UNIVERSITÄT HOHENHEIM

Haplotype structure and recent selection signatures in the Piétrain pig genome

Patrick Stratz[1], Jörn Bennewitz[1] and Theo Meuwissen[2]

[1]Institute of Animal Husbandry and Animal Breeding, University of Hohenheim
[2] Department 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 Stephans 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 EHH_t:
  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:

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

- 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:

$$REHH_t = EHH_t / \overline{EHH}$$

- 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).

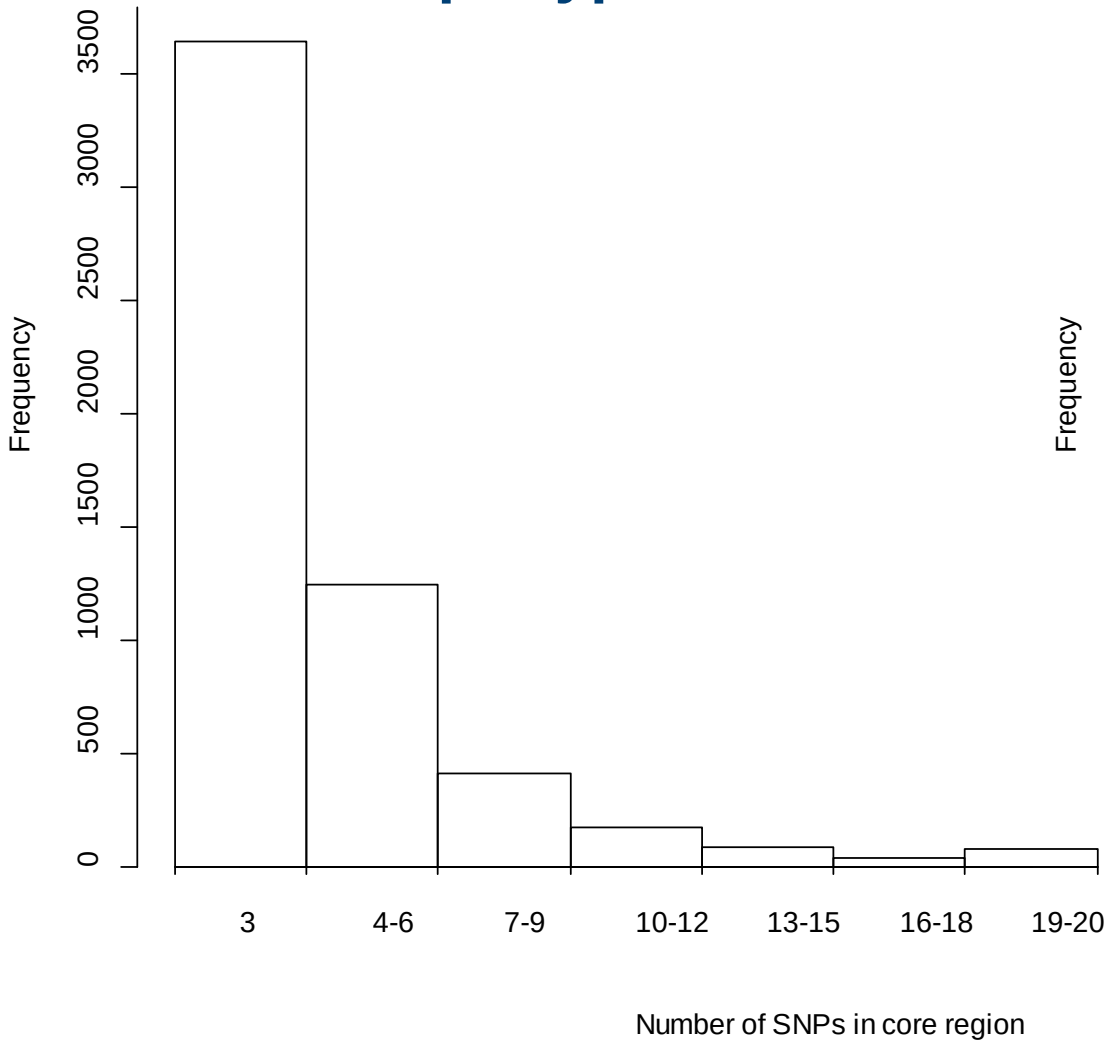
Results and Discussion

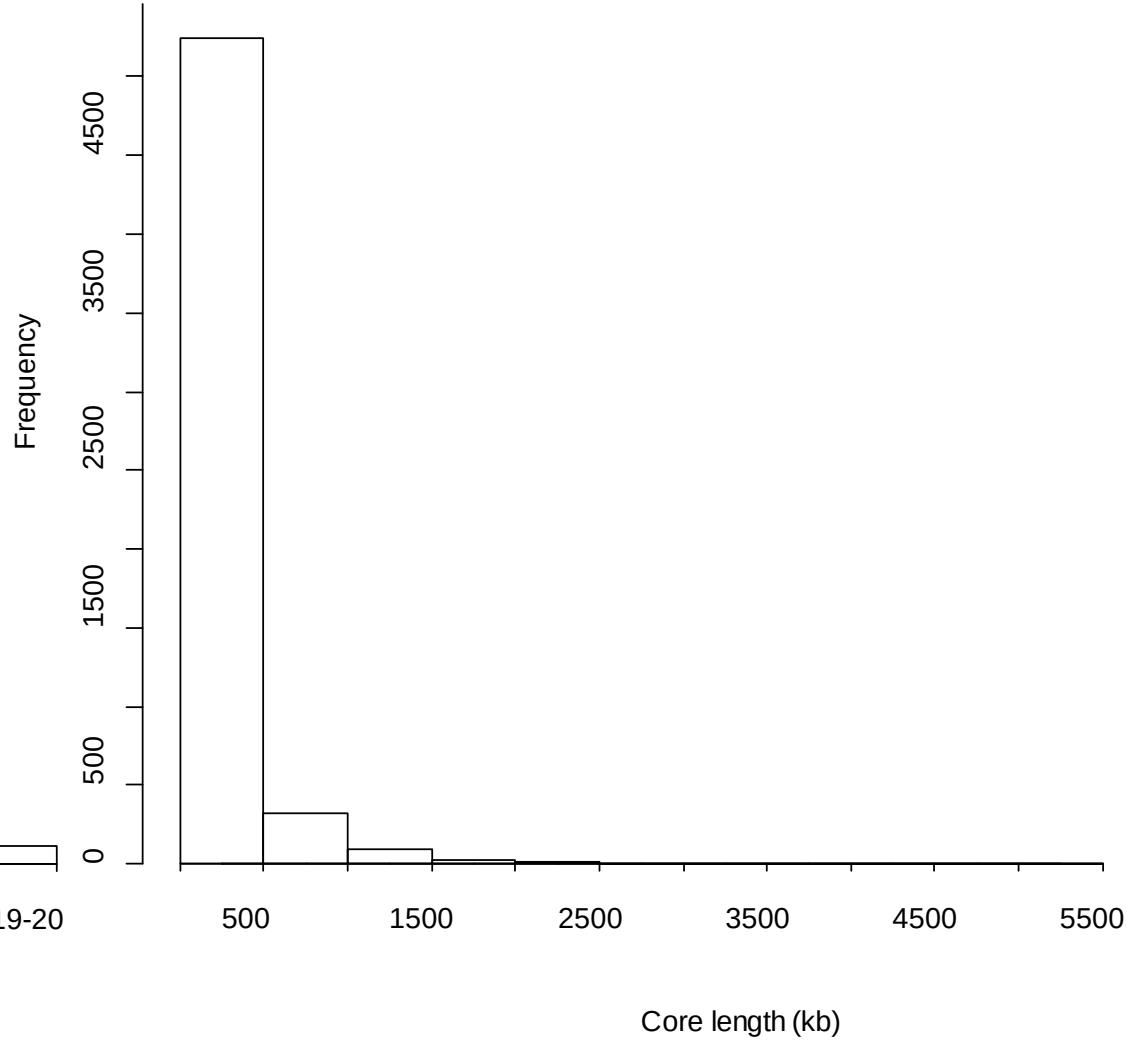
Pattern of haplotype blocks

- 22854 core haplotypes within 5700 core regions spanning 954 010 kb (32%) of the entire genome were detected.
- Mean core region length: 167.4 ± 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

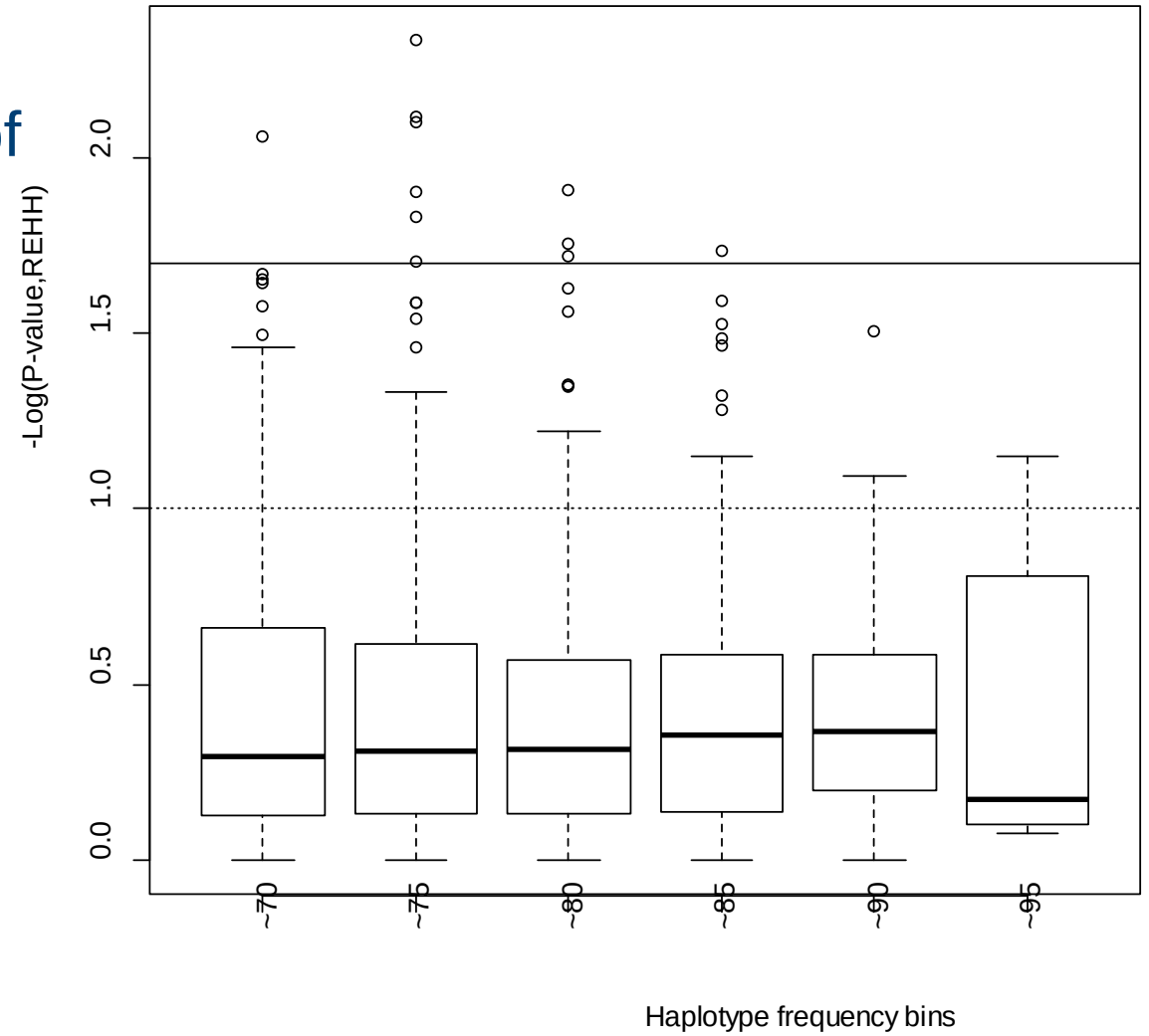




Results and Discussion

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



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:

Chr	Position	Core length (kb)	Hap freq	EHH	REHH	REHH *p-Value*	Candidate genes
8	128 652 415-128 724 560	72.15	0.75	0.51	3.25	0.00461	*PPP3CA*
18	31 531 915-31 769 165	237.25	0.76	0.38	2.53	0.01253	*CAV1/CAV2*
8	114 555 082-117 361 292	2806.21	0.76	0.82	2.42	0.01474	*CaMk2D/ANK2*

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 signficant 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 Stratz 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’).


UNIVERSITÄT HOHENHEIM
1818


UNIVERSITÄT HOHENHEIM
1818

Supplemental slides

Materials und Methods

'Extended Haplotype Homozygosity' test

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

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

- 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.