10. Genomic selection

Hello. The topic of this lecture is genomic selection. The lecture is part of Module 3, Animal Breeding. The creation of this presentation was supported by the ERASMUS+ KA2 grant within the project ISAGREED, Innovation of content and structure of study programs in the field of management of animal genetic and food resources using digitization.

To begin, let's define a genome. A genome is the complete set of genetic information of an organism. In eukaryotic organisms, the genome is contained in the haploid set of chromosomes. After 2000, the development of next-generation sequencing technologies allowed for faster and cheaper sequencing of whole genomes.

The first sequencing of the whole genomes of major livestock species took place around 2009. Cattle have a genome size of 2.7 Gbp (30 chromosomes / 1n), pigs 2.5 Gbp (19 chromosomes / 1n), and chickens 1.05 Gbp (40 chromosomes / 1n). Generally, birds have about half the genome size of mammals.

Thanks to the sequencing of whole genomes, tremendous variability has been found in short variants (mainly SNP, indel - insertion-deletion). Cattle have 97 million, pigs have approximately 71 million, sheep have 58 million, and chickens have 22 million of these polymorphisms. There are several genomic databases available, and one of them is ENSEMBL www.ensembl.org/, or on the NCBI server.

This made it possible to perform whole-genome association analyses (GWAS) to identify QTL regions using thousands of SNP markers that evenly cover the entire genome. Results from GWAS in livestock species and humans lead to the conclusion that the individual QTL effect on complex traits is very small, and therefore a large number of QTL is needed to explain the genetic variance in these traits.

The gains from MAS programs using a small number of DNA markers to detect a limited number of QTL are small, and alternative technologies needed to be developed to use denser information through genomic SNPs, called genomic selection.
Genomic selection uses a panel of whole-genome markers where QTL are in linkage disequilibrium with one or more SNP. Estimated genomic breeding values are predicted as the sum of the effects of these SNPs across the entire genome.

Genomic selection is a higher version of MAS. This is possible thanks to the use of many SNPs discovered during genome sequencing and new methods of genotyping a large number of SNPs (DNA microarrays, SNP chips containing 60,000 or 700,000 SNPs).
The ideal method for estimating BVs from genomic data is to calculate the conditional mean of the breeding value of a given genotype of the animal at each QTL (~ marker). In practice, markers (SNPs) are used instead of QTL genotypes, but the more we approach larger sequences and SNP data, the more ideal the method will be.

The idea of genomic selection is quite old. In 2001, Meuwissen et al. proposed the concept of genomic selection as the use of a large number of genetic markers covering the entire genome to predict the genetic value of individuals.

The application of genomic selection is currently used in dairy cattle, beef cattle, and pigs. However, this varies between different countries. In the Czech Republic, it is currently only used in dairy cattle.

The basic principle of genomic selection involves genotyping of whole-genome SNP markers and utilizing this information to estimate the true genetic variability. Subsequently, based on genomic SNP markers, a kinship matrix is calculated and incorporated into the equations for Best Linear Unbiased Prediction (BLUP), which have various variants and are constantly evolving. Further, it is necessary to estimate the genomic heritability of the trait, resulting in an estimation of genomic estimated breeding value (GEBV).

The reason why genomic selection is expanding to other groups of animals is that it enables more efficient breeding by reducing costs, improving accuracy of predicted breeding values, shortening generation intervals, and reducing inbreeding.

Is there any difference in estimating breeding values between traditional and genomic methods? Estimating breeding values using traditional methods requires calculation of the additive genetic relationship matrix (A) based on the expected proportion of shared genes from parents, so knowledge of pedigree data is necessary.

Estimating genomic breeding values (~genomic selection), on the other hand, also requires knowledge of relatedness, but this time based on genotypes of genetic markers (SNPs), from which a genomic relationship matrix is calculated. Its elements express the estimates of the realized proportion of genome that two individuals share from their parents. Basically, there is no difference between traditional and genomic estimation of breeding values. However, the latter is more advantageous.

The genomic selection system is depicted here. It is always necessary to create a prediction equation for estimating genomic breeding values on a reference population and test it on reference populations with known SNP genotypes and phenotypes (performance traits). Subsequently, the resulting prediction equation is used to evaluate candidate parents (often very young, theoretically even embryos) for which we do not need to know phenotypes, only SNP genotypes. The result is a very accurate estimation of genomic breeding values for candidate parents.

In the following diagram, we can see its application in cattle. As you can see, the principle of breeding has not really changed. We start with the reality of the population from which we want to select individuals, we must know phenotypes, genetic parameters, relatedness, and in the end, from a lot of data, we have only one number, which is the value of genomic breeding value.

The advantages of genomic selection include improved accuracy of statistical estimates, shortened generation interval, increased genetic and therefore economic gain, and the ability to effectively manage inbreeding. It is particularly advantageous for traits with low heritability, sex-linked traits, or traits determined post mortem.

What about the future? Due to constant technological advancement in NGS and TGS, it will be possible to obtain more and more information from more individuals - longer sequence reads in a shorter time, which will improve SNP identification - increasing the number of detectable SNPs and reducing their false detection, leading to more informative markers. It will also be economically advantageous due to decreasing costs of whole genome resequencing to a few

hundred USD. Furthermore, the estimation of whole-genome heritability will be accurate down to the nucleotide level.

In conclusion, genomic selection is very successful in cattle as it provides greater genetic gain at similar or lower costs.

Genomic selection is a very recent innovation, yet it is quickly being implemented. Genomic selection is rapidly evolving, including reducing the cost of genotyping, strategies for phenotyping new traits, approaches for creating or replacing reference populations, and increasing the robustness and stability of genomic predictions through the identification of causal mutations from genome sequences or genomic predictions of genetics × environment interactions.

And thank you for your attention.