



Analysis of the AnGR biodiversity status using genomic data

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

Nina Moravčíková

Slovak University of Agriculture in Nitra

Faculty of Agrobiolgy and Food Resources

Tools for the analysis of genome variability

Genome:

- haploid chromosome set/complete DNA sequence

DNA analysis:

- Genetic markers
- Whole-genome sequencing
- Genotyping arrays (chips)



Co-funded by
the European Union

Tools for the analysis of genome variability

Genetic marker

- any characteristic trait or an organism characteristic that can be used to identify a specific chromosome, cell or individual
- genes, short segments of DNA, chromosomal heterochromatin regions, or other manifestations of genotype, chromosomes or karyotype
- a polymorphic trait (variant) that exhibits mendelian inheritance and is correlated with genetic variation of a trait that is significant from a breeding point of view



Co-funded by
the European Union

Tools for the analysis of genome variability

Perspective markers of production traits of organisms

- candidate genes - their different alleles and genotypes influence quantitative traits
- QTL - quantitative trait loci



Co-funded by
the European Union

Tools for the analysis of genome variability

DNA markers

- direct ability to detect alleles in nucleotide sequences
- ↑ polymorphism
- dominant/codominant heritability
- common occurrence in the genome
- easy and fast testing
- high reproducibility



Co-funded by
the European Union

Tools for the analysis of genome variability

SNP – single nucleotide polymorphism

- is caused by a nucleotide substitution in DNA at a specific location (point mutation)
- occurrence in genome: 100-300 bp →
the most common type of DNA polymorphism
- occurrence in populations with frequency $> 1\%$
- biallelic marker
- provide information about gene variation in families, lineages and populations



Co-funded by
the European Union

Tools for the analysis of genome variability

Whole-genome sequencing

- DNA sequencing is the determination of the exact order of individual nucleotides in the DNA strand, i.e. the determination of the primary structure of DNA
 - classical sequencing methods: the Maxam-Gilbert and Sanger method of sequencing
 - new sequencing technologies - Next Generation Sequencing (NGS): These are several platforms (Illumina, Ion Torrent, 454 and others) that differ in their technical approach to sequencing but produce comparable outputs



Co-funded by
the European Union

Tools for the analysis of genome variability

SNP arrays (chips)

- allow simultaneous analysis of several thousand SNP markers uniformly distributed in the genome
- available for most species of farm and companion animals (cattle, horses, sheep, goats, pigs, poultry, dogs or cats)
- currently the world's most popular tool for testing e.g.:
 - parentage analysis
 - genomic diversity status
 - predisposition of individuals and populations
 - genome-wide association studies
 - estimation of genomic breeding values
 -



Co-funded by
the European Union

Genome homozygosity and genomic inbreeding

- **Genome autozygosity**– all alleles or chromosomal segments of DNA that are identical by descent (IBD) → derived from common ancestor
- **runs of homozygosity (ROH)** - genomic regions with a specific number of consecutive homozygous SNP markers
- effect of natural and artificial selection, recombination, linkage disequilibrium, mutations....
- the length of ROH segments in the genome corresponds to the distance of ancestors in the pedigree of an individual



Co-funded by
the European Union

Genome homozygosity and genomic inbreeding

Utilization of ROH segments screening:

estimation of genomic inbreeding based on the proportion of homozygous segments in the autosomal genome

testing the impact of the section on the genome

identification of causal mutations involved in the control of preferred traits and characteristics

Genome homozygosity and genomic inbreeding

Estimation of inbreeding coefficient (McQuillan et al., 2008):

$$F_{ROH} = \frac{\sum L_{ROH}}{L_{autosome}}$$

where $\sum L_{ROH}$ expresses the total length of ROH segments in the genome of an individual containing a specific number of consecutive homozygous SNP markers, and $L_{autosome}$ is the length of the autosomal genome derived based on the physical position of the SNP markers tested

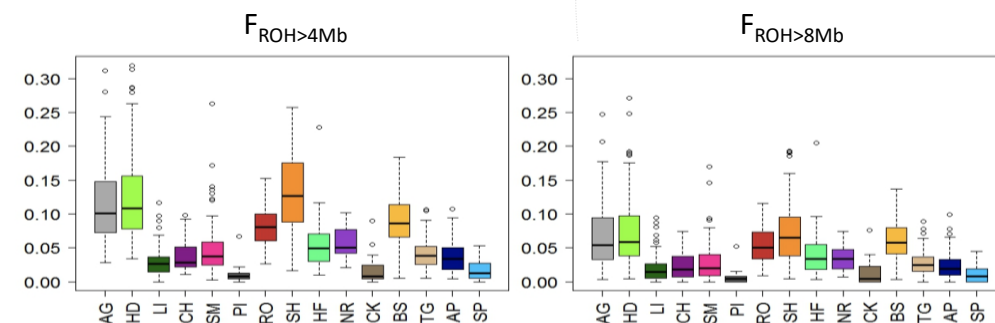
- F_{ROH} : from 0 to 1 (from 0 to 100%)
- $\Delta F > 1\%$ in small populations (4% in large populations) can lead to a significant loss of diversity and may affect their long-term survival

The most commonly used programs:

PLINK v1.9

detectRUNS (program R)

cgaTOH



Genomic inbreeding derived from homozygous segments based on the number of generations from a common ancestor ($F_{ROH>4Mb}$ – 12-13 generácií a $F_{ROH>8Mb}$ – 6-7 generácií) in 15 European cattle breeds (Kukučková et al., 2017)

McQUILLAN, R. et al. 2008. Runs of homozygosity in European populations. In *Am J Hum Genet*, vol. 83, pp. 359-372.

KUKUČKOVÁ, Veronika - MORAVČÍKOVÁ, Nina - FERENČAKOVIĆ, Maja - SIMČIČ, Mojca - MÉSZÁROS, Gábor - SÖLKNER, Johann - TRAKOVICKÁ, Anna - KADLEČÍK, Ondrej - CURIK, Ino - KASARDA, Radovan. Genomic characterization of Pinzgau cattle: genetic conservation and breeding perspectives. In *Conservation Genetics*. ISSN 1566-0621, 2017, vol. 18, no. 4, s. 893-910.



Co-funded by
the European Union

Linkage disequilibrium and effective population size of AnGR

Linkage disequilibrium (LD)

- non-random association of alleles at different loci in a given population
- influence of various factors: selection, recombination, mutations, genetic drift, mating system
- use of LD: analysis of evolutionary history of populations, impact of selection on the genome, estimation of effective population size



Co-funded by
the European Union

Linkage disequilibrium and effective population size of AnGR

Linkage disequilibrium (LD)

- is most commonly expressed using the Pearson product-moment correlation coefficient (Hill and Robertson, 1968):

$$r_{LD}^2 = \frac{(p_{AB} - p_A p_B)^2}{p_A(1 - p_A)p_B(1 - p_B)}$$

where p_{AB} is the difference between the frequency of gametes carrying alleles A and B at the two loci, and p_A and p_B are the resulting frequencies of these alleles

- takes values from 0 (linkage equilibrium) to 1 (complete linkage disequilibrium)

HILL, W.G. – ROBERTSON, A. 1968. Linkage disequilibrium in finite populations. In *Theor Appl Genet*, vol. 38, pp. 226-231.



Co-funded by
the European Union

Linkage disequilibrium and effective population size of AnGR

Effective population size (N_e)

- the relationship between LD and N_e variability is most commonly used to estimate both historical and current effective population size $\rightarrow N_{eLD}$
- N_{eLD} of the real population X with observed LD for a given segment in the genome (length interval in Mb or kb) is defined as the size of a hypothetical ideal population in equilibrium that has the same level of LD for that segment as in the observed real closed population



Co-funded by
the European Union

Linkage disequilibrium and effective population size of AnGR

Effective population size (N_e)

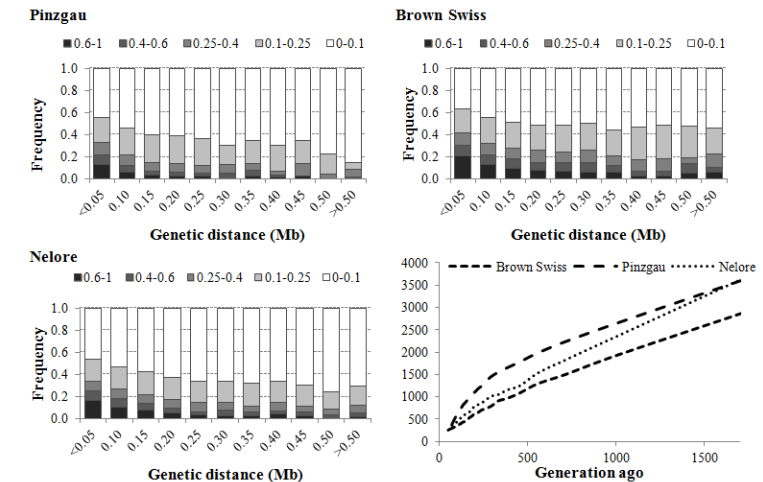
- can be calculated as follows (Corbin et al., 2012):

$$N_{eLD} = \frac{1}{kc} \left[\frac{1}{r_{LD}^2 - \frac{1}{n_g}} - \alpha \right]$$

where k expresses the heritability model ($k = 4$ for autosomal loci), c is the physical distance between SNP markers in Morgans, n_g is the gamete sample size (twice the number of individuals tested), and α is a factor accounting for the effect of mutations

historical N_{eLD} : function of time and physical distance between two markers assuming constant linear growth of N_{eLD} with time expressed by past generations

- can take values from 0 to n
- the most commonly used programs: LDNE, SneP, GONE



Comparison of the proportion of markers within the five categories of r_{LD}^2 and historical effective population size (Kasarda et al., 2016)

CORBIN, L.J. et al. 2012. Estimation of historical effective population size using linkage disequilibrium with marker data. In *J Anim Breed Genet*, vol. 129, pp. 257-270.

KASARDA, Radovan - MORAVČÍKOVÁ, Nina - MESZÁROŠ, G. - TRAKOVICKÁ, Anna - KADLEČÍK, Ondrej. Genetic divergence of cattle populations based on genomic information. In *Scientia agriculturae bohemia*. ISSN 1211-3174, 2016, vol. 47, no. 3, s. 113-117.



Co-funded by
the European Union

Genetic structure of populations

- diversity of populations at both intra- and inter-population levels
- genetic distance - the degree of genetic divergence (genomic divergence, e.g. in allele frequencies) between species or populations that can be quantified by various statistical approaches
- most commonly used parameters:
 - Nei's genetic distance and Wright's F_{ST} index
 - Principal component analysis
 - Bayesian analysis of the degree of genetic admixture/gene flow between populations



Co-funded by
the European Union

Genetic structure of populations

Nei's genetic distances

- assumes that if two populations displaying low genetic distances are similar, they share common ancestors with a high degree of reliability
- can be calculated (Nei, 1972):

$$D = -\ln I ; I = \frac{J_{xy}}{\sqrt{J_x J_y}}$$

where I is the normalized gene identity (or genetic identity) between population X and Y

- takes values from 0 (populations/individuals are genetically identical) to 1 (populations/individuals are genetically different)
- the R program can be used for the calculation (packages StAMPP, Poppr etc.)

NEI, M. 1972. Genetic distance between populations. In Am Nat, vol. 106, pp. 283-285.



Co-funded by
the European Union

Genetic structure of populations

Wright's F_{ST} index

- indicator of the intensity of population fragmentation expressed as a decrease in heterozygosity in subpopulations due to genetic drift
- can be calculated according to formula (Weir and Cockerham, 1984):

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

where H_T is the expected heterozygosity of the metapopulation and H_S is the average heterozygosity in the subpopulations

- takes values from 0 (populations are genetically identical) to 1 (populations are genetically different) $\rightarrow F_{ST} > 0.25$ populations can be considered genetically differentiated
- the R program can be used for the calculation (packages StAMPP, Poppr etc.)

WEIR, B.S. – COCKERHAM, C.C. 1984. Estimating F-statistics for the analysis of population structure. In *Evolution*, vol. 38, pp. 1358-1370.



Co-funded by
the European Union

Genetic structure of populations

Principal Component Analysis

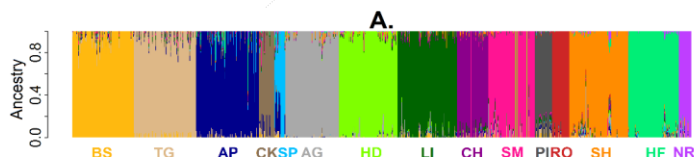
- a popular multivariate statistical method that has been applied in various scientific fields, including population genetics
- representation of high-dimensional data, e.g. genomic information about individuals or populations, in a reduced number of dimensions

Bayesian analysis of genetic admixture

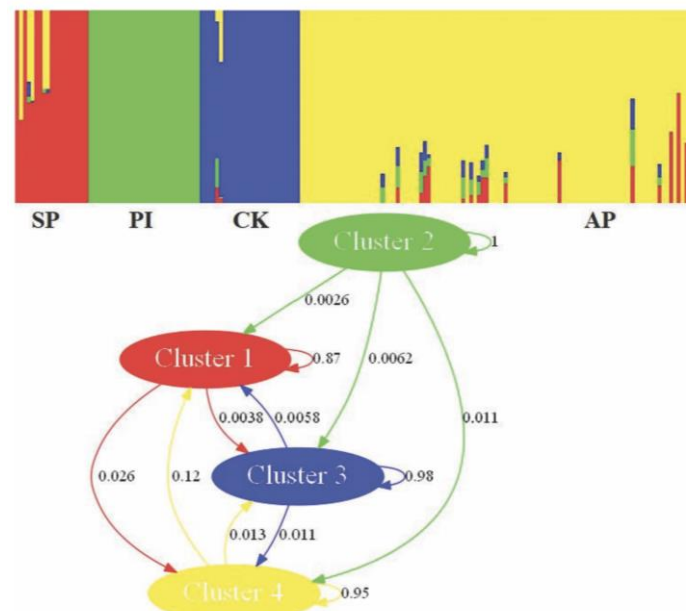
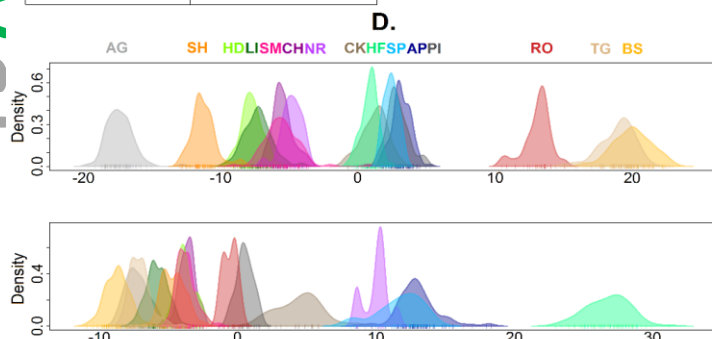
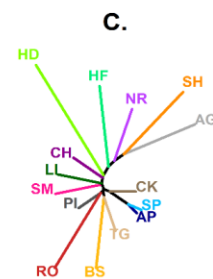
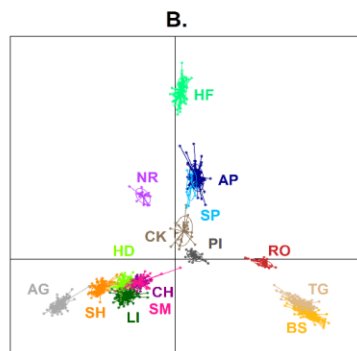
- admixture or mixing of populations is a phenomenon that can occur as a result of introgression and hybridization of individuals, populations or species
- Bayesian statistics - a field of modern statistics that works with conditional probability and allows to refine the probability of the initial hypothesis in the order in which other relevant facts appear



Genetic structure of populations



Graphical visualisation of population structure and genetic relationships between 15 European cattle breeds (Angus-AG, Austrian Pinzgau-AP, Brown Swiss-BS, Cika-CK, Hereford-HD, Holstein-HF, Charolais-CH, Limousin-LI, Norwegian Red-NR, Piedmontese-PI, Romagnola-RO, Shorthorn-SH, Slovak Pinzgau-SP, Simmental-SM, Tyrol Grey-TG) (Kukučková et al., 2017)



Analysis of admixture and gene flow between 4 breeds (SP: Slovak Pinzgau Cluster 1, PI: Piedmontese Cluster 2, CK: Cika Cluster 3, AP: Austrian Pinzgau Cluster 4) (Kukučková et al., 2018)

KUKUČKOVÁ, Veronika - MORAVČÍKOVÁ, Nina - FERENČAKOVIĆ, Maja - SIMČIČ, Mojca - MÉSZÁROS, Gábor - SÖLKNER, Johann - TRAKOVICKÁ, Anna - KADLEČÍK, Ondrej - CURIK, Ino - KASARDA, Radovan. Genomic characterization of Pinzgau cattle: genetic conservation and breeding perspectives. In *Conservation Genetics*. ISSN 1566-0621, 2017, vol. 18, no. 4, s. 893-910.

KUKUČKOVÁ, Veronika - MORAVČÍKOVÁ, Nina - CURIK, Ino - SIMČIČ, Mojca - MÉSZÁROS, Gábor - KASARDA, Radovan. Genetic diversity of local cattle. In *Acta Biochimica Polonica*. ISSN 0001-527X, 2018, vol. 65, iss. 3, s. 421-424.

Effect of selection on genome structure

- quantification without access to phenotypic information
identification of so-called selection signals
- depends on the intensity of selection, duration of its effect on the genome, recombination rate...
- two basic approaches:
 - inter-population/inter-breed differences
 - intra-population differences

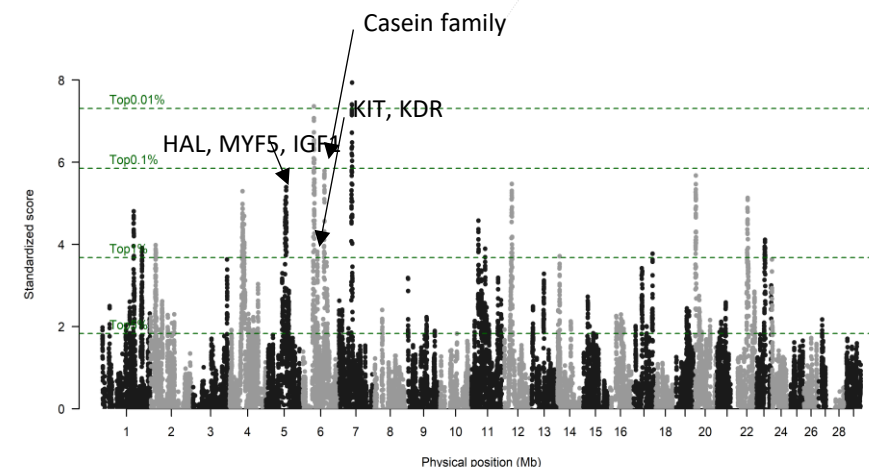


Co-funded by
the European Union

Effect of selection on genome structure

Inter-population/inter-breed differences

- genome-wide screening of the Wright's F_{ST} index
- analysis of variability in linkage disequilibrium
- principal component analysis
- integrated haplotype scores
- can be calculated by programs: PLINK v1.9, varLD, pcadapt, rehh...



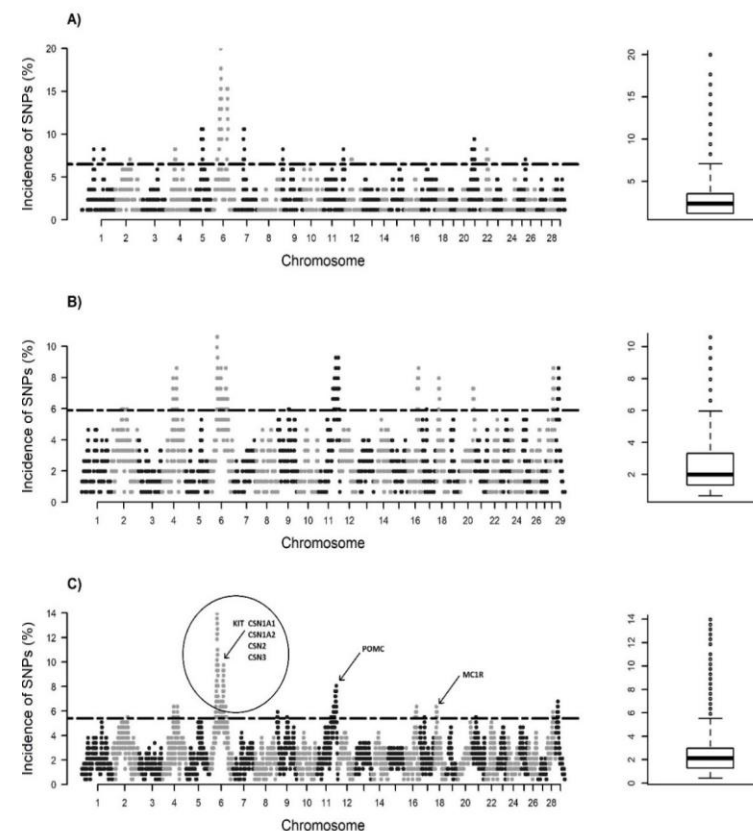
Differences in linkage disequilibrium in the genome of the Slovak Pinzgau and Slovak Spotted breeds (Moravčíková et al., 2019)

Effect of selection on genome structure

Intra-population differences

- distribution of runs of homozygosity in the genome
- level of linkage disequilibrium
- integrated haplotype scores
- can be calculated by programs: PLINK v1.9, detectRUNS, rehh...

KASARDA, Radovan - MORAVČÍKOVÁ, Nina - KADLEČÍK, Ondrej - TRAKOVICKÁ, Anna - CANDRÁK, Juraj. The impact of artificial selection on runs of homozygosity in Slovak Spotted and Pinzgau cattle. In *Slovak journal of animal science*. ISSN 1337-9984, 2018, vol. 51, no. 3, s. 91-103.



Frequency of occurrence of SNP markers in ROH segments (%) for Slovak Spotted (A) and Slovak Pinzgau breeds (B) and occurrence of SNP markers in ROH within the genome of both breeds (C) (Kasarda et al., 2018).



Siedlce University
of Natural Sciences
and Humanities



Czech University
of Life Sciences Prague

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