
<td>0.76</td>
<td>0.82</td>
<td>2.42</td>
<td>0.01474</td>
<td>CaMk2D/ANK2</td>
</tr>
</tbody>
</table>
Conclusion

• 32% of the entire genome is covered by core regions.
• 49% of all investigated SNPs participate in forming core regions.
• 922 core haplotypes had a frequency >0.7.
• On SSC8, SSC18, SSC6 and SSC1, the most significant core regions (p<0.02) were found.
• Significant SNPs found on SSC1 for carcass length content and on SSC6 for daily gain in GWA studies by Stratza et al. (2013) could not be allocated to core regions.
• Following candidate genes on SSC8 and 18 were annotated to core regions and are getting investigated in an ongoing study: PPP3CA, CAV1/CAV2 and CaMk2D/ANK2
Acknowledgements

PS received funding from the Carl-Zeiss-Stiftung, Stuttgart, Germany as well as from the DAAD (‘German Academic Exchange Service’).
Supplemental slides
‘Extended Haplotype Homozygosity’ test

- The decay on all core haplotypes, except the tested core haplotype \( t \) is mathematically calculated as:

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

- \( c_{j} \): number of samples of a core haplotype \( j \) (unequal to \( t \)).
- \( n \): number of core haplotypes in a particular core region.
- \( e_{i} \): number of samples of a unique EH \( i \).
- \( s \): number of unique EH belonging to one core haplotype \( j \).
- \( j \): a particular core haplotype.
EHH test over the genome

- Genome-wide map of P-values for core haplotypes with frequency >0.7.
- Core haplotypes with P-values lower than 0.1 and 0.02 are separated by dashed and continuous threshold lines respectively.