## 9. Genetic markers in breeding

Hello. Molecular genetics influences new breeding methods. This lecture focuses on genetic markers and their applications. The lecture is part of Module 3, Animal Breeding. The creation of this presentation was supported by ERASMUS+ KA2 grant within the ISAGREED project, Innovation of content and structure of study programs in the field of management of animal genetic and food resources using digitization.

With the development of molecular genetics since the 1970s, and especially molecular genetic methods working with DNA molecules, it is possible to identify real genetic variability (genotypes) using molecular genetic markers. This is used, for example, for mapping genes or regions in the genome (QTL) where genes for complex utility traits might be located. Thanks to these analyses, the use of this information in breeding is possible, specifically for improving estimates of genetic parameters (such as heritability) and breeding values, which leads to the efficiency of breeding through increased genetic gain and shortened generation interval.

At the beginning (i.e. since the 1980s), the idea of using genetic markers as a selection criterion emerged, i.e. after identifying the genotype of individuals, subsequent selection of a genotype-appropriate parent. This is called marker-assisted selection (MAS). However, this approach is limited in its actual use. Such selection has been used, for example, for the stress gene in pigs (CRC gene) or BLAD in cattle. These were genes with monogenic inheritance, mutations of which caused a reduction in individual fitness. For utility, complex, quantitative traits, direct selection based on the genotype of a single gene is inappropriate, or multiple markers had to be included in statistical models and breeding values were estimated. If the variability in the causal gene was directly determined by a marker, it is called gene-assisted selection (GAS). A great revolution in these approaches was the advent of new technologies for massive genome sequencing and identification of millions of markers (SNPs) and more precise determination of QTL regions and genes in the genome for quantitative traits. Since 2009, after sequencing the genomes of cattle, and gradually other economically important animal species, genomic selection has been gradually introduced. This is a method that incorporates genomic SNP markers (tens of thousands to hundreds of thousands) into a genomic relationship matrix and into equations such as BLUP (various variants) in order to estimate breeding values - GEBV (genomic estimated breeding value). This is again just one number that is easily usable in breeding practice.

Because we use real genetic variability, breeding is significantly streamlined - costs are reduced, breeding value estimates are improved, and the generation interval is shortened.

So what is a genetic molecular marker? It is a detectable polymorphism (multiple alleles) with a known position in the genome. There are three types of these markers: Type I are coding genes, so-called candidate genes (e.g. CRC gene, ESR in pigs). Type II markers are microsatellites, i.e. short tandemly repeating base sequences (STR) - found outside coding sequences. Type III markers are biallelic single nucleotide polymorphisms (SNPs) in coding or more commonly in non-coding intronic or intergenic regions. An example of one SNP is shown in the bottom image.

The diagram shows polymorphism in SNPs for a single nucleotide pair, and polymorphism in microsatellites, which is caused by length variation of tandem repeats (especially dinucleotide

repeats). SNPs usually have two alleles, while microsatellites can have up to twenty alleles in a population, although there are significantly fewer of them in genomes compared to SNPs.

The main significance of a genetic marker is that it can be associated with phenotypic variability of an important production trait in breeding, even though it may not have a direct biological effect on the trait. Such a marker is then referred to as indirect. It is called indirect because it is in linkage (i.e. close proximity on the chromosome) with another gene (QTL region) that directly influences the trait. The marker's linkage disequilibrium with QTL is then utilized. The stronger the linkage, the more informative the marker is for breeding.

An example of a SNP that is also a candidate gene is a SNP on the 4th chromosome, located in the leptin gene, where there is a polymorphism of adenine and guanine substitution. Furthermore, we can see that the mutation is in the coding sequence and causes a change in reading, substituting the 40th amino acid from threonine to alanine. This SNP is therefore a direct marker. Subsequent association analysis must determine whether this SNP affects the variability in the effect of leptin and, in this case, the indicators of fat in cattle muscle.

The relationship between the marker and QTL is therefore related to their distance from each other and the possibility of crossing over between them. If the marker is not direct and causal, it may be in linkage with the causal locus QTL. In the figure, we can see different combinations of linkage disequilibrium between marker alleles (M1 and M2) and QTL alleles (L - worse-low and H - better-high performance). Depending on the combination on the chromosome, allele M1 may be advantageous in one case, according to which selection can be made, while allele M2 may be advantageous in another case. In the third case, no marker allele is suitable because both are linked to the same H allele in QTL.

In brief, we will describe the principle of marker-assisted selection (MAS). MAS in selection programs for livestock allows for increased accuracy in selecting specific DNA variations associated with measurable differences in economically important traits. The degree of genetic improvement achieved through MAS can be substantially higher than improvement achieved through selection based on breeding values for traits with low heritability values in populations or traits determined post-mortem. Therefore, MAS has the potential to significantly increase the effectiveness of animal breeding for these traits.

MAS phases. Detection phase, evaluation phase, and implementation phase are distinguished. In the detection phase, DNA polymorphisms are used as direct or linked markers to determine specific allele frequencies within QTL segregation populations. During this phase, markers associated with QTL are identified, and the size of allele effects and the location of QTL in the genome can be estimated.

In the evaluation phase, linked markers are tested in target populations to determine whether QTL segregate within the population.

The implementation phase uses predictive linked markers within families, and direct markers are used across families to create a genotype database. These data are combined with pedigree and phenotypic information during genetic evaluation to predict individual genetic values.

MAS is suitable for direct selection of individuals based on the genotype of a genetic marker/gene for simple traits - monogenic traits (most commonly monogenic diseases). For example, the CRC stress gene and the BLAD disease gene in cattle.

For markers associated with quantitative traits, the use of MAS is limited, with a smaller effect - there are not as many described candidate genes, and there are not many traits with simple genetic determinism. For these traits, it is necessary to include whole-genome SNP markers in the genomic selection system.

And thank you for your attention.