Topic 1: Analysis of the AnGR biodiversity status using pedigree data Lecture

Hello, welcome to the new lecture of the module Conservation and Sustainable Use of Animal Genetic Resources: Biodiversity analysis of animal genetic resources using pedigree data.

The study of genetic diversity is possible based on two types of data: data obtained from pedigrees, so-called genealogical data, and data obtained from molecular genetic analyses. This lecture will discuss genealogical data and working with them.

Pedigree data can be kept in different formats; here are various graphical representations of pedigrees and relationships between individuals. As can be seen, this type of pedigree may only sometimes be suitable for further processing.

Therefore, the presented type of data record is most often used for work with pedigree data, which is also suited for further "machine" processing of pedigrees. The so-called "three" column pedigree primarily contains the columns of individual, father, and mother. These columns are the basis of the whole pedigree file. However, the file may contain other necessary and important information, such as birth year and sex. Relationships are supported by the fact that each individual listed as a father or mother is simultaneously listed as an individual. If the pedigree is compiled correctly, the individual in the position of father or mother is always listed as an individual in one position before. It follows from the logic that an individual must first be born and then become a parent, as shown here for individual one. It may be that the relevant data for a given individual is unknown; for example, we do not know the individual's parents, the year of birth, or the sex. These unknown values are usually given as 0, as seen here.

Inbreeding coefficient is one from the elementary genetic diversity indicators.

The basic parameter on which the inbreeding coefficient is based is Identity by descent (IBD). This describes the state for copies of an allele that may be traced back through an arbitrary number of generations without mutation to a common ancestor of the organisms that carry the copies. The opposite of IBD is Identity by state - Identity by state (IBS), the coincidental possession of identical alleles (chemically similar). Last but not least, we should mention the term "Autozygosity", which is the state of the genotype where two alleles are identical by descent.

As already mentioned, the most important indicators of genetic diversity is "Inbreeding coefficient". Most commonly, the inbreeding coefficient is defined as "the probability that two homologue genes in an individual are identical by descent.

The female horse karyotype, which consists of 32 pairs of chromosomes, including XX chromosomes, is presented on the slide. As the picture shows, the chromosomes are coded by size. For the following examples, we will select only loci on the first autosomal chromosome. However, the patterns shown apply to an individual's entire genotype.

This slide shows two parent pairs, with the male being identical in both cases. The genotype of each individual is defined by a pair of chromosomes: one from the father, shown in pink, and one from the mother, shown in orange. The parental generation is not an autozygous set, as indicated by the different chromosome numbers that define the different founder individuals from which these chromosomes came, the so-called founders. In the generation of descendants of these parents, the individuals are related because they have the same father. However, they are non-inbred individuals because they have always obtained different parental chromosomes. In the generation of grandchildren, that is in the second generation of offspring, inbred individuals may already occur if they have inherited chromosomes identical by descent from their parents, that is either chromosome 1 or chromosome 2 from both father and mother.

This example can be made more concrete by marking certain alleles at a specific locus, presented by a horizontal line. The slide shows that the female in the first generation inherited identical pinkmarked alleles from both father and mother. Still, they come from different founders and are identical by state (IBS). The male has kept different allele types from both father and mother, pink from the father and orange from the mother. In the second generation of offspring, the following combinations occurred: the first individual obtained identical alleles by state (IBS), the same colour variant, but the alleles come from different founders. Thus, this is a non-inbred individual. The second individual inherited different types of alleles, presented here by different colours, so-called different by state (DBS) and different by descent (DBD), because both come from different founders. Again, this is a non-inbred individual. The last individual inherited alleles identical by descent (IBD), marked with the same colour and coming from the same founder. This individual is, therefore, inbred.

As has already been said. The IBD value is "unobservable", so we use pedigree records to derive the probability that two alleles at homologue genes in an individual are IBD. Here in the figure above, individual X would only be inbred if he inherited either the A1 or A2 allele from his father 1 or mother 3; in this case, the common ancestor would be his grandfather or the B1 or B2 allele the common ancestor would be his grandmother. In other cases, the individual would be non-inbred. The probability of individual X inheriting either the A1 or A2 allele is shown by the red segments, and the probability of inheriting either the B1 or B2 allele is shown by the blue segments. In both cases, the probability is one-half to three because there are three allele transmission paths. And since both cases can occur, the two probabilities sum together.

This image shows a case where an individual can be inbred with a probability of 0.125. In this case, the individual is Y. In the second cycle, for a given individual Y, the probability of transmitting the identical allele 'A' will no longer be 0.5 because with a probability 0.125 this individual is inbred, it means that this animal has a probability 0.125 that his alleles are IBD, and means that with this probability (0.125) it will transmit the identical allele A with 100% probability. This fact must be considered in the estimation of the inbreeding coefficient of individual X. This corresponds to the second part of the formula (1+FA), which takes this condition into account.

Here, we have a similar example. The estimation of the inbreeding coefficient for individual X is shown. Individual X has one common ancestor, ancestor D. The probability that he will get the IBD allele from his parents from this ancestor is 0.5 to the fifth because there are five possible transmission paths. It is also necessary to consider that individual D has a common ancestor, A, and the inbreeding coefficient of individual D is 0.125. Considering all these facts, the coefficient of inbreeding of individual X is equal to 0.0352 after rounding.

Based on these facts, the overall formula for estimating the inbreeding coefficient is composed of these three parts, where the sum represents all possible combinations of alleles that can occur in an individual in the IBD state.

The inbreeding coefficient can be defined in two ways. Either as "The probability that part of an individual's genome is autozygous (IBD)" or the other way, as "The probability that two randomly selected alleles at a single locus are autozygous."

Another fundamental parameter defining the level of genetic diversity based on pedigree records is the relationship coefficient. The assessment of relationship assumes that two individuals are related if they share a chromosome, chromosome segment, allele, or point mutation that is identical by descent (IBD). It implies that two individuals must have a common ancestor. Again, several approaches are used to estimate the level of relationship between two individuals. The first is the so-called Kinship or Coancestry Coefficient. This coefficient is defined as the probability that an allele randomly selected in individual X is identical by descent to a randomly selected allele in individual Y.

The coancestry coefficient can be estimated from the following relationship, where it is defined as the probability that an allele drawn at random from one individual X is identical by descent to a random allele from the other individual Y. In this case, individual Z inherited different alleles identical to descent. The relationship also shows that the coancestry coefficient of two individuals is the same as the inbreeding coefficient of their hypothetical offspring.

If we wanted to estimate the value of the coancestry coefficient between this male and female, we would get a value of 0.125. Because the only allele A is IBD. The other alleles are different by descent. For the individuals to be related, they must have the same allele, either A1 or A2. And since the probability that both the male and the female will inherit the A1 allele is 0.25 or that they will inherit the A2 allele is again 0.25. Thus, the probability of an individual inheriting either the allele A1 or the A2 allele from both the mother and the father is 0.5. Furthermore, the probability that the A1 allele

acquired by an individual from both father and mother is identical by descent is 1. It is also true for the A2 allele. Therefore, the relationship between these individuals is equal to 0.125.

Another indicator of the level of relationship between individuals is the Relationship Coefficient, according to Wright (1922). This coefficient is defined as the correlation of additive genetic value between two individuals. The relationship between the coancestry coefficient and the Wright's relationship coefficient is such that the Wright's relationship coefficient is equal to twice the coancestry coefficient, taking into account the coefficient of inbreeding of the two individuals.

This image shows that the relationship coefficient, according to Wright (1922), corresponds to twice the original coefficient. This is the same pedigree as in one of the previous slides. The estimate of the coefficient of relationship, according to Wright (1922), is 0.25, which corresponds exactly to twice the value of the coancestry coefficient.

Last but not least, it is necessary to mention the additive relationship coefficient, which is based on constructing the additive relationship matrix. This coefficient expresses a quantitative measure of similarity between two individuals, which refers to the number of alleles shared between these individuals. This coefficient takes values from zero to two. The first value expresses zero relatedness between individuals. The last value, a value of two, may occur in an individual related to itself.

One way to estimate the value of the additive relationship coefficient is to use the Tabular Method to construct an additive relationship matrix. In this matrix, the value of the inbreeding coefficient of the individual is expressed as the diagonal element of the additive relationship matrix for a given individual minus one, and the relationships between individuals are expressed on the off-diagonal element of the given additive relationship matrix.

An example of an additive relationship matrix is shown in this slide. The first column and the first row in the matrix represent our individuals and their relationship to their parents. For example, individual 3 has parents 2 and 3, which are listed in brackets. Inside the matrix, the relationships are shown. The diagonal elements of the matrix show the values of the inbreeding coefficient + one. The matrix shows that only individual five has a higher value of the diagonal elements than one, it has a value of 1.25. This means that the value of the inbreeding coefficient of this individual equal to 0.25. The off-diagonal elements marked in green, or blue represent the relationships between individuals. This means that, for example, the relationship between individual 4 and 5 has a value of 0.75.

As mentioned, the Wright's relationship coefficient is twice the coancestry or additive relationship coefficients if the individual's parents are non-inbred. Otherwise, these different approaches give different values.

It means that, if individuals X & Y are not inbred, relationship coefficient is then equal to the numerator i.e. additive genetic relationship and, also, twice kinship coefficient of those two individuals as well as twice inbreeding coefficient of their potential offspring.