



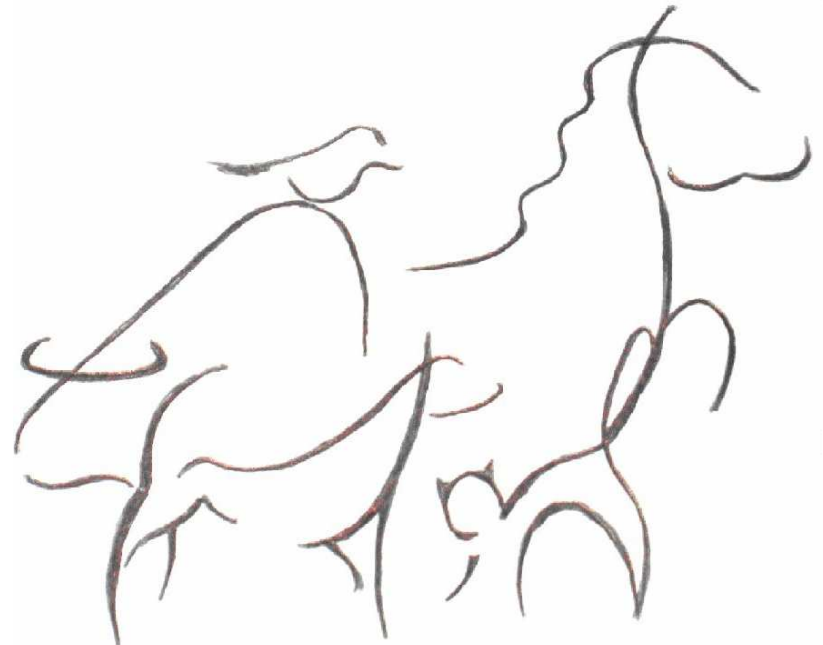
Using genome scan of DNA polymorphism to identify regions exhibiting positive selection

Animal Breeding and Genetics Group, Department of Animal Sciences,
Georg August University Göttingen, Germany

Saber Qanbari

QTL MAS Workshop, Poznan - Poland

17th May 2010





Background

Highly productive cattle breeds have been intensively selected during the last decades.

Consequently, genomic regions controlling traits of economic importance are expected to exhibit signatures of selective breeding.

Motivation:

This study aimed to scan the genome of a diverse set of cattle breeds for finding the positions that may have been targets of recent positive selection.

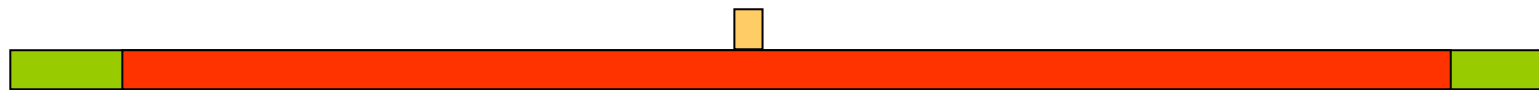


Signatures of selection-Theory

No Selection

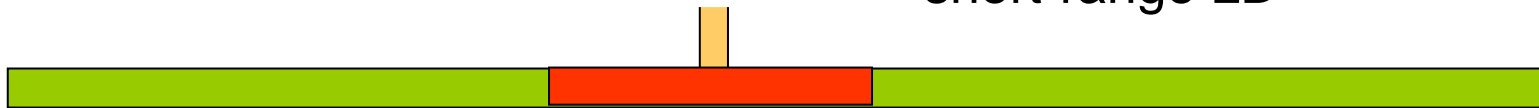
Young alleles:

- low frequency
- long-range LD (long haplotypes)



Old alleles:

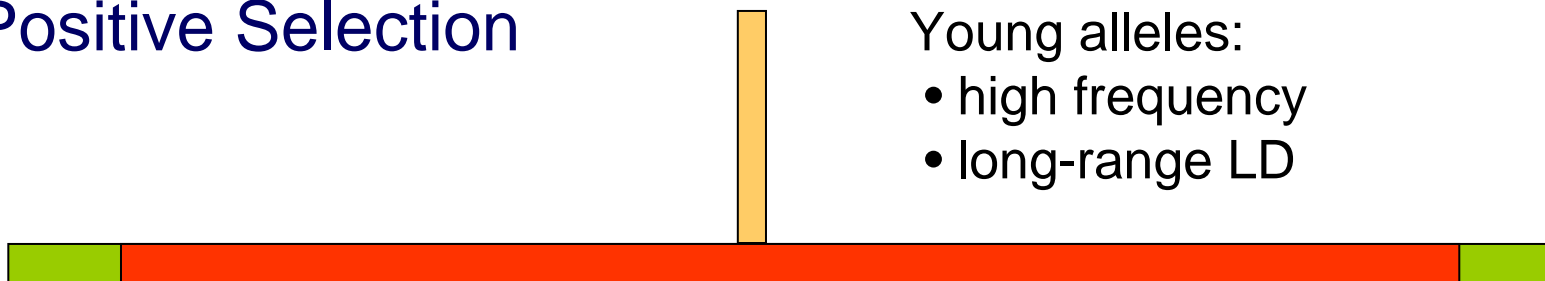
- low or high frequency
- short-range LD



Positive Selection

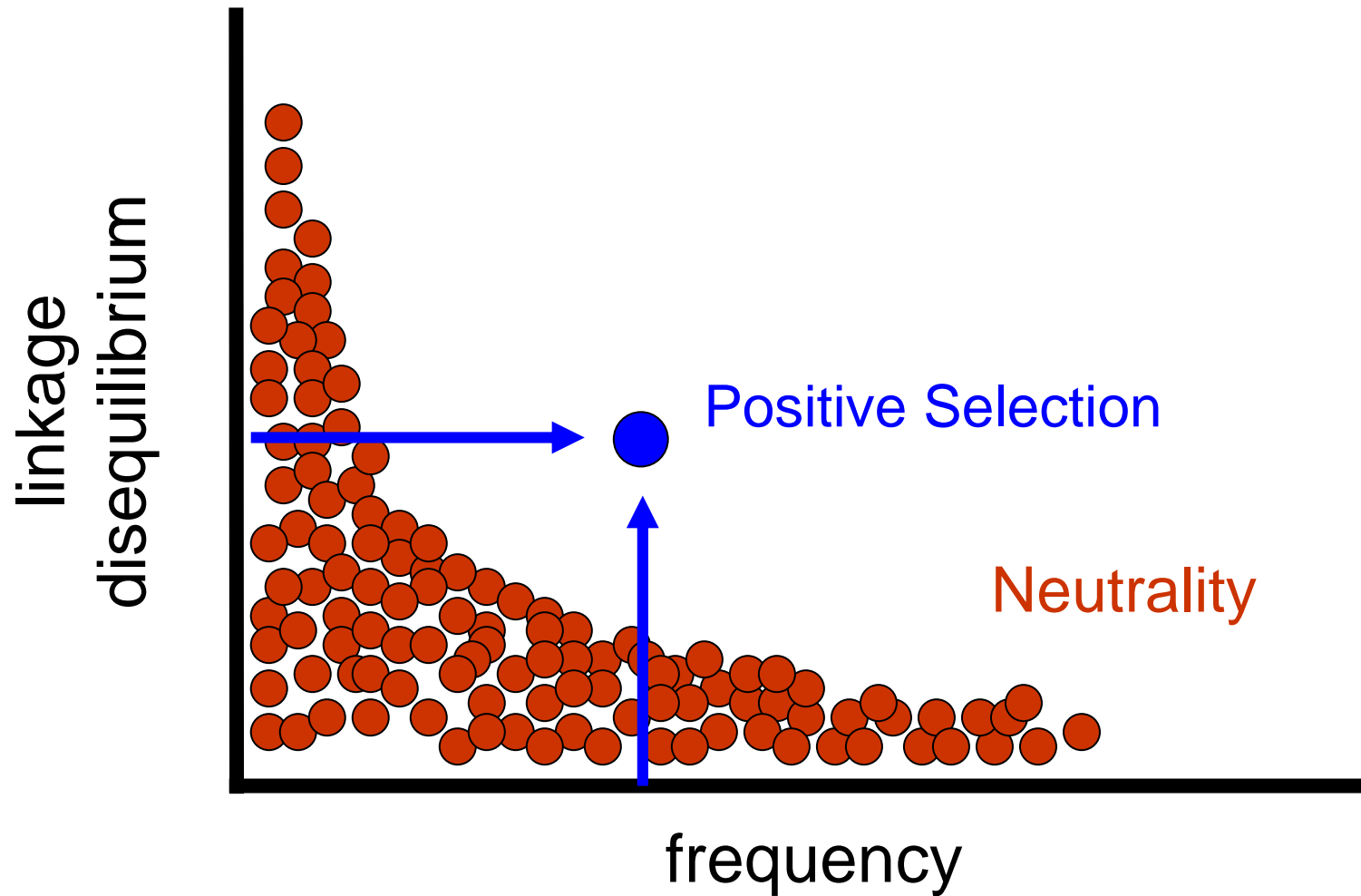
Young alleles:

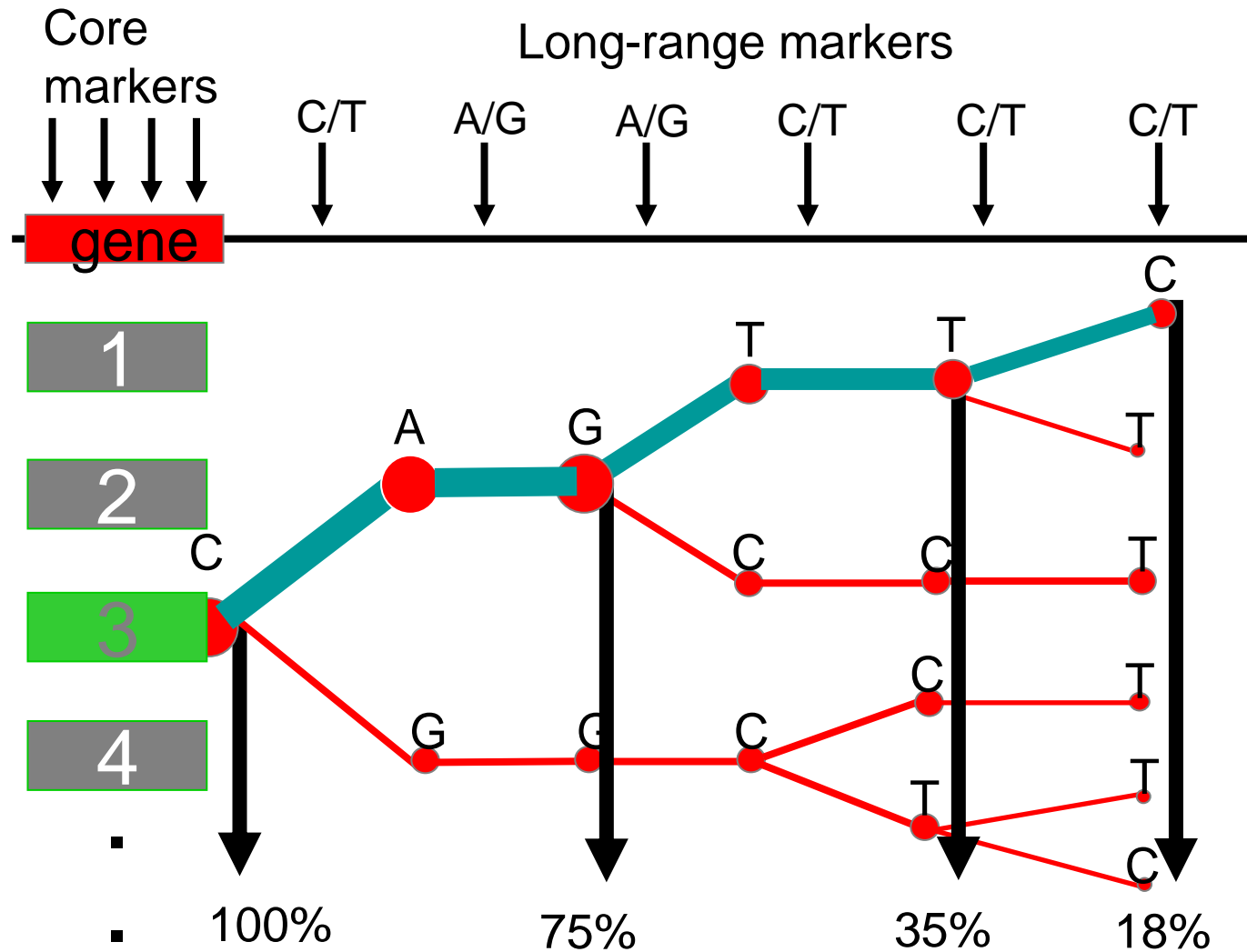
- high frequency
- long-range LD





Principle: Allele frequency vs. extent of LD





Extended Haplotype Homozygosity (EHH)

The probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical by descent for the entire interval from the core region to a distance x (*Sabeti et al. 2002*).



II. “integrated Haplotype Homozygosity Score” (iHS)

The **integrated EHH (iHH)** is denoted iHH_A or iHH_D , depending on whether it is computed with respect to the **ancestral** or **derived** core allele (Voight et al. 2006).

$$\text{unstandardized } iHS = \ln \left(\frac{iHH_A}{iHH_D} \right)$$

This quantity is standardized such that it has a mean of 0 and variance of 1 irrespective of allele frequency at the core SNP

$$iHS = \frac{\ln \left(\frac{iHH_A}{iHH_D} \right) - E_p \left[\ln \left(\frac{iHH_A}{iHH_D} \right) \right]}{\text{SD}_p \left[\ln \left(\frac{iHH_A}{iHH_D} \right) \right]}$$



Application of |iHS| test

Description of samples

Breed	Code	(n)	Country	Purpose
Holstein	HS	2091	Germany	Dairy
Brown Swiss	BS	277	Germany	Dairy
Simmental	SI	462	Germany	Dual-purpose
Angus	AN	232	Australia	Beef
Brahman	BR	80	Australia	Beef
Belmond Red	BE	166	Australia	Beef
Hereford	HR	158	Australia	Beef
Murray Gray	MG	57	Australia	Beef
Santa Gertrudis	SG	126	Australia	Beef
Shorthorns	SH	81	Australia	Beef



Genetic DATA

- Samples genotyped by Illumina 50K bead chip = 54'001 SNPs
- List of ancestor and derived alleles was used from (Matukumalli et al. 2009).

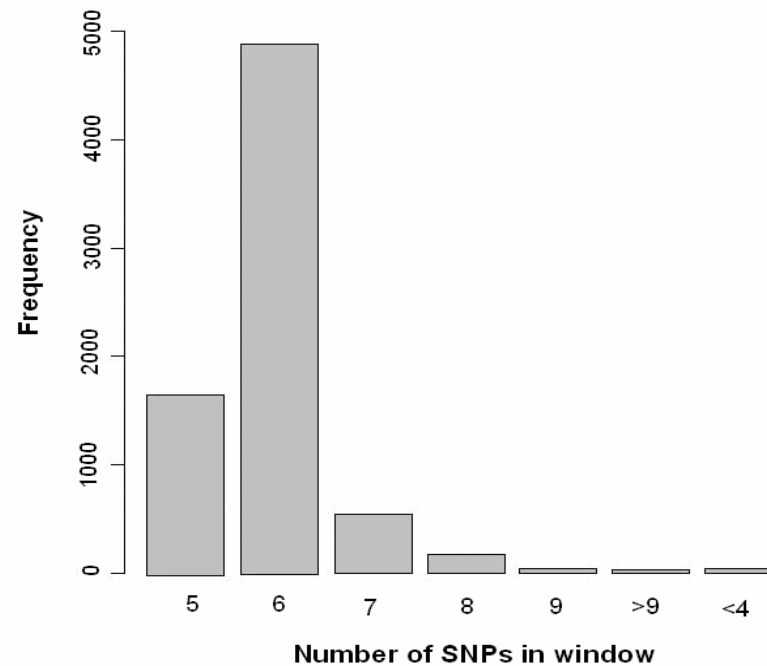
Markers were filtered to exclude

- Markers assigned to un-positioned contigs
- Loci with missing genotypes (>5%)
- Loci with a minor allele frequency (MAF) < 0.05



Calculation of $|iHS|$

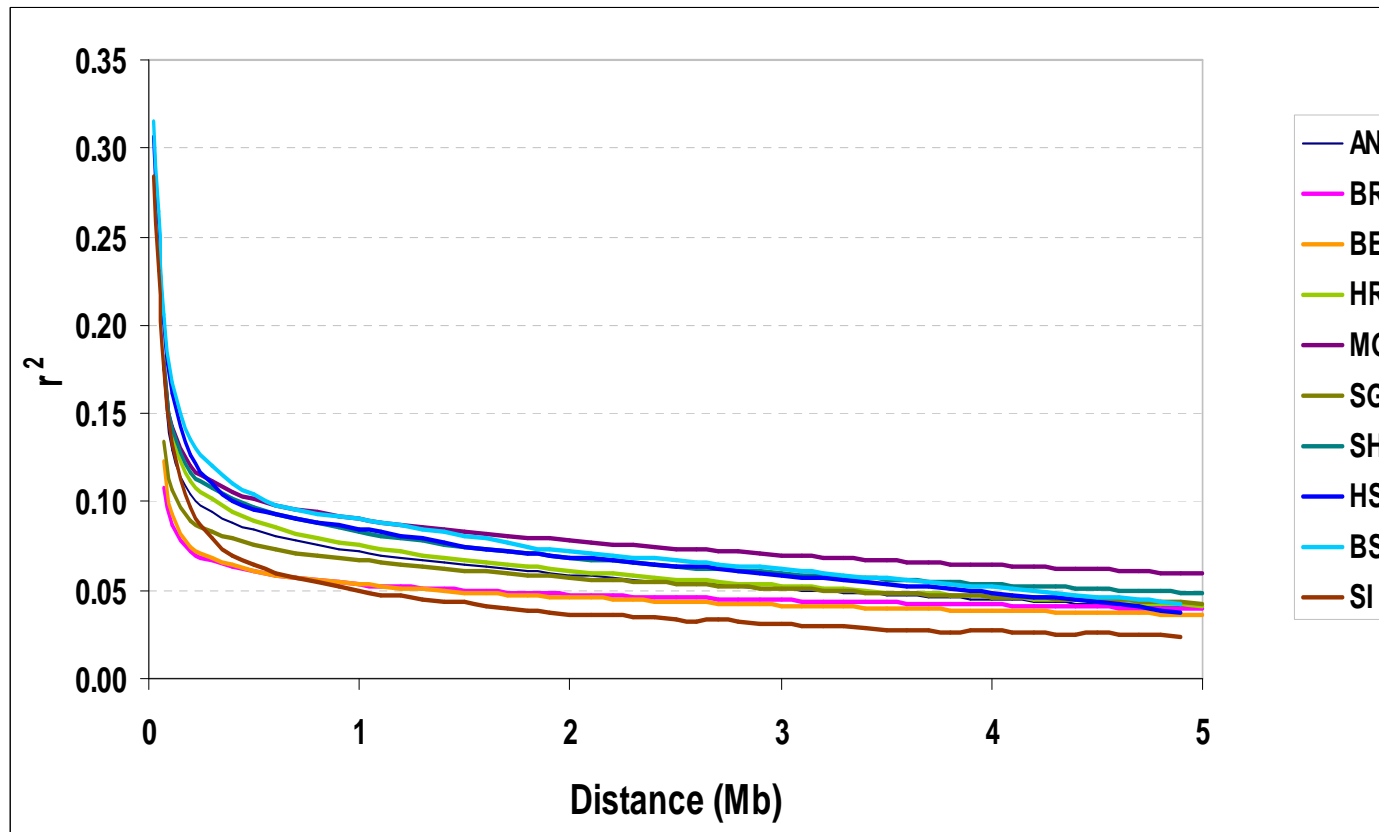
- $|iHS|$ was calculated for each SNP across the genome of the breeds.
- Genome was split into the segments of 500 kb and $|iHS|$ scores averaged over the SNPs located in each window.



Distribution of the SNPs in 500 kb windows sliding the genome of breeds.



Decay of LD as a function of inter-marker distance



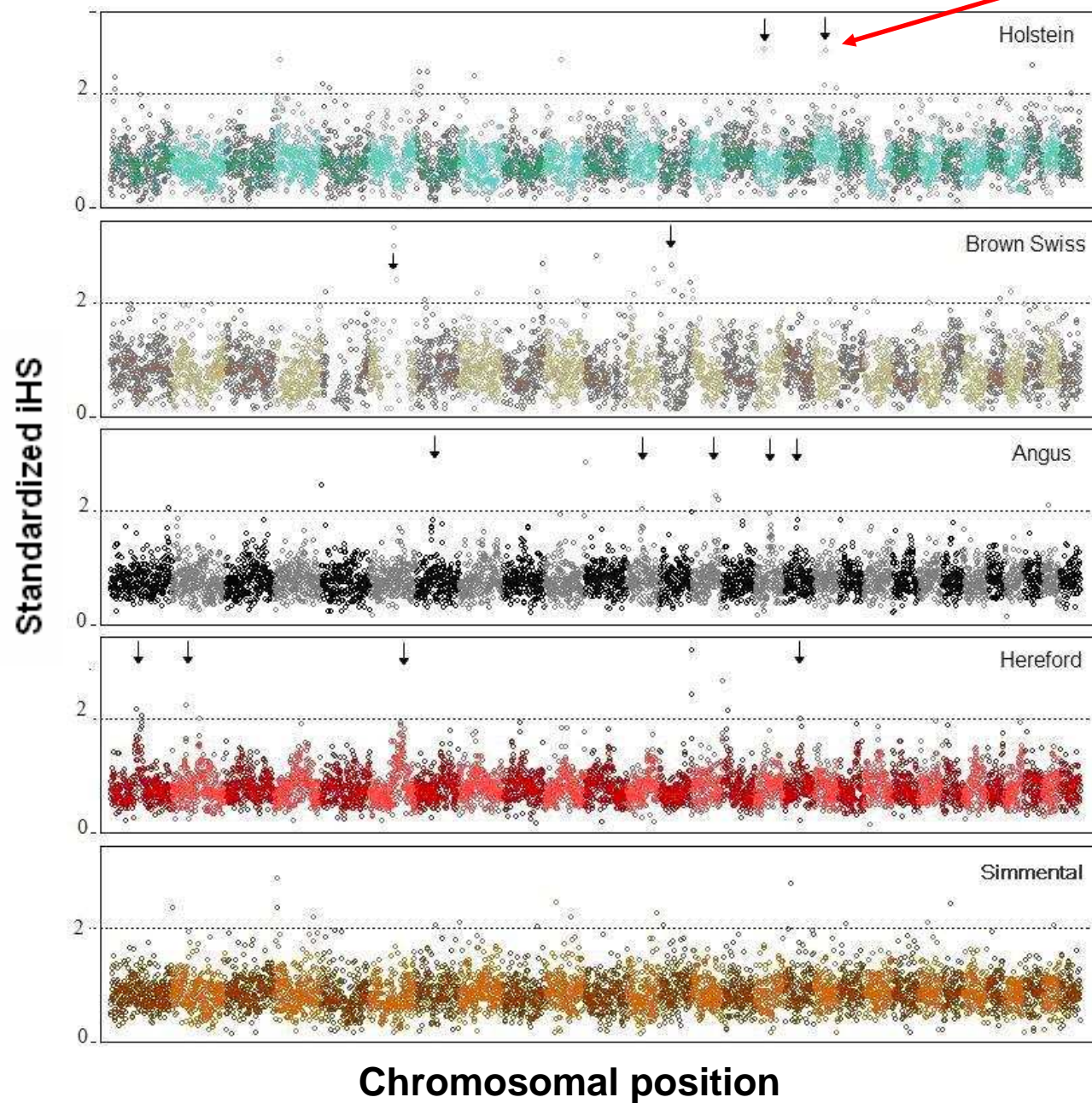


Some results from iHS analysis

- Final number of SNPs analyzed, ranged from 35'226 (BS) to 47'416 (BR)
- Mean MAF over breeds = 0.24 ± 0.03 and mean SNP distance = 60.3 K
- Overall mean $|iHS|=0.74$ and max $|iHS|_s = 3.41$
- $|iHS|_s$ in the extreme 5% of the empirical distribution (n=109 across breeds) were assumed as evidences of the positive selection.

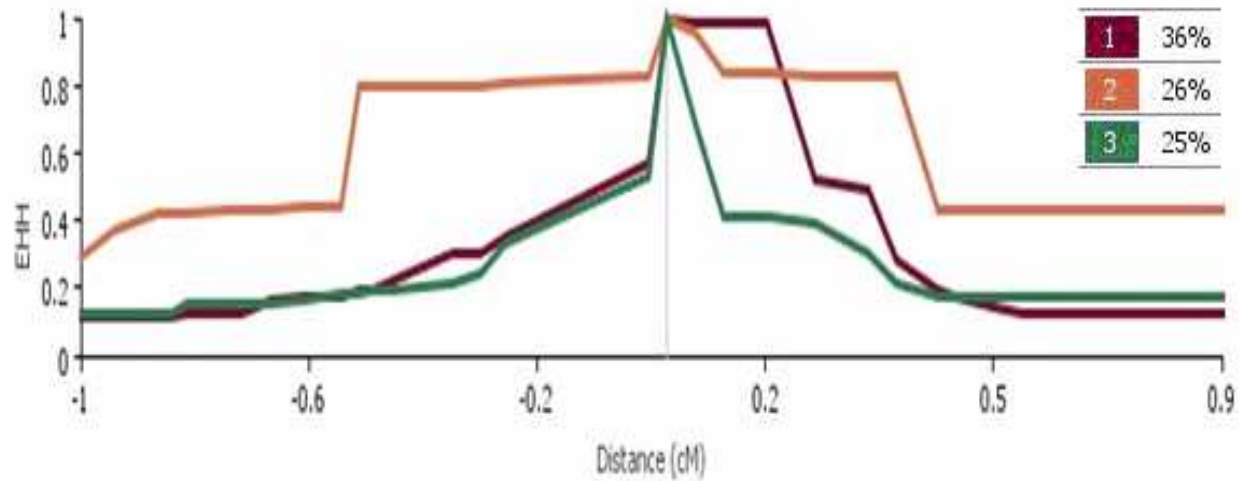


Distribution of $|iHS|$ signals across the genome



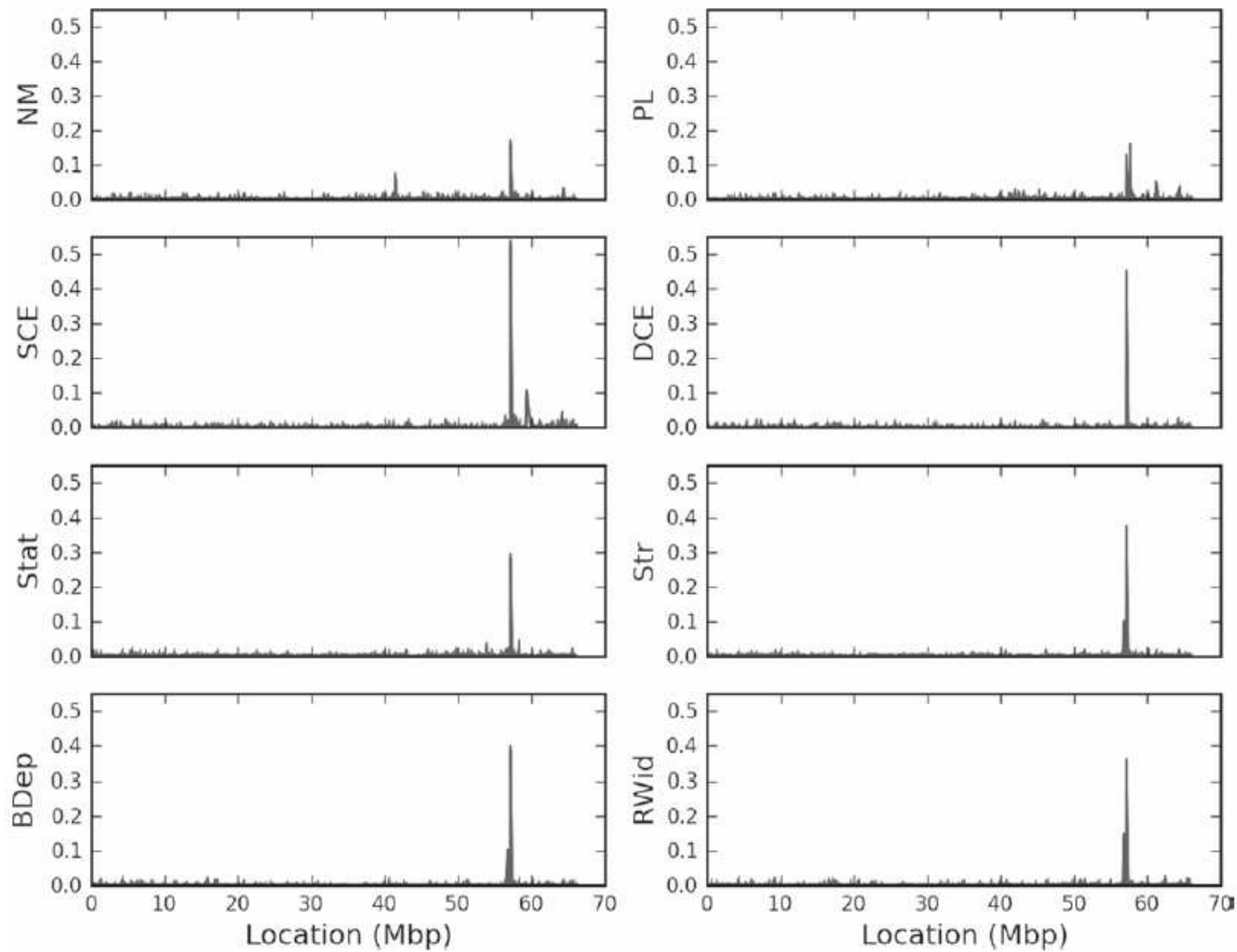


Extent of haplotype homozygosity in the interval 57.25–57.75 Mb on BTA18 in Holstein cattle revealed by $|iHS|$.





COLE ET AL.



Size (in additive genetic SD) and location of marker effects on *Bos taurus* autosome 18 affecting net merit (NM), longevity (PL), sire (SCE) and daughter calving ease (DCE), stature (Stat), strength (Str), body depth (BDep), and rump width (RWid).

Interrogating positively selected regions revealed by |iHS|



Chr	iHS	Breed	Gene	Function
18	2.2	HS	SIGLEC8	Sialic acid binding Ig-like lectin 8 and 10
16	2.6	HS	SPATA17	Spermatogenesis associated 17
6	3.41	BS	UGDH	UDP-glucose dehydrogenase
13	2.68	BS	TRDMT1	Cysteine and methionine metabolism
1	2.10	HR	SST	Somatostatin
2	2.26	HR	GCG FAP	Glucagon Fibroblast activation protein, alpha
6		HR	SRD5A2L	Converts testosterone into dihydrotestosterone
7	1.9	AN	COL23A1 MGAT1	Collagen, type XXIII, alpha 1 Fertilization and early development of the embryos
12	2.03	AN	ATP12A	ATPase activity
14	2.02	AN	MATN2	Developing cartilage rudiments
16	1.98	AN	NMNAT1	Methylene tetrahydrofolate reductase (NADPH) activity
17	2.05	AN/HR	PGRMC2	Progesterone receptor membrane component 2
2	2.06	MG/BE/ SH/BR	-	-
10	2.24	BE/SH	ACTC1	Actinin, involved in the formation of filaments



Conclusions

1. We present a genome-wide maps of iHS signal in cattle genome.
2. Our results confirmed the longer than expected haplotype consistency in the vicinity of SIGLEC5 gene on BTA18 which was recently reported as a strong QTL in the Holstein cattle (Cole et al. 2009).
3. Many of the regions showing extreme values for the metric used seem to play important roles in economically important traits in cattle (specially reproduction).
4. We conclude that high-resolution genome scan using dense markers is capable to identify selection signature.



REMAINING CHALLENGES

- Additional studies with a higher SNP density are needed to confirm our results.
- The metric used lacks a proper testing philosophy.
- Efficient methods to differentiate effects of drift and selection must be developed.
- The causal targets driving the signature of selection must be Identified.



ACKNOWLEDGEMENTS

This study is part of the project FUGATO-plus GenoTrack and was financially supported by the German Ministry of Education and Research, BMBF, the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, and Lohmann Tierzucht GmbH, Cuxhaven. I thank to H. Wilhelm Schaumann Stiftung, Hamburg for financial support.



Bundesministerium
für Bildung
und Forschung



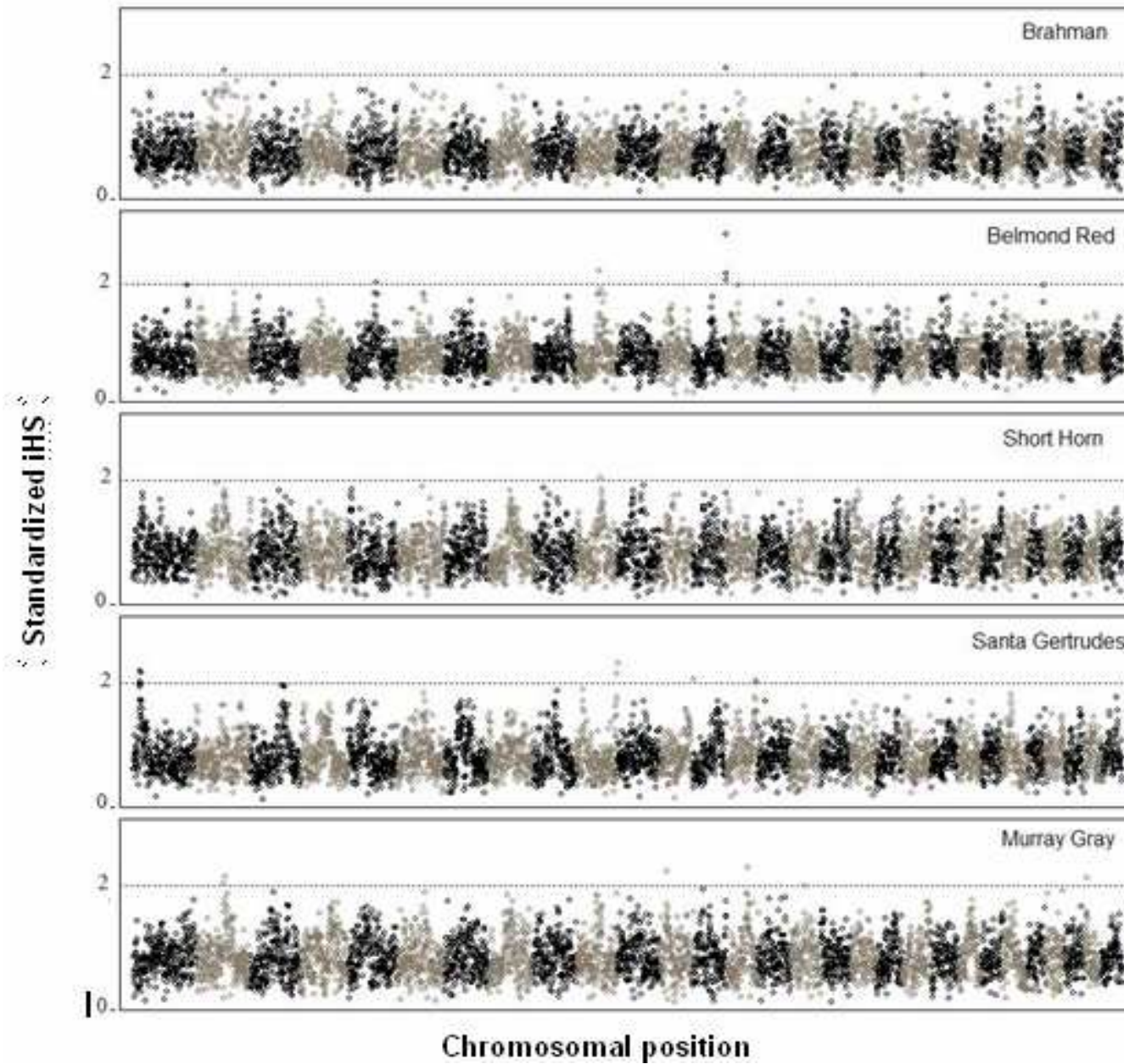
LOHMANN
TIERZUCHT



Thank you for attention !



Distribution of iHS signals across the genome



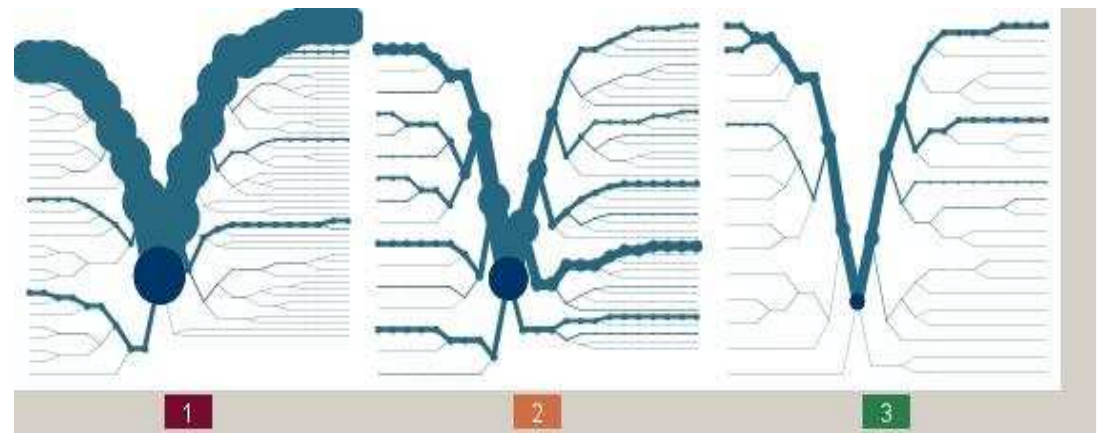
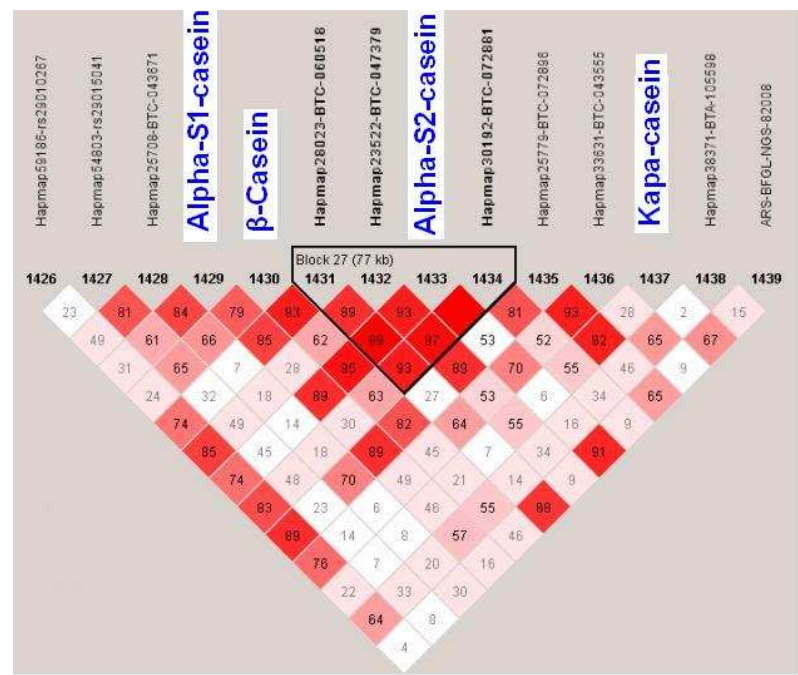
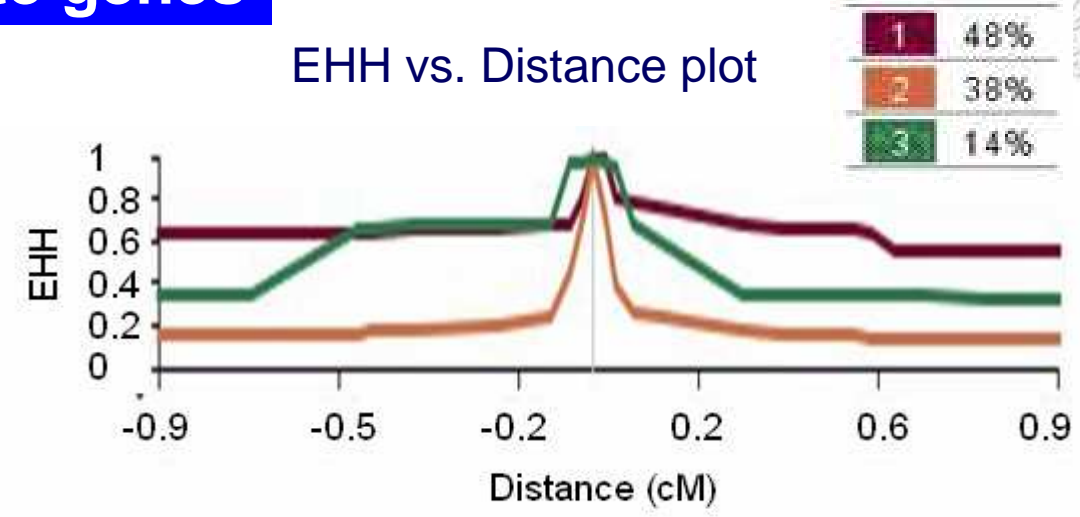


1. EHH test on candidate genes

Casein cluster :
A graphical overview

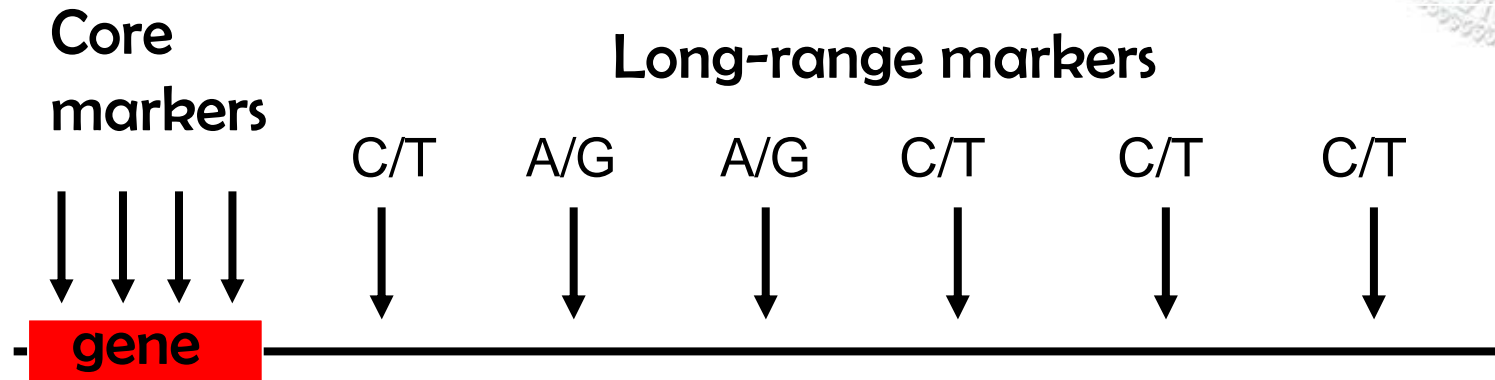
Casein Cluster as Core Region

EHH vs. Distance plot

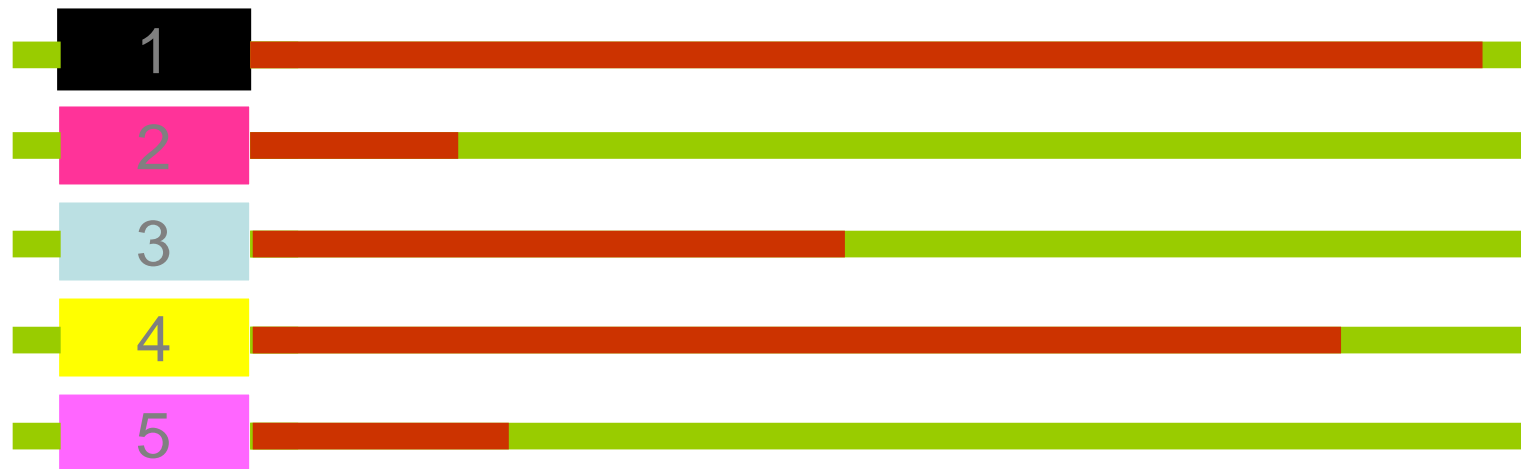


Haplotype bifurcation plots

Long-range multi-SNP haplotypes



Core Haplotypes



Decay of LD

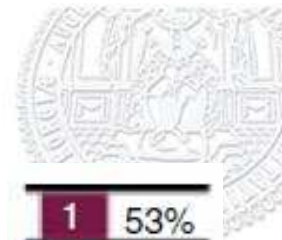
Software

We used:

- **R** for Data Processing
- **fastPHASE** (Scheet & Stephens, et al; 2001) for haplotype reconstruction of data.
- **HAPLOVIEW v4.1** (Barrett *et al.* 2005) for LD statistics and plotting the LD blocks and.
- **SWEEP** (Sabeti, 2002) for measuring positive selection based on Extended Haplotype Homozygosity (EHH) approach.

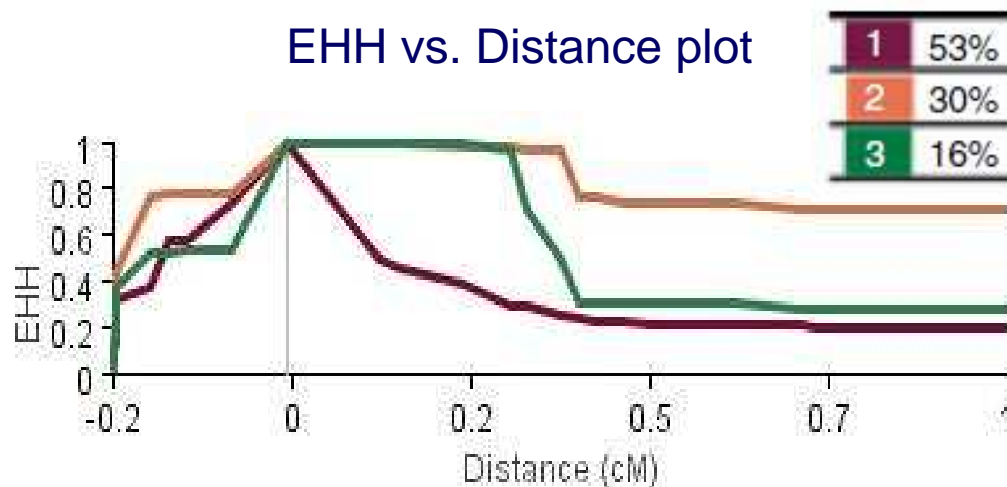


I. Application of EHH test on Holstein genome

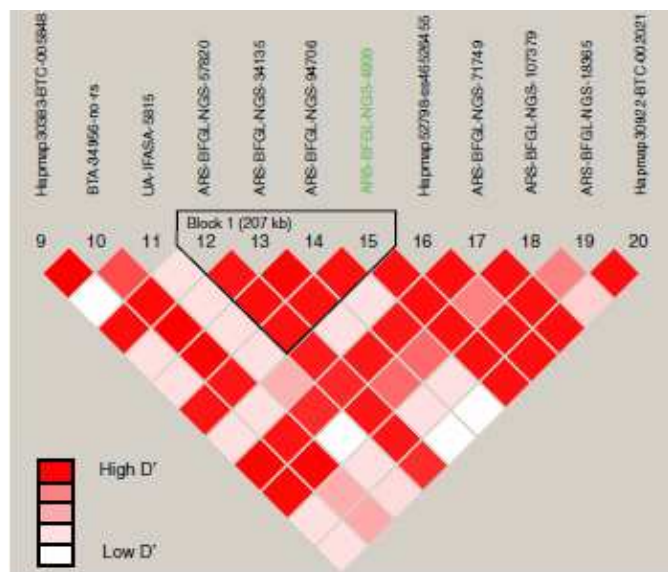


DGAT1 gene:
A graphical overview

EHH vs. Distance plot



DGAT1 as Core Region



Haplotype bifurcation plots

