

# Application example: IDENTIFICATION OF GENOMIC REGIONS SUBJECT TO SELECTION

**Juha Kantanen**

**Natural Resources Institute Finland**

**[juha.kantanen@luke.fi](mailto:juha.kantanen@luke.fi)**





# Selection patterns and animal genetic resources

---

- Domestication, adaptation, breed formation and selective breeding have left specific genetic patterns, “selection signatures,” in genomic regions of farm animal breeds.
- The identification of selection signatures has importance in characterization of animal genetic resources.
- Breeds displaying special genetic adaptive traits have typically high priority in the conservation of animal genetic resources.
- The selection signature studies can provide knowledge of quantitative trait loci for production characters and important causal mutations.
- Genome-wide selection-mapping scans may promote our understanding of function of genomes (selection has operated in non-coding genomic regions too).

# Effects of selection on genomes in animal populations

## The various forms of selection have specific effects

- ✓ *Allelic and genotypic frequencies*: positive selection increases frequency of an advantageous allele while negative selection removes unfavorable alleles (mutations). Balancing selection, maintains multiple alleles in a locus and high genetic diversity.
- ✓ *Hard sweep*: A beneficial mutation arises and rapidly increases in a frequency.
- ✓ *Soft sweep*: A beneficial mutation has existed already in a population and increases in a frequency.
- ✓ *Hitchhiking*: An advantageous allele approaches fixation, other alleles (neutral and even mildly deleterious) near the selected allele may sweep too.
- ✓ *Linkage disequilibrium*: The population can be left with highly correlated alleles in regions of the chromosome, leading to long haplotypes.
- ✓ The identification of chromosomal regions *with this structure* is the basis for detecting sweeps.

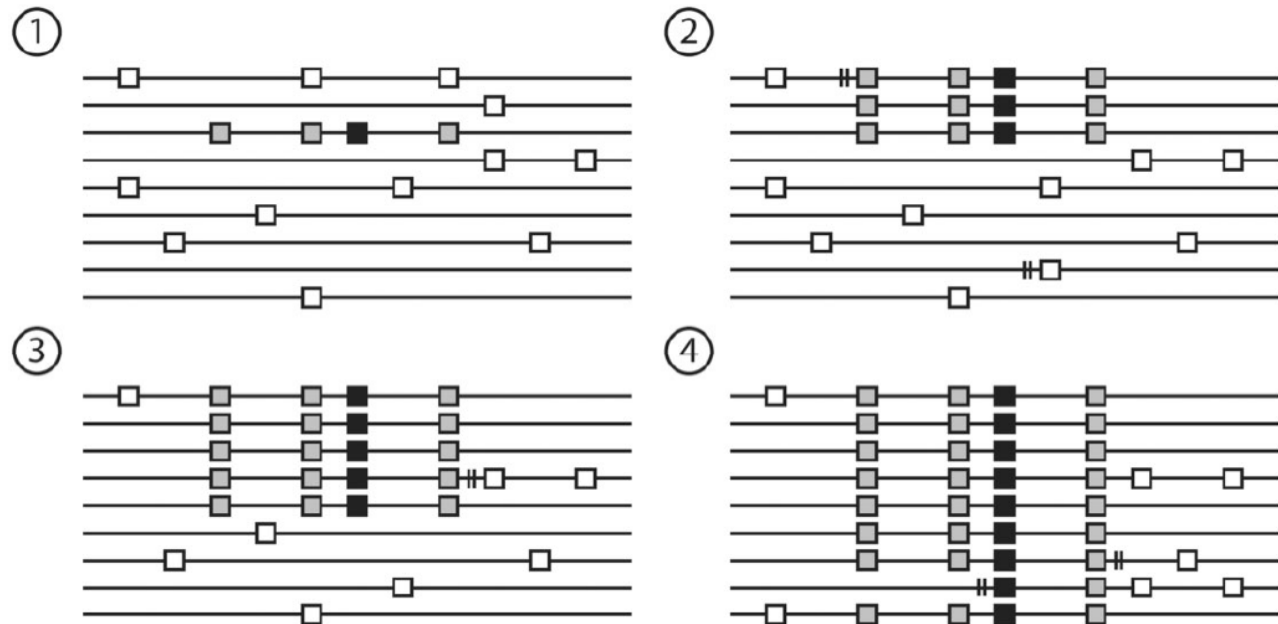
# The impact of selection on a 'population' of nine chromosomes

Step 1: A beneficial mutation (black square) occurring on the 3<sup>rd</sup> chromosome.

Step 2: The mutation begins to spread through the population, with nearby variants (grey) hitchhiking along with it.

Step 3: Haplotypes with the beneficial mutation becoming longer.

Step 4: Fixation of the beneficial mutation.



# Methods of detecting selected loci

- The methods commonly used in livestock studies can be classified into two main groups:
  - methods based on **intra-population** statistics (genomic or DNA-marker data are compared within a population); and
  - those on **inter-population** statistics (genomic or DNA-marker data are compared between populations).
- **The intra-population** statistics methods are based on the site frequency spectrum (the *site frequency spectrum* is the distribution of the allele frequencies of a given set of loci, e.g. SNPs, in a population or sample), LD or the identification of genomic regions with reduced variation compared to the genome average.
- **The inter-population** approaches are based on allele frequency differences and the degree of differentiation among the populations. The level of genetic differentiation can be analysed using single site differentiation or haplotype-based differentiation.

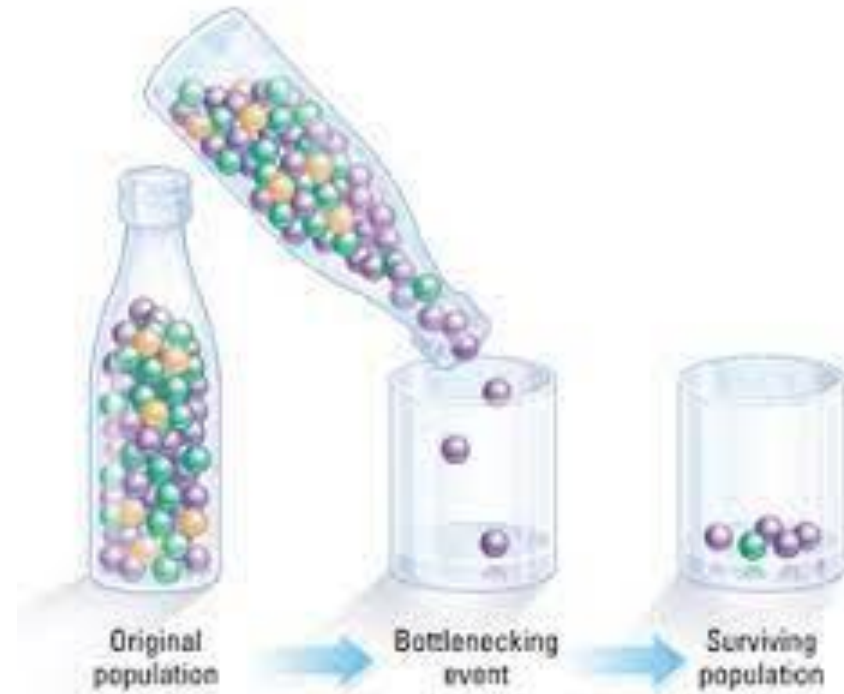



**In the FAO Guidelines, examples of tools for detecting sweeps are given**

Program/tool	Method	Reference
<b>Intra-population statistics</b>		
SweeD	Site frequency spectrum (SFS)	Pavlidis P. et al. 2013.
Sweep	relative extended haplotype homozygosity (rEHH)	Sabeti P.C. et al. 2002.
Selscan	Extended Haplotype Homozygosity (EHH), Integrated Haplotype Score (iHS)	Szpiech Z.A., Hernandez R.D. 2014.
cgaTOH	Runs of homozygosity (ROH)	Zhang L, Orloff MS, Reber S, Li S, Zhao Y, Eng C (2013) cgaTOH: Extended Approach for Identifying Tracts of Homozygosity. PLoS ONE 8(3): e57772.
<b>Inter-population statistics</b>		
HierFstat	$F_{ST}$	de Meeûs T., Goudet J. 2007.
Selscan	Cross population extended haplotype homozygosity (XP-EHH)	Szpiech Z.A., Hernandez R.D. 2014.
HapFLK	Haplotype-based extension of the FLK statistics (hapFLK), an extension of the Lewontin and Krakauer (LK) test considering a hierarchical population structure (FLK)	Fariello M.I. et al. 2013. Bonhomme et al. 2010. Lewontin R., Krakauer J. 1973.

# Genomic signals which may not be signatures of selection

- Frequent recombinations may confound efforts to identify sweeps.
- In animal populations, demographic events such as population expansions, migrations, genetic bottlenecks and population subdivision can lead to false selection signals that mimic signatures of selection.





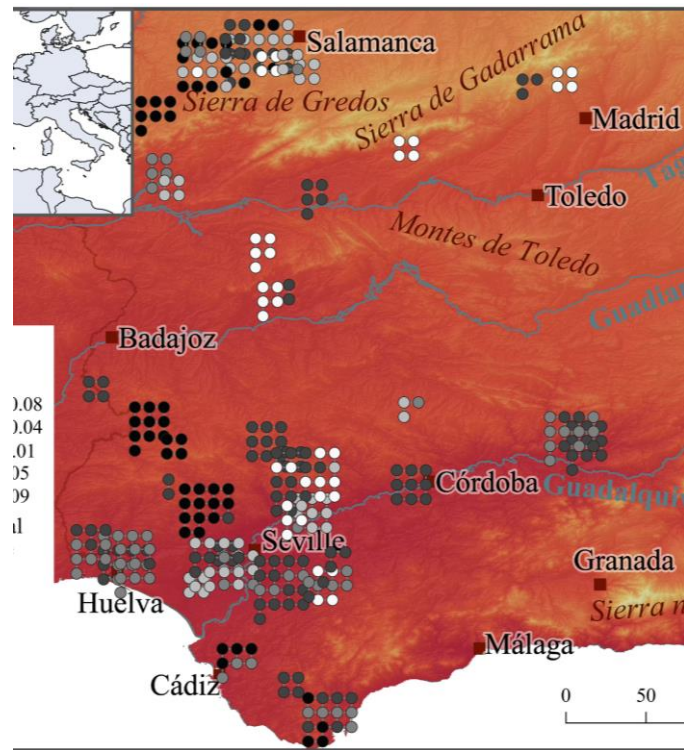
# Composite methods

---

- It is postulated that a composition method combining distinct approaches into a single metric may improve the resolution in detecting sweeps.
- For example, a composite evaluation test,  $\mu$  statistic for detection of positive selection, which examines genomic regions by quantifying the site frequency spectrum, the levels of LD and the amount of genetic diversity along a chromosome (the tool *RAiSD*; Alachiotis N. & Pavlidis P. 2018. *Communications Biology* 1: 79.)



# Landscape genetics/genomics



- The landscape genomics approach consists of linking genome-wide information to environmental variables to identify valuable genomic regions associated with adaptive advantages.
- Data obtained from genome-wide scans carried out on a number of animals from populations living in different habitats or across ecological clines will be compared to geo-environmental information characterizing these habitats (e.g. yearly amount of precipitation, monthly temperatures, number of days with ground frost, etc.).
- Example of tools: R.SamBada, LFMM2, BAYESCENV

# Bibliography

Ahrens C.W., Rymer P.D., Stow A. et al. 2018. The search for loci under selection: trends, biases and progress. *Molecular Ecology* 27: 1342– 1356.

Bank C., Ewing G.B., Ferrer-Admettla A. et al. 2014. Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends in Genetics* 30 (12): 540-546,

de Simoni Gouveia J.J., da Silva M.V., Paiva S.R., de Oliveira S.M. 2014. Identification of selection signatures in livestock species. *Genetics and Molecular Biology* 37: 330-42.

Horscroft C., Ennis S., Pengelly R.J. et al. 2019. Sequencing era methods for identifying signatures of selection in the genome, *Briefings in Bioinformatics* 20: 1997–2008.

Passamonti M.M., Somenzi E. Barbato M. et al. 2021. The Quest for Genes Involved in Adaptation to Climate Change in Ruminant Livestock. *Animals* 11: 2833.

**Thank  
you!**



# You can find us online

 [luke.fi](https://luke.fi)

Subscribe to our newsletter to stay informed!  
[luke.fi/newsletter](https://luke.fi/newsletter)



Natural Resources Institute Finland (Luke)  
Latokartanonkaari 9, FI-00790 Helsinki

