Topic 4: Genomics of farm animals/Sequencing demonstration

Lecture

In this lecture, we will deal with genomics in farm animals and, above all, with sequencing. The lecture is part of module 1, animal genetics. The creation of this presentation was supported by the ERASMUS + KA2 grant within the ISAGREED project, Innovation of the content and structure of study programs in the field of animal genetic and food resource management using digitization.

DNA sequencing

DNA sequencing is a basic molecular genetic method. The method makes it possible to read genetic information, which is stored in the form of a sequence of nucleotides in a DNA molecule. In this figure, I present an overview of the most important sequencing methods from the oldest to the most modern. A chemical method based on the specific cleavage of the DNA chain and referred to by the discoverers as Maxam-Gilbert was the first sequencing method, but it is no longer used today. The method discovered by Frederick Sanger, after whom it is named, is still used today. It is also referred to as the dideoxy method because it uses specially modified nucleotides, so-called end terminators. These are added to the mixture with the normal nucleotides, and if they are incorporated into the chain, the synthesis is terminated. Pyrosequencing and sequencing by hybridization are alternative methods, the use of which is currently small. Conversely, next-generation sequencing (NGS) is very widespread and used due to its high speed and sequencing capacity, as it can also be used to sequence entire genomes and is referred to as 2nd generation sequencing. Methods based on the sequencing of a single DNA molecule, referred to as 3rd generation sequencing, are increasingly being used.

A great advantage of sequencing is that it enables simple and accurate identification of a polymorphic site or mutation. Modern automatic sequencers are also suitable for direct detection of polymorphisms.

Sanger sequencing

The principle of the method was discovered in 1977 by the English biochemist Frederick Sanger. T. The basis of Sanger sequencing are dideoxynucleotides, called ddNTPs or end terminators. It is a modified nucleotide that has removed the OH binding group on the 3rd carbon of deoxyribose, which is necessary for binding another nucleotide. Thus, upon incorporation of the ddNTP, chain synthesis stops, producing a chain of size and colour corresponding to the corresponding nucleotide on the template.

Overview of sequencing

The sequencing process consists of several steps. It starts with sample preparation, which is most often a PCR product. This must be purified to contain only template DNA. Furthermore, we need to know at least its approximate concentration. The sequencing reaction itself is a linear cyclic enzymatic reaction that must contain the listed components. The sequencing primer defines the start of sequencing, only one must be used, not two as in PCR. The reaction is catalysed by DNA polymerase inside the reaction buffer with the addition of standard dNTPs (similar to PCR) and additionally ddNTPs. This is followed by purification of the sequencing

reaction, when free labelled ddNTPs must be removed. The mixture of fragments is divided by capillary electrophoresis into sequencers and the result evaluated.

Four-color terminator sequencing

Currently, a variant known as four-color terminator sequencing is routinely used. Individual termination nucleotides are marked with fluorescent colours according to the nitrogenous base they carry, or which nucleotide on the template they pair with. By the action of the polymerase, chains complementary to the template are synthesized, while usually normal dNTPs are incorporated and the chain is lengthened, but occasionally ddNTPs are also incorporated, which then terminate the synthesis of the chain. The resulting fragment is labelled with a fluorescent colour that corresponds to the nucleotide at the appropriate position of the template.

Sequencer

A mixture of single-chain molecules of different lengths and different colours is formed. In order to determine at which position which base occurs, an accurate electrophoretic separation of this mixture of fragments must take place. To do this, it uses fluorescent capillary electrophoresis, which is the basis of genetic analysers, or sequencers. This device can not only divide and sort the fragments according to size, but also directly read the sequence of nucleotides based on different coloured peaks. E.g. the 100 nucleotides long chain emits green light, i.e. there was adenine in position 100 of the template, the chain 101 long lights up red, i.e. in position 101 was thymine etc.

Advantages and disadvantages of Sanger sequencing

The Sanger method is the most widely used sequencing method. Among its advantages are the reading of relatively long DNA chains (approx. 800 bp) and at the same time high accuracy and reliability. If we need to know the sequence or detect a mutation in a specific section of the genome in one or a few individuals, this is the most cost-effective option. Compared to NGS methods, however, the performance and price per 1 sequenced base is very high, so the method is not suitable for sequencing large sections of the genome, or even entire genomes.

New generation sequencers

For whole genome sequencing, it is more appropriate to use another method, the so-called next-generation sequencing, NGS for short. The method is also referred to as second generation sequencing. The method is based on DNA fragmentation and sequencing of short fragments, but in huge (hiudž) quantities at the same time, and is thus referred to as massively parallel sequencing. The method enables a generally very high sequencing capacity, however, you can choose sequencer variants with a capacity according to your needs from 1 GB to 8 TB.

NGS Illumina

Common NGS sequencers include Illumina devices. The basis of the method is the so-called bridge PCR and sequencing during synthesis using 4-color fluorescence. The actual sequencing takes place in clusters of the same sequence - the corresponding spot lights up in colour according to the base. Sections from 150 to 300 nucleotides in size are sequenced. A more detailed explanation is beyond the time frame of this lecture, and I recommend watching videos on the Internet for those interested in understanding this rather complex method.

NGS summary

Next-generation sequencing is today a widely used tool in genetics for both research and diagnosis. The high capacity of the method makes it possible to obtain sequences of even large genomes of animals, e.g. mammals, the size of which is around 3 billion nucleotides, within a few hours to days, which would take several years of work using the Sanger sequencing method. It is possible to obtain the entire genetic information of an individual in this way. It therefore allows finding a large number of new or detecting all known polymorphisms and mutations. However, there is a problem with the processing and evaluation of a large amount of data, which is why it is necessary to use very powerful bioinformatics tools. Whole-genome data are stored in the genome databases of individual organisms, and can be compared with a specific sample. The method is also important for sequencing an individual, e.g. a patient, which is the field of personal genomics.

3rd generation sequencing

Sequencing techniques based on the sequencing of a single molecule are referred to as 3rd generation sequencing. The advantage is long reads, in the case of de novo sequencing up to 7 kb. This makes it possible to correctly identify variants on one chain, the so-called haplotypes, which reading of short sequences in the 2nd generation does not allow. The method is gradually being improved, as the reading error rate is still higher compared to the two previous generations. A method known as Single molecule real time sequencing from Pacific Bioscience or nanopore sequencers from Oxford Nanopore is available.

Physical map from genome sequencing

The results of whole-genome sequencing are stored in genomic databases and often freely available, for example the NCBI internet database allows viewing down to the sequence level of individual gene nucleotides.

Species identification using sequencing

Another practical use of sequencing in animals is species identification. Molecular taxonomy uses so-called DNA barcoding, which makes it possible to identify a species where it is not possible with the classical method, e.g. in insect larvae. A fragment of the gene for cytochrome C oxidase I, which is located in mitochondrial DNA, is used for this purpose. This fragment is usually sequenced by the Sanger method, and species identification is performed by comparison with determined sequences stored in the Bold or Blast databases. The advantage of mtDNA is more copies, i.e. greater amounts of DNA per amount of biological material and higher stability, it is also used for the determination of older or museum samples. The disadvantage is the possibility of contamination with a bacterial genome (Wolbachia, etc.).

DNA barcoding

This image shows a sample of a specific sequence obtained by mtDNA sequencing of an old museum butterfly specimen.

Species Identification (Blast)

By inserting the sequence into the internet alignment search tool called Blast, it was possible to identify the species based on homology.

The lecture is finished. I believe that you have understood the basic principles of one of the most important molecular-genetic methods - sequencing. I also recommend watching the follow-up presentation Laboratory examples. Thank you for your attention.