ISAGREED

Fine-scale analysis of population structure based on genomic data and quantification of selection effect on livestock genome

Modul no. 2: Conservation and Sustainable Use of Animal Genetic Resources

Nina Moravčíková, Radovan Kasarda
Slovak University of Agriculture in Nitra
Faculty of Agrobiology and Food Resources

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED

Co-funded by the European Union

Content of the lecture

- Quality control of genomic data
- Approaches and tools for population structure analysis
- Approaches and tools for evaluating the impact of selection on the livestock genome
- Functional annotation of regions significantly affected by selection pressure

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

Quality control of genomic data

SNP data (genotyping using SNP chips)

- QC is a crucial step to ensure the accuracy and reliability of the results of subsequent analyses
- Incorrect or low-quality data can lead to errors in analyses and interpretation of results
- data quality indicators – call rate of SNP markers overall and within individuals, frequency of the minor allele frequency, deviation from the Hardy-Weinberg equilibrium, degree of linkage disequilibrium ...
- the choice of indicators used in QC depends on the objective/type of follow-up analyses

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Quality control of genomic data

SNP data (genotyping using SNP chips)

- Standard QC for population structure analysis:
  - call rate across SNP markers – min. 90%
  - call rate across animals – min. 90%
  - minor allele frequency (MAF) – min. 1% in the population
  - Hardy-Weinberg equilibrium - 1 × 10–6
  - Software tools: e.g. PLINK

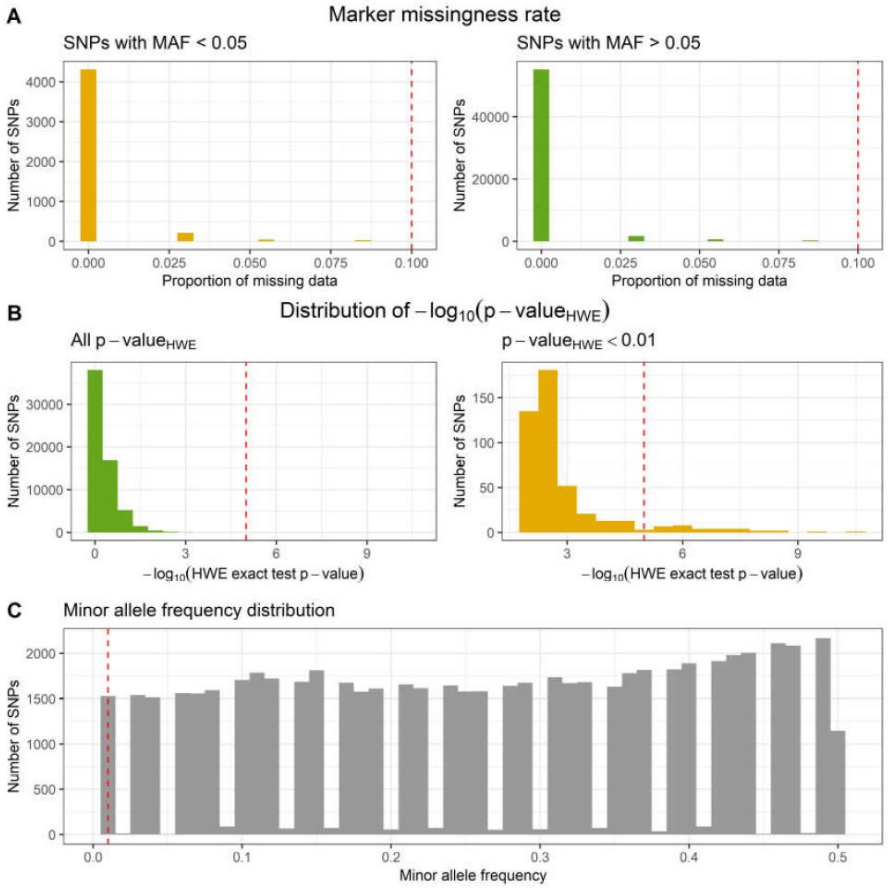


Fig. 1: Graphical visualization of SNP data quality control (Moravčíková et al., 2020)

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

SNP data (genotyping using SNP chips)

The study of population structure using genomic data allows:

- analyse genetic differentiation within and between populations
- evaluate the degree of their genetic admixture as well as changes in their gene pool, which have arisen, for example, due to selection, migration or genetic drift
- estimate the genomic relationship matrix and optimize mating plans

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

SNP data (genotyping using SNP chips)

The most common types of analyses:

- calculation of the Wright's F_{ST} index,
- calculation of genetic distances and relationship matrices
- Principal Component Analysis
- Discriminant analysis of principal components
- Bayesian analysis of genetic admixture and gene flow between populations
- construction of phylogenetic trees and genetic networks

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

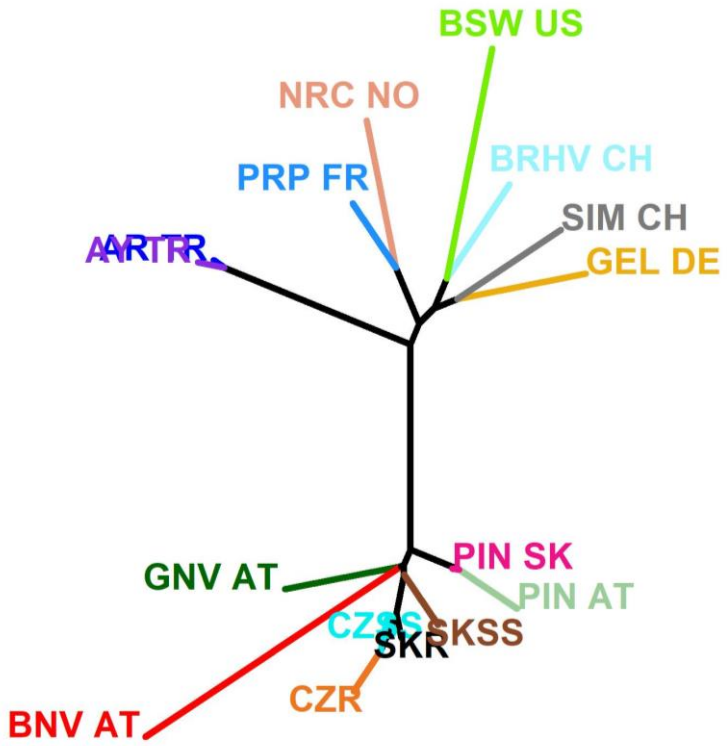
Wright's F_{ST} index

- is one of the most commonly used indicators to determine the degree of genetic differentiation between populations
- its value ranges from 0 (populations are genetically identical) to 1 (populations are genetically completely differentiated)
- advantages: ease of interpretation, fast computational method for detecting differentiation between populations
- disadvantages: cannot be used at the individual level, lower reliability in populations with a low level of diversity

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis



Wright's F_{ST} index

- Software tools: Arlequin, Genepop and GenAlEx (limited number of SNP markers), R (e.g. StAMPP packages)

Fig. 2: Dendrogram constructed on the basis of the F_{ST} matrix showing the genetic relationships between 16 cattle breeds cattle (unpublished results)

AR_TR - Anatolyan Red, AY_TR - Anatolyan Yellow, BNV_AT - Braunvieh (Austria), BNV_CH - Braunvieh (Switzerland), BSW_US - Brown Swiss, CZR - Czech Red, CZSS - Czech Spotted, GEL_DE - Gelbvieh, GNV_AT - Tyrol Grey, NRC_NO - Norwegian Red, PIN_AT - Pinzgau (Austria), PIN_SK - Slovak Pinzgau (Slovakia), PRP_FR - French Red Pied, SIM_CH - Simmental, SKR - Slovak Red, SKSS - Slovak Spotted

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

Relationship matrices

- express genetic similarities and relationships between individuals within a population
- each element of the matrix represents a measure of genetic similarity between a pair of individuals
- are most often calculated based on the frequency of alleles, while the calculation itself can be based on various approaches, e.g. calculation of the IBD (identity by descent) matrix or Nei's genetic distances
- advantages: relatively accurate estimates of genetic relationships, suitable for the study of intrapopulation relationships
- disadvantages: the calculation is time-consuming in the case of large databases

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

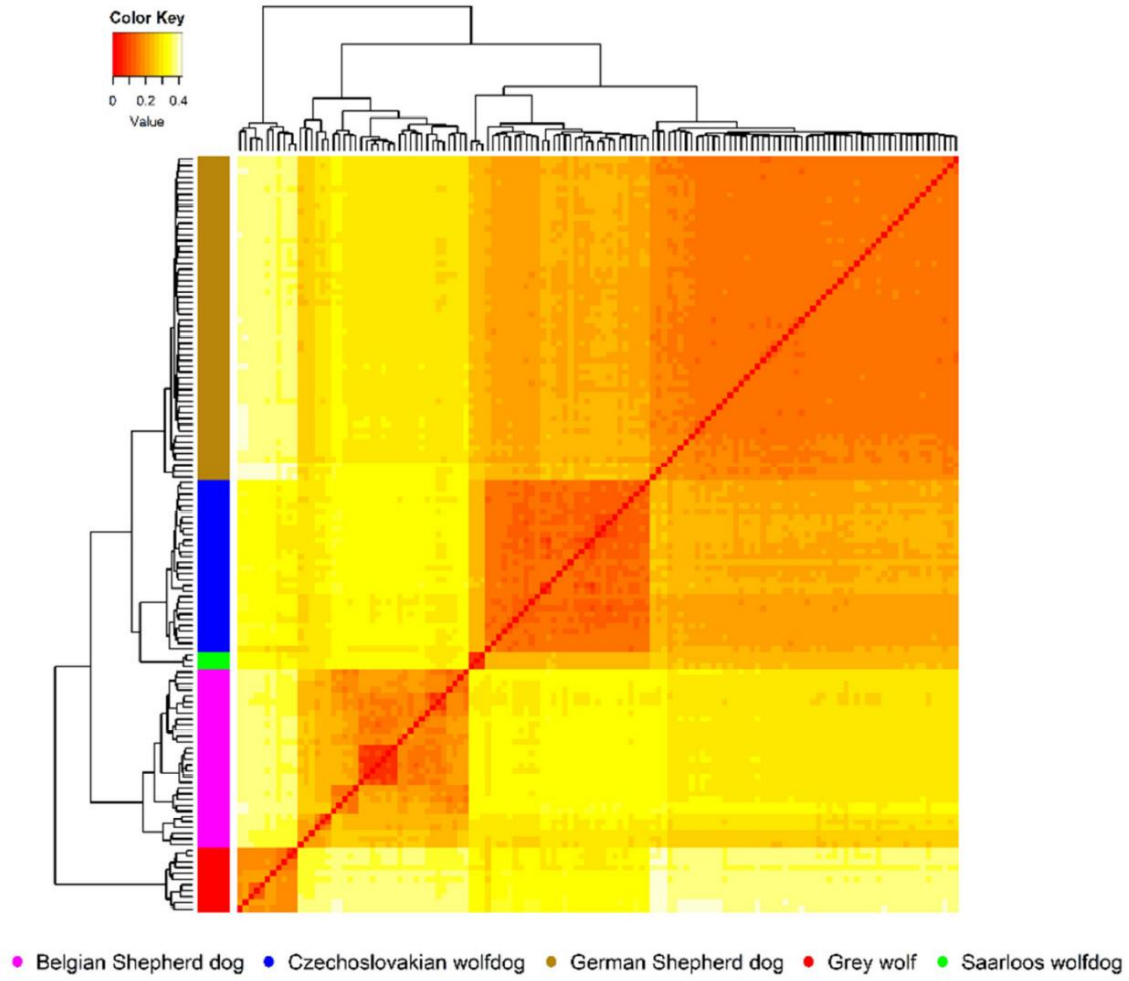


Fig. 3: Intrapopulation genetic structure within 5 dog breeds derived from a Nei's genetic distance (Moravčíková et al., 2021)

Relationship matrices

- Software tools: PLINK, R (e.g. StAMPP package)

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

Principal component analysis (PCA)

- a multivariate statistical method that decomposes a covariance matrix of genetic data and extracts the principal components that reflect the variability of the data in the dataset
- the first two principal components usually capture the highest proportion of variability
- it provides basic information about the genetic structure, which is useful when testing databases with a large number of individuals
- advantages: time-saving method for assessing the state of genetic differentiation, simple and easily interpretable visualization
- disadvantages: low sensitivity for genetic admixture

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

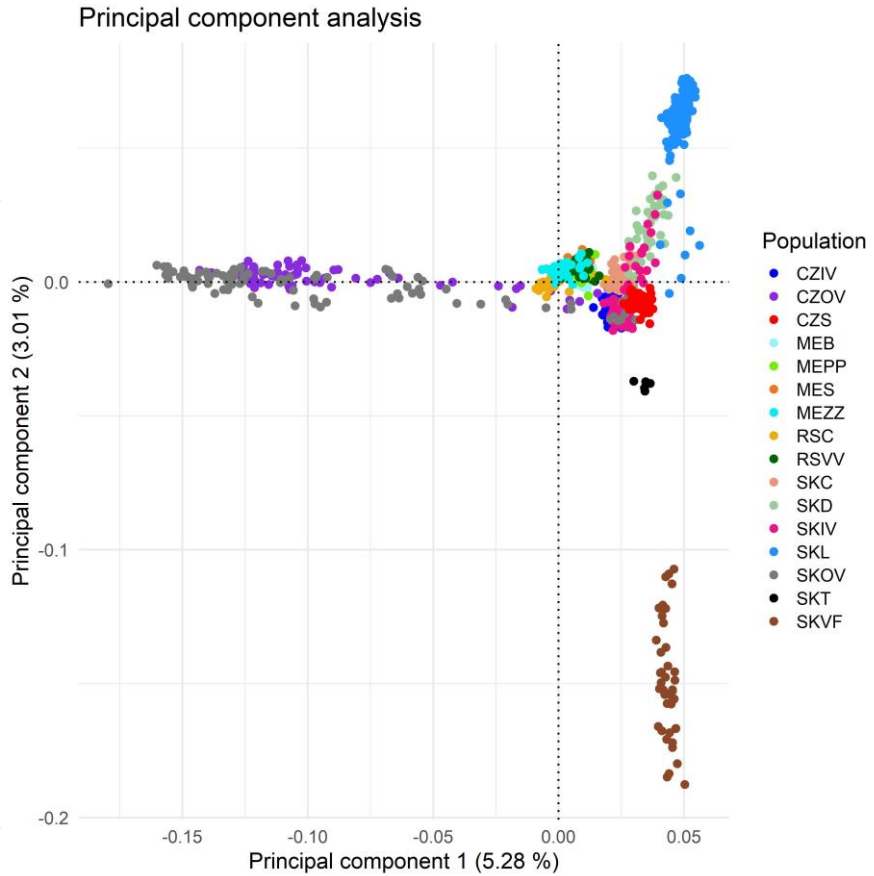


Fig. 4: Genetic differentiation within 16 sheep breeds based on visualization of the first and second principal components of the PCA (unpublished results)

Principal component analysis (PCA)

- Software tools: PLINK, R (e.g. adegenet package)

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

Discriminant analysis of principal components (DAPC)

- DAPC is a method of discriminant analysis aimed at visualizing the genetic structure between predefined groups or clusters. It uses PCA to reduce the dimension of data and then discriminant analysis to maximize the resolution between populations
- a more accurate representation of the genetic structure between predefined clusters, such as distinct genetic subpopulations or subpopulations that exhibit genetic admixture
- advantages: high accuracy in detecting differences between populations, simple and easily interpretable visualization
- disadvantages: requires predefined groups, may be sensitive to low levels of diversity or high levels of genetic connectedness

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

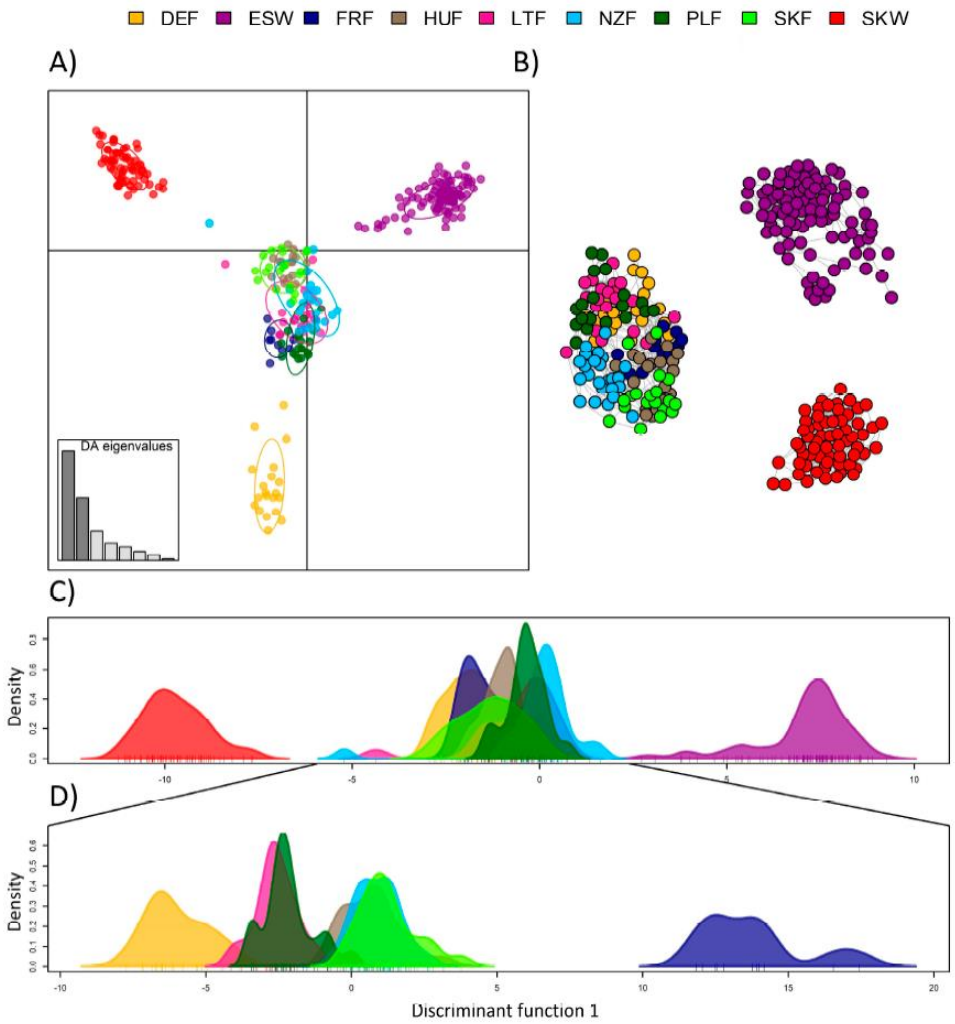


Fig. 5: Genetic relationships between 7 farmed and 2 wild populations of red deer based on the first and second discriminant functions of the DAPC analysis (A), genetic distances (B) and the first discriminant function separately for the whole dataset (C) and farm populations (D) (Moravčíková et al., 2023)

Discriminant analysis of principal components (DAPC)

- Software tools: R (adegenet package)

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

Approaches and tools for population structure analysis

Admixture analysis – Bayesian approach

- Population structure detection and identification of genetic clusters without the need to predefine groups
- It allows the identification of genetic groups and the degree of admixture within individuals, which is useful in the study of migration and differentiation
- advantages: accurate identification of genetic clusters, flexible method for complex structures
- disadvantages: time-consuming calculation, reliability of results depends on the number of SNP markers used and tested individuals

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

Approaches and tools for population structure analysis

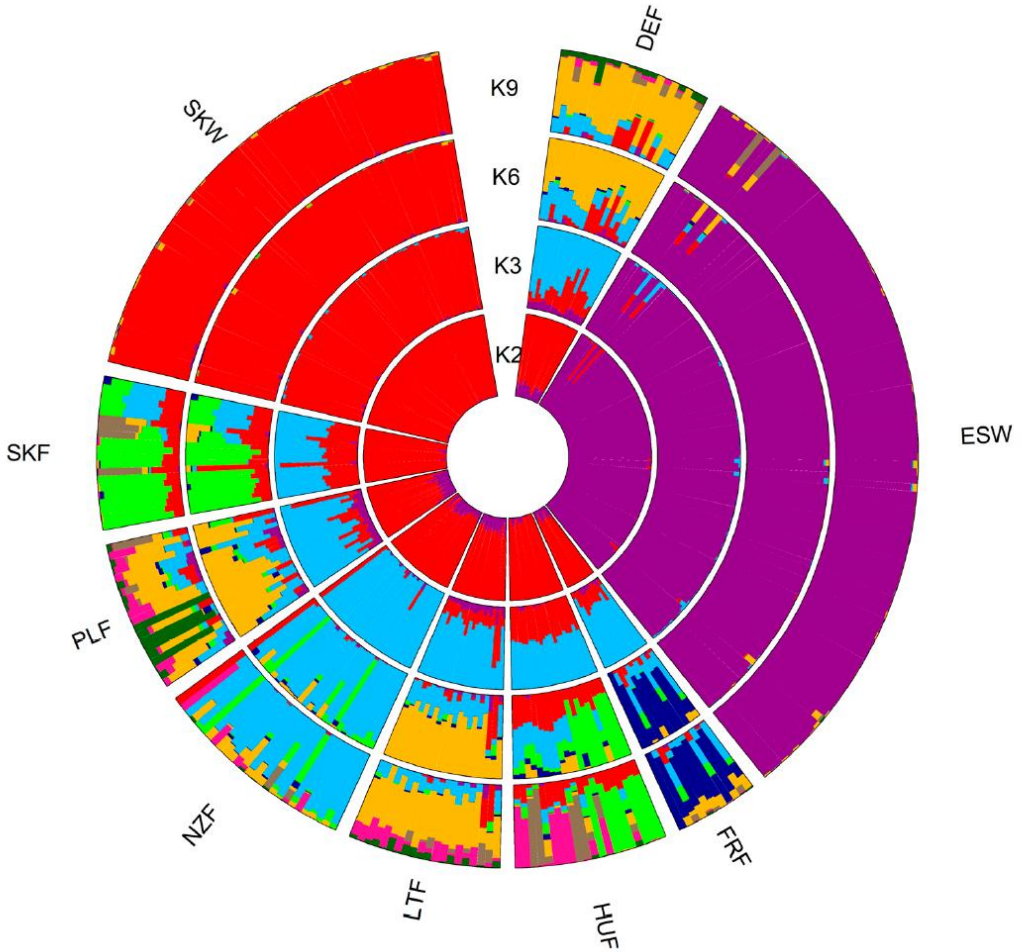


Fig. 6: Representative results of genetic admixture testing between farmed and wild populations of red deer using a Bayesian approach for K=2, K=3, K=6 and K=9 (Moravčíková et al., 2023)

Admixture analysis – Bayesian approach

- Software tools: e.g. STRUCTURE, ADMIXTURE, BAPS, FASTSTRUCTURE

Co-funded by the European Union

Approaches and tools for population structure analysis

Gene flow between populations – TreeMIX program

- based on allele frequencies, it creates phylogenetic trees with the possibility of testing the intensity of migration between populations
- this method is based on maximum probability and allows simultaneous estimation of phylogenetic relationships and migration between populations
- advantages: allows detection of the intensity of migration and gene flow in the past
- disadvantages: the reliability of the results depends on the amount of available genotypic data as well as the reliability of the allele frequency estimation

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

Gene flow between populations – Bayesass program

- uses a Bayesian approach
- allows to determine the intensity of gene flow between and within populations

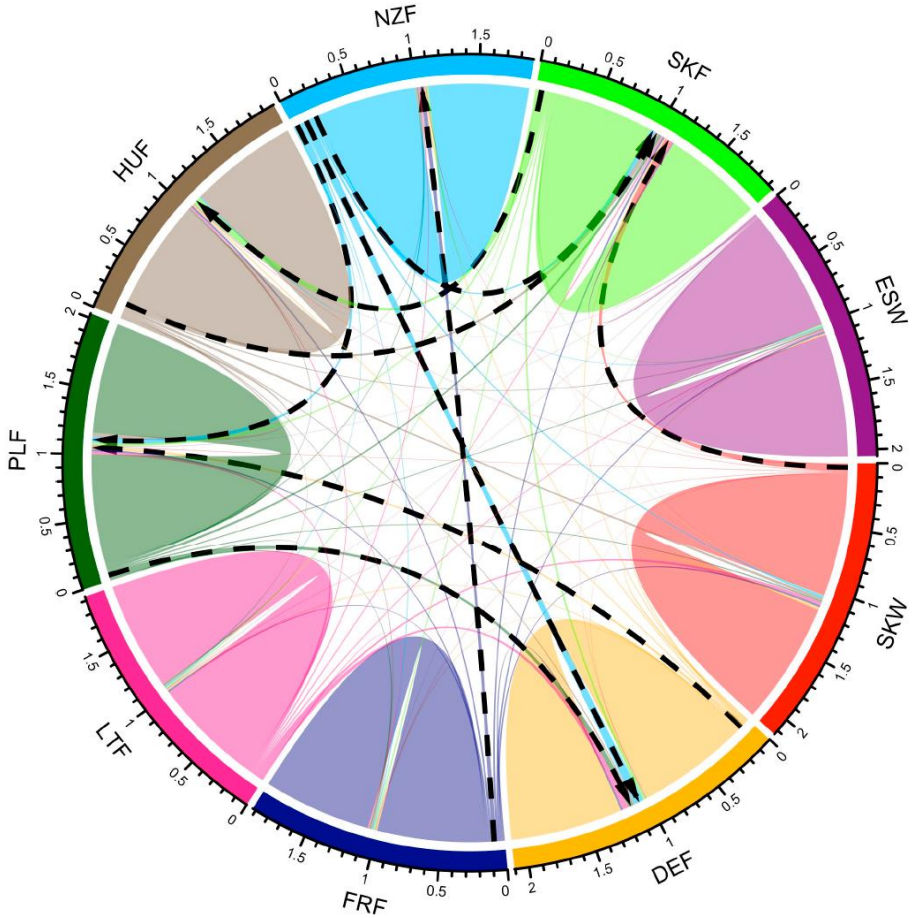


Fig. 7: Graphical visualization of gene flow intensity between farmed and wild populations of red deer (Moravčíková et al., 2023)

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

Constructing genetic networks – Netview package

- is a visualization tool that uses genetic networks to show relationships between individuals or populations
- creates genetic networks that show genetic relationships and gene flow between populations
- is suitable for assessing complex relationships as well as the impact of migration
- advantages: intuitive visualization, suitable for displaying admixture and differentiation
- disadvantages: limited use in populations with large numbers of individuals

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

Constructing genetic networks – Netview package

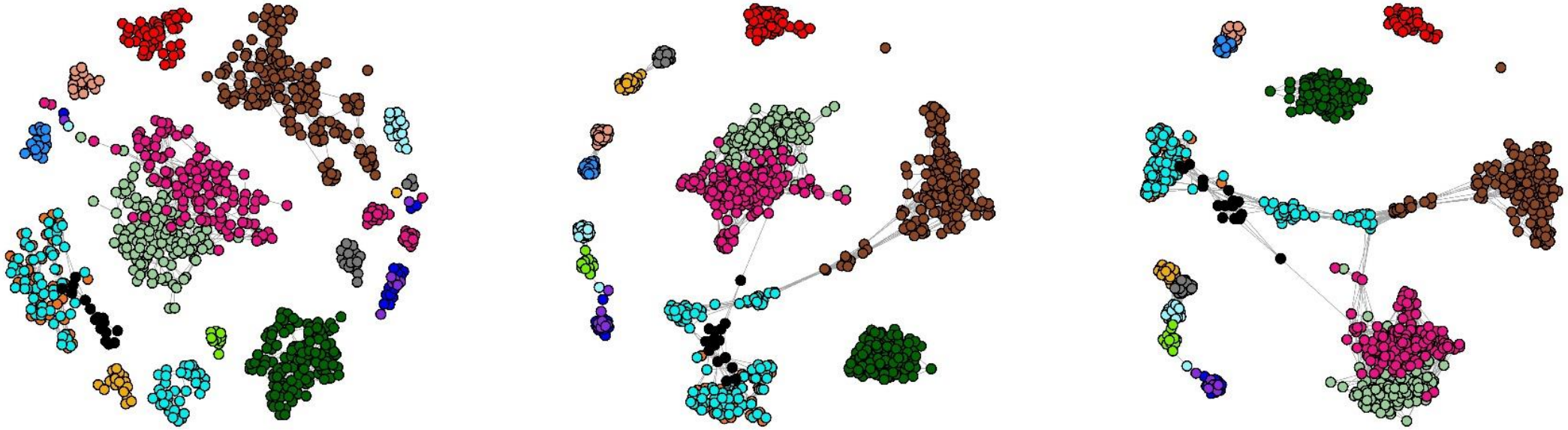


Fig. 8: Graphical visualization of the results of testing 3 different scenarios of development of intra-population and inter-population genetic relationships within 16 cattle breeds using Netview (unpublished results)

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

Construction of phylogenetic trees

- graphical representations of evolutionary relationships between populations or species based on their genetic data
- they are used to visualize genealogical or genetic relationships, model evolutionary processes, and track population differentiation and migration
- they can be created using a variety of algorithms and models, most commonly based on genetic distances (e.g., Nei's genetic distance) or probabilistic models (maximum reliability and Bayesian methods)

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for population structure analysis

Construction of phylogenetic trees

- Software tools: e.g. MEGA, SplitsTree, various R packages

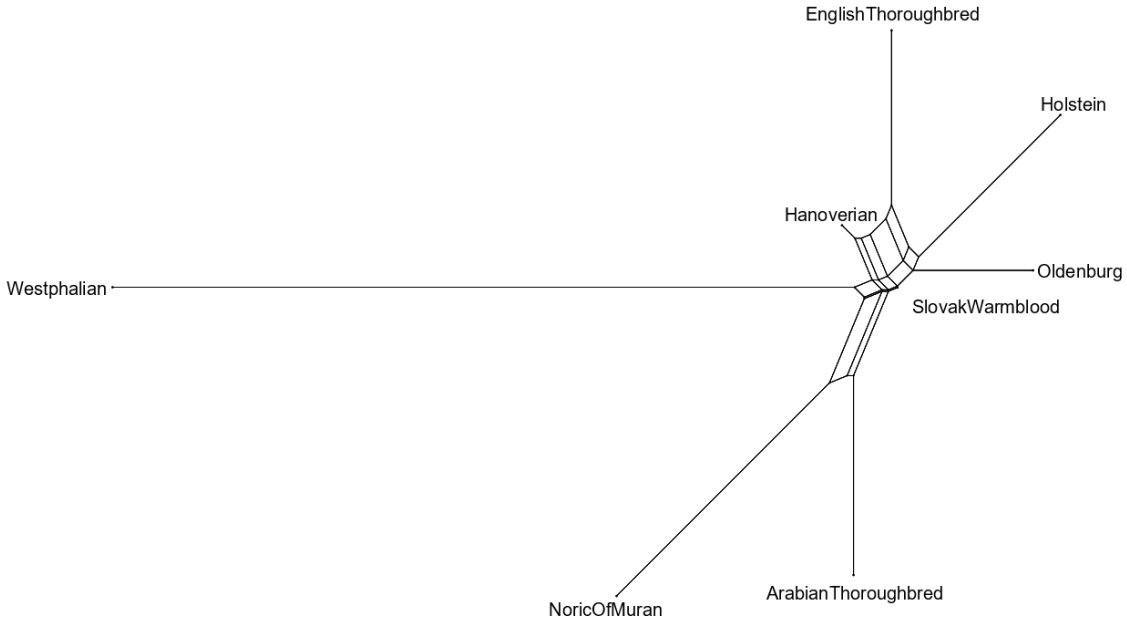


Fig. 9: Phylogenetic tree constructed based on a matrix of Nei's genetic distances reflecting relationships between 8 horse breeds (unpublished results)

Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

- genomic regions under significant selection pressure – so-called selection signals
- analysis of selection signals allows for a better understanding of:
  - evolutionary processes and the impact of domestication
  - the impact of natural and intensive artificial selection on specific genomic regions controlling preferred phenotypic traits, whether in terms of adaptability, resilience or performance of individuals, populations and livestock species
- identify genomic regions showing a decrease or increase in genetic variability
- phenotypic information is not necessary

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

- Two groups of approaches:
  - Evaluation of inter-population/inter-breed differences: e.g.
    - Wright's F_{ST} index at the genome-wide level
    - Differences in linkage disequilibrium (LD) – analysis based on haplotype structure
    - PCA analysis
  - Evaluation at the intra-population level: e.g.
    - Distribution of runs of homozygosity
    - Distribution of heterozygosity-rich regions
    - Level of linkage disequilibrium
    - RDA analysis
    - Tajima's D statistics

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation of inter-population/inter-breed differences - Wright's F_{ST} index

- one of the most commonly used approaches
- selection signals are identified based on differences in allelic frequencies between populations, which arose as a result of, for example, different breeding goals or breed standards
- two basic types of signals - the different type of selection corresponds to the regions (represented by several loci or SNP markers) with a high value of the F_{ST} index and, conversely, the regions with a low value represent genomic regions that were subject to the same type of selection in the given breeds

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation of inter-population/inter-breed differences - Wright's F_{ST} index

- threshold value defining the signal – e.g. 1% of the highest values
- advantages: relatively simple method of calculation and wide use in population genetics
- disadvantages: cannot be used for the identification of signals at the intrapopulation level

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation of inter-population/inter-breed differences - Wright's F_{ST} index

- Software tools: e.g. PLINK

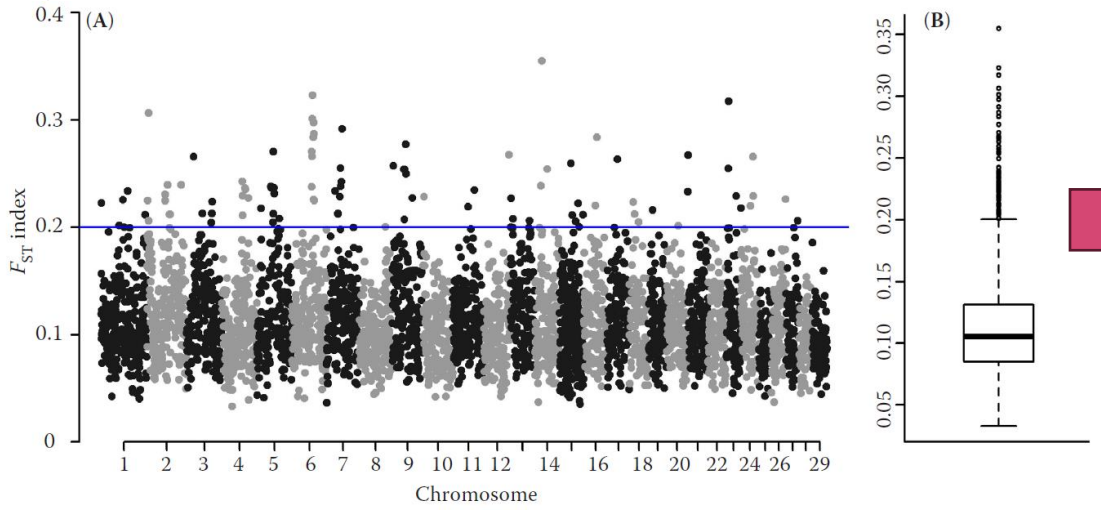


Fig. 10: Distribution of F_{ST} index values in the autosomal genome of beef cattle breeds (A) and overall using a box plot (B) (Moravčíková et al., 2019)

Tab. 1: Description of selected selection signals (Moravčíková et al., 2019)

BTA	Start (Mb)	End (Mb)	Length (Mb)	Genes *n*	QTL in region
1	62.046	75.434	13.388	164	live weight at age of 365 days, carcass weight, BSE resistance
	80.063	91.358	11.295	132	
2	5.277	5.664	0.387	6	
	61.736	62.266	0.530	9	marbling score, milk production, carcass weight
	70.560	77.289	6.729	45	marbling score, length of production life, birth weight, IMF
3	84.966	88.207	3.241	28	marbling score
4	70.452	71.554	1.101	10	LMA (*Longissimus thoracis* cut)
	80.303	81.434	1.131	2	
5	48.460	49.399	0.939	11	dressing percentage, birth weight, LMA, FSH concentration, EBV for backfat thickness
	55.652	56.618	0.966	38	
	57.470	60.557	3.086	154	
6	67.644	69.185	1.541	21	growth, strength, body frame, ham thickness, claw angle, claw quality, tits placement, udder quality, udder depth, fat and protein content
	70.001	72.907	2.906	35	
	74.085	76.210	2.125	6	
7	38.904	39.327	0.423	8	ovulation intensity, SCS
	45.439	48.909	3.469	70	
	51.021	53.468	2.447	67	
9	40.401	43.145	2.744	35	fat content
	46.531	69.144	22.613	148	milk yield, fat and protein yield, ham angle, conformation traits 2 and 6

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation of inter-population/inter-breed differences – Integrated haplotype score (iHS)

- selection signals are derived from a change in the linkage disequilibrium in the genome of the evaluated breeds and the emergence of specific haplotypes due to the linkage disequilibrium
- The iHS value can be defined simply as a measure of how unusual a haplotype consisting of specific SNP markers is, compared to the rest of the genome

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation of inter-population/inter-breed differences – Integrated haplotype score (iHS)

- iHS is a particularly sensitive method for detecting the effect of recent selection that leads to an increase in the frequency of a certain allelic variant in a population, but has not yet had eliminate other variants at a given locus
- the analysis begins with the calculation of the EHH value (extended haplotype homozygosity), which quantifies the decrease in homozygosity of the haplotype from a certain SNP marker and then continues with the calculation of the iHS value, which is based on the logarithm of the ratio of integrated EHH values for two allelic variants

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation of inter-population/inter-breed differences – Integrated haplotype score (iHS)

- iHS can reach positive values (a haplotype carrying a single allele is longer and has a higher EHH, indicating a significant effect of positive selection) or negative values (an alternative allele has a higher EHH, which can also reflect selection, but in the opposite direction)
- threshold value defining the signal – e.g. 1% of the highest positive values
- advantages: suitable approach for detecting the effect of recent selection and identification of signals arising e.g. as a result of adaptation, possibility of haplotype structure analysis
- disadvantages: the need for high-quality and robust genomic data, the need to define or determine haplotype frequencies

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation of inter-population/inter-breed differences – Integrated haplotype score (iHS)

- Software tools: e.g. REHH program package, Haploview

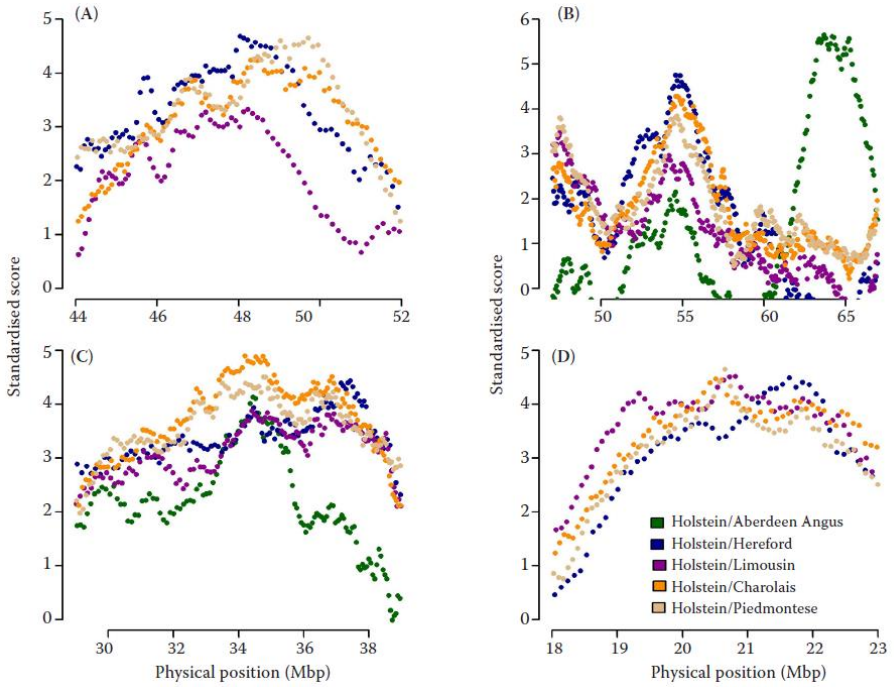


Fig. 11: Differences in linkage disequilibrium on chromosomes 7 (A), 13 (B), 20 (C) and 26 (D) between the Holstein breed and beef cattle breeds (Moravčíková et al., 2019)

Tab. 2: Autosomal regions under selection pressure identified based on variability in linkage disequilibrium (Moravčíková et al., 2019)

Breed	BTA	Start (Mb)	End (Mb)	Region length (Mb)	Genes *n*	QTL in region
Aberdeen Angus	13	62.825	65.859	3.034	82	dressing percentage
Hereford	7	47.201	49.331	2.130	39	
	13	54.208	55.559	1.350	70	dressing percentage
Limousin	26	19.088	22.122	3.034	47	dressing percentage, IMF, milk yield
Charolaise	20	32.621	37.325	4.704	39	
Piedmontese	7	48.308	50.202	1.894	53	
	20	33.418	35.197	1.778	7	milk production, fat and protein content
	26	20.572	21.075	0.503	11	dressing percentage, IMF, milk yield
Romagnola	6	38.791	41.800	3.009	25	birth weight, growth, strength, milk yield, fat and protein content

Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation of inter-population/inter-breed differences – PCA analysis

- assumes that the selection signals in the genome arose as a result of the local adaptation of individuals to environmental conditions
- an alternative method for identifying selection signals to the F_{ST} index
- detection of selection signals is based on the assumption of the existence of a correlation between genetic variants and principal components, which reflects the local adaptation of populations to the production environment

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation of inter-population/inter-breed differences – PCA analysis

- to identify selection signals, e.g. Mahalanobis distance test evaluating the distance of points from the mean could be used - the identification of SNP markers showing association with positive selection is then based on the construction of a z-score vector, obtained by regression analysis of the relationship between SNP markers and the principal components of K
- the threshold value defining the signal can be determined, for example, based on the false discovery rate FDR test
- advantages: efficient visualization of complex genetic data
- disadvantages: alternative, not so often used approach, complicated interpretation

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation of inter-population/inter-breed differences – PCA analysis

- Software tools: PCAdapt R Package

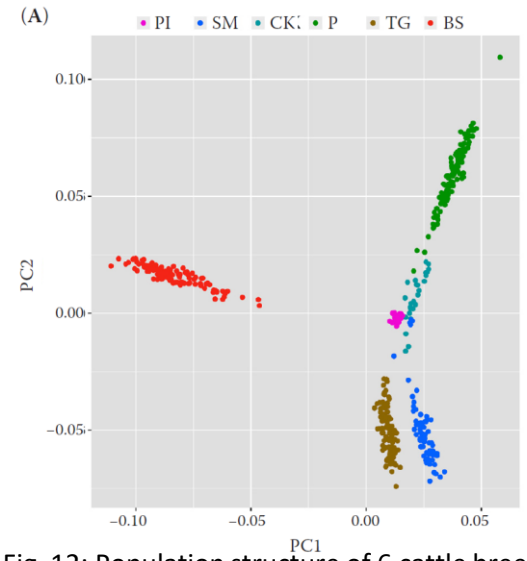


Fig. 12: Population structure of 6 cattle breeds based on PCA analysis (Moravčíková et al., 2018)

Tab. 3: Genomic regions showing the strongest selection signal (Moravčíková et al., 2018)

Region	BTA	Start position (Mb)	End position (Mb)	No. of SNPs	QTL traits	No. of genes
1	1	94.59	133.94	11	Resistance to BSE; Birth weight; Adjusted weaning and yearling weight	244
2	2	4.06	7.49	15	Yearling weight; Kidney, pelvic and heart fat	41
3	3	43.69	64.87	13	Marbling score; Estimated kidney, pelvic and heart fat	159
4	6	37.87	62.69	21	Longissimus muscle area; Hot carcass weight; Birth weight; Yearling weight; Marbling score	95
5	9	40.12	56.41	10	Marbling score; Canonical conformation trait 2	110
6	11	57.13	72.84	14	Yield grade	92
7	13	41.95	56.69	9	Canonical conformation trait 9	226
8	22	20.18	28.71	8		23

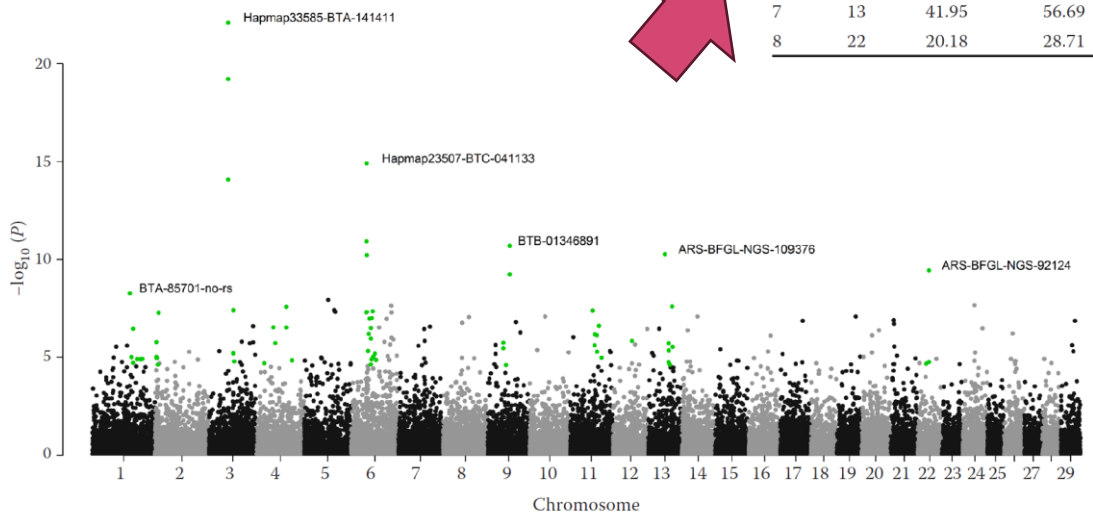


Fig. 13: Selection signals identified through PCA analysis (Moravčíková et al., 2018)

Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation at the intra-population level – Distribution of runs of homozygosity (ROH)

- this approach assumes that regions in the genome showing strong selection signals are the result of an increase in local homozygosity due to intensive breeding to traits defined in the breed standard of each breed
- ROH regions forming selection signals located in the genome are formed by alleles derived from common ancestors, which can be inherited from generation to generation in an unchanging form

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation at the intra-population level – Distribution of runs of homozygosity (ROH)

- selection signals are determined based on the frequency of SNP in ROH in a specific regions across individuals in the population
- threshold value defining the signal – e.g. 1% of the highest values
- advantages: this method allows to detect regions where there has been a decrease in diversity, a good indicator of the effect of positive selection
- disadvantages: the need for high-quality and robust genomic data

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation at the intra-population level – Distribution of runs of homozygosity (ROH)

- Software tools: e.g. PLINK, R package detectRUNS

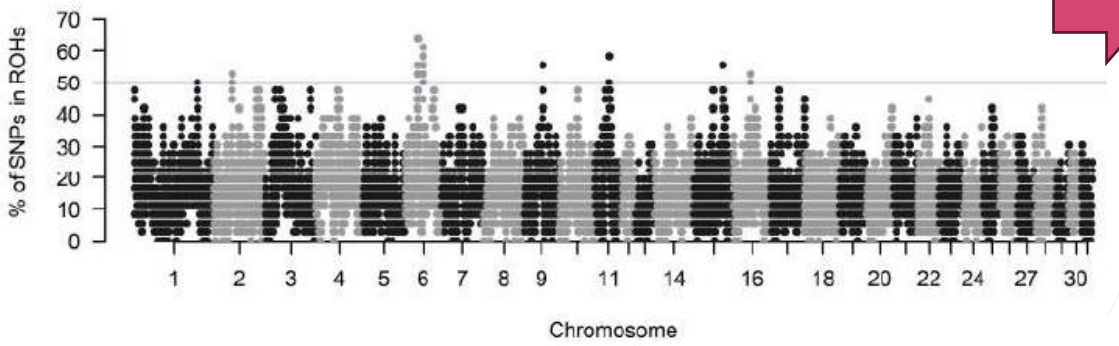


Fig. 14: Overlapping ROH regions in the autosomal genome of the Slovak warmblood horse (Moravčíková et al., 2020)

Tab. 4: Selection signals in the genome of the Slovak warmblood horse derived from the extreme values of the frequency of SNP in ROH regions (Moravčíková et al., 2020)

ECA	Start position (Mb)	End position (Mb)	Region size (Mb)	Protein-coding genes
1	147.61	148.06	0.44	STARD9, TTBK2, CDAN1, HAUS2, LRRC57, SNAP23, ZNF106, CAPN3, GANC, VPS39
2	42.74	43.12	0.38	RERE, SLC45A1, ERRFI1, PARK7, TNFRSF9
6	29.36	30.02	0.66	CACNA2D4, DCP1B, LRTM2, ADIPOR2
	41.18	42.71	1.54	LRP6, MANSC1, BORCS5, DUSP16, CREBL2, GRP19, CDKN1B, APOLD1, DDX47, GRPC5A, GRPC5D, HEBP1, FAM234B, GSG1, EMP1, GRIN2B
9	44.25	44.53	0.27	UQCRB, MTERF3, PTDSS1, SDC2
11	32.32	33.58	1.26	MSI2, CCDC182, MRPS23, CUEDC1, VEZF1, SRSF1, DYNLL2, EPX, MKS1, LPO, MPO, TSPOAP1, MIR142, RNF43, SUPT4H1, HSF5, MTMR4, TEX14, RAD51C PPM1E, TRIM37, SKA2
15	67.26	67.83	0.57	LBH, YPELP
16	39.90	40.62	0.71	SLC26A6, TMEM89, UQCRC1, MIR711, PFKFB4, SHISA5, TREX1, ATRIP, CCDC51, PLXNB1, FBXW12, SPINK8, NME6, ECATH-3, ECATH-2, CDC25A, MAP4

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation at the intra-population level – Distribution of heterozygosity-rich regions (HRR)

- this method is used to detect regions showing a high degree of heterozygosity, which may be important in terms of adaptability, response to environmental changes or the occurrence of pathogens – heterozygous individuals have higher fitness than homozygous individuals
- a high level of heterozygosity may be the result of balancing selection – the preservation of genetic diversity within a population
- similar approach to ROH – but selection signals form different lengths of continuous stretches of heterozygous genotypes in the genome

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation at the intra-population level – Distribution of heterozygosity-rich regions (HRR)

- selection signals are determined based on the frequency of SNP markers in HRR in a specific regions across individuals in the population
- threshold value defining the signal – e.g. 1% of the highest values
- advantages: allows to detect regions in which there is an increased proportion of heterozygous genotypes – an indicator of genomic regions important in terms of adaptation or evolutionary potential
- disadvantages: the need for high-quality and robust genomic data

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation at the intra-population level – Distribution of heterozygosity-rich regions (HRR)

- Software tools: R package detectRUNS

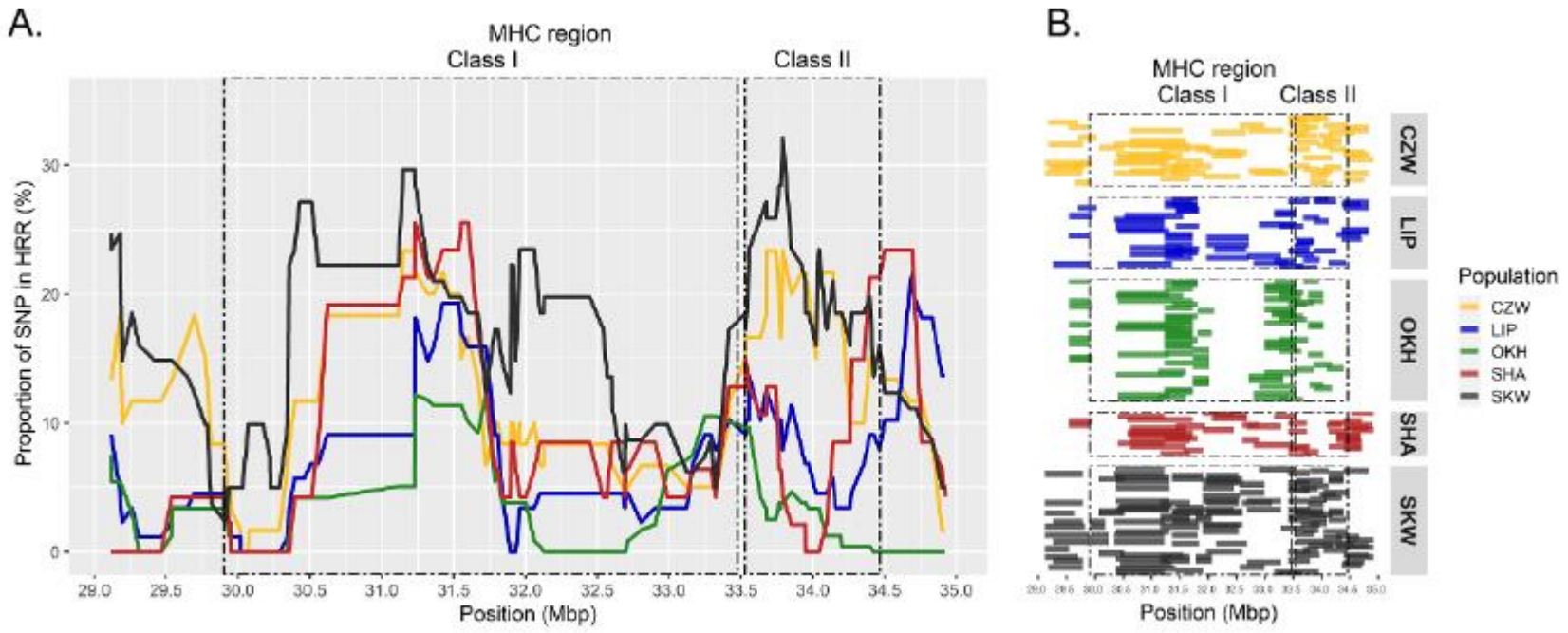


Fig. 15: Graphical visualization of the frequency of SNP markers in HRR regions (A) and distribution of HRR regions in the genomic coordinates of the major histocompatibility complex (B) (Moravčíková et al., 2024)

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation at the intra-population level – RDA analysis

- RDA (Redundancy analysis) tests the relationship between genetic variability and environmental factors – the influence of natural selection
- a method of evaluating genotype-environment associations (GEA) that simultaneously evaluates the percentage of genomic variability explained by environmental variables and detects loci under selection pressure
- two-step analysis in which genetic and environmental data are evaluated using multivariate linear regression
- advantages: a comprehensive approach to evaluate the relationships between genetic variability within a population and environmental factors
- disadvantages: the need for high-quality and robust data
- Software tools: e.g. R package vegan, DeepGenomeScan

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation at the intra-population level – Tajima's D statistics

- evaluates population diversity and can serve as one of the indicators of balancing selection
- positive values indicate balancing selection - allele frequencies usually showed intermediate frequency, resulting in more variability than would be expected based on the theory of neutral evolution
- negative values, on the other hand, are associated with the effect of positive selection, which leads to a decrease in genetic diversity and nucleotide variability
- threshold value defining the signal – e.g. 1% of the highest positive values
- advantages: this method allows to detect regions in which there is an increased proportion of heterozygous genotypes – an indicator of regions important e.g. in terms of adaptation
- disadvantages: the need for high-quality and robust genomic data

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Approaches and tools for evaluating the impact of selection on the livestock genome

Evaluation at the intra-population level – Tajima's D statistics

- Software tools: e.g. VCFtools, R package SnpR

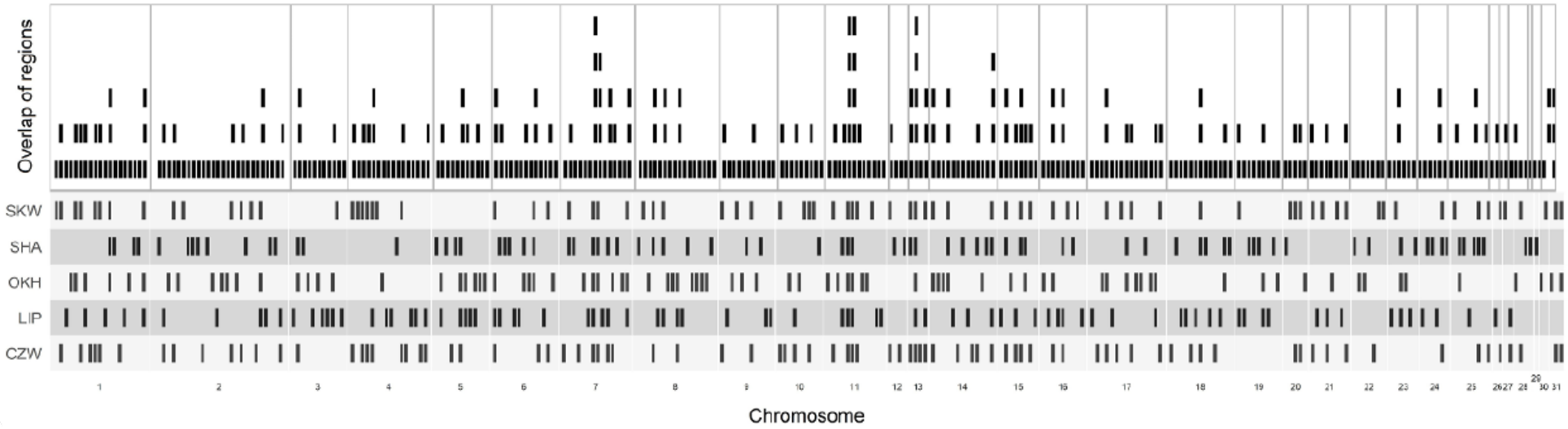


Fig. 16: Signals of balancing selection in the genome of 5 horse breeds identified based on the Tajima's D index (Moravčíková et al., 2024)

ISAGREED
Co-funded by the European Union

Functional annotation of regions significantly affected by selection pressure

- searching for quantitative trait loci (QTLs) in the region of selection signals
- search for genes located directly or near selection signals – potential candidate genes for further analysis
- functional annotation: identification of biological functions of genes using tools such as GO (Gene Ontology) or KEGG (Kyoto Encyclopedia of Genes and Genomes)

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

Functional annotation of regions significantly affected by selection pressure

Databases for QTL and gene identification

- Animal QTLdb - database of information on QTLs associated with significant phenotypic traits of different livestock species
- BioMart – Ensembl – a simple web-based tool for obtaining genome data (e.g. gene positions) without the need for programming knowledge

Functional analysis of genes

- DAVID: a web-based tool for functional annotation and analysis of the role of genes, including classification by biological processes and pathways

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Functional annotation of regions significantly affected by selection pressure

- advantages: detailed analysis of regions in the genome significantly affected by selection allows the identification of specific genes and biological pathways responsible for phenotypic traits, can help in further research of genes and QTL loci with potential use in breeding programs
- disadvantages: the overlap between selection signal regions and functional regions does not always imply a causal relationship, the information in the available databases is limited to current knowledge and may not always cover all relevant genes or QTL loci

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

References

- MORAVČÍKOVÁ, Nina - SIMČIČ, Mojca - MESZÁROŠ, G. - SÖLKNER, Johann - KUKUČKOVÁ, Veronika - VLČEK, Michal - TRAKOVICKA, Anna - KASARDA, Radovan. Genomic response to natural selection within alpine cattle breeds. In *Czech journal of animal science*. ISSN 1212-1819, 2018, vol. 63, iss. 4, s. 136-143. Dostupné na internete: http://dx.doi.org/10.17221/62/2017-CJAS.
- MORAVČÍKOVÁ, Nina - KASARDA, Radovan - VOSTRÝ, Luboš - KRUPOVÁ, Zuzana - KRUPA, Emil - LEHOCKÁ, Kristína - OLŠANSKÁ, Barbora - TRAKOVICKÁ, Anna - NÁDASKÝ, Rudolf - ŽIDEK, Radoslav - BELEJ, Ľubomír - GOLIAN, Jozef. Analysis of selection signatures in the beef cattle genome. In *Czech journal of animal science*. ISSN 1212-1819, 2019, vol 64, no. 12, s. 491-503. Dostupné na internete: https://doi.org/10.17221/226/2019-CJAS.
- MORAVČÍKOVÁ, Nina - KASARDA, Radovan - HALO, Marko - KADLEČÍK, Ondrej - TRAKOVICKÁ, Anna - LEHOCKÁ, Kristína - OLŠANSKÁ, Barbora - CANDRÁK, Juraj. Genome-wide distribution of autozygosity islands in slovak warmblood horse. In *AGROFOR*. ISSN 2490-3434, 2020, vol. 5, no. 1, s. 77-84.
- MORAVČÍKOVÁ, Nina - KASARDA, Radovan - ŽIDEK, Radoslav - VOSTRÝ, Luboš - VOSTRÁ VYDROVÁ, Hana - VAŠEK, Jakub - ČILOVÁ, Daniela. Czechoslovakian wolfdog genomic divergence from its ancestors canis lupus, german shepherd dog, and different sheepdogs of european origin. In *Genes*. ISSN 2073-4425 online, 2021, vol. 12, iss.6, article 832, p. 1-18. Dostupné na internete: https://doi.org/10.3390/genes12060832.
- MORAVČÍKOVÁ, Nina - KASARDA, Radovan - ŽIDEK, Radoslav - MCEWAN, John Colin - BRAUNING, Rudiger - LANDETE-CASTILLEJOS, Tomas - CHONCO, Louis - CIBEREJ, Juraj - POKORADI, Jaroslav. Traces of Human-Mediated Selection in the Gene Pool of Red Deer Populations. In *Animals*. ISSN 2076-2615, 2023, vol. 13, iss. 15, art. no. 2525 [17] s. Dostupné na internete: https://doi.org/10.3390/ani13152525.
- MORAVČÍKOVÁ, Nina - KASARDA, Radovan - VOSTRÁ VYDROVÁ, Hana - VOSTRÝ, Luboš - KARÁSKOVÁ, Barbora - CANDRÁK, Juraj - HALO, Marko. Genomic variability of the MHC region: Empirical evidence from five horse breeds. In *Livestock science*. ISSN 1871-1413, 2024, vol. 284, art. no. 105480, [10] s. Dostupné na internete: https://doi.org/10.1016/j.livsci.2024.105480.

Co-funded by the European Union

ISAGREED

Partners:

Thank you for your attention!

This presentation has been supported by the Erasmus+ KA2 Cooperation Partnerships grant no. ***2021-1-SK01-KA220-HED-000032068 "Innovation of the structure and content of study programs in the field of animal genetic and food resources management with the use of digitalisation - Inovácia obsahu a štruktúry študijných programov v oblasti manažmentu živočíšnych genetických a potravinových zdrojov s využitím digitalizácie"****. The European Commission support for the production of this presentation does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.*

Nina Moravčíková

nina.moravcikova@uniag.sk

Co-funded by the European Union