Genomics of farm animals/ Sequencing demonstration

Lecture

Modul no. 1: Animal Genetics

Aleš Knoll

Mendel University in Brno, Faculty of AgriScience

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED

Co-funded by the European Union

DNA sequencing

= determining the sequence of DNA nucleotides

a) chemical method (formerly)

b) ddNTP method, cyclic reaction (asymmetric PCR), today prevails, also Sanger sequencing (so-called 1st generation)

c) pyrosequencing

d) sequencing by hybridization

e) NGS - sequencing of the new generation, modern large-format applications, parallel sequencing of fragments (2nd generation)

f) single molecule sequencing (3rd generation)

- accurate identification of the polymorphic site
- automatic sequencers (also for direct detection of polymorphisms)

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

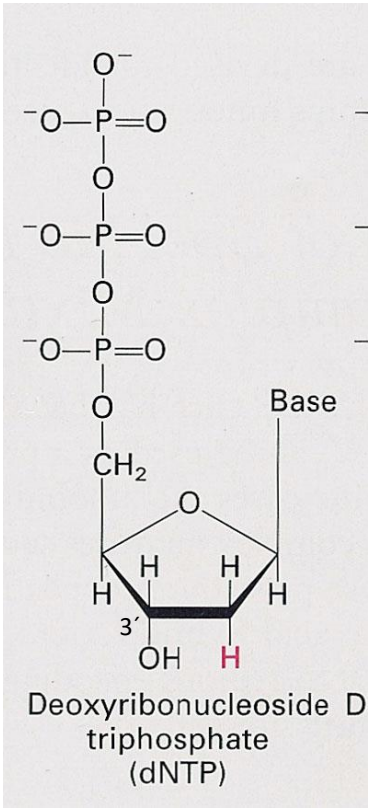
Co-funded by the European Union

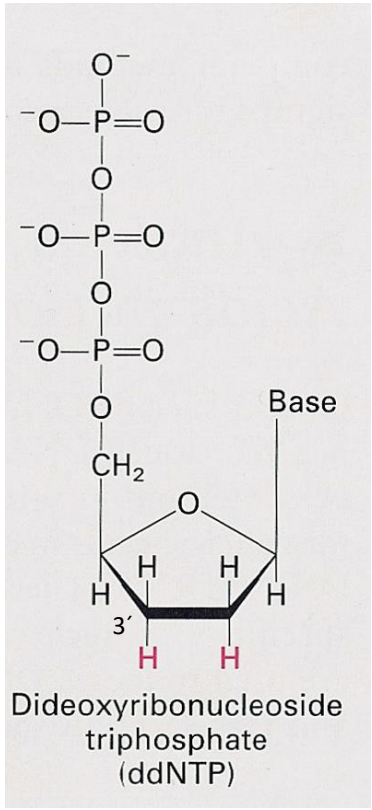
Sanger sequencing

chain termination method: dideoxyribonucleoside triphosphates (ddNTPs)

they are labeled with 4 fluorescent colors according to the base

Frederick Sanger





standard nukleotide (dNTP) enables the binding of another nucleotide to the 3' end

modified nucleotide (ddNTP) terminates chain synthesis after binding

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

Overview of sequencing process

1. Sample preparation: usually PCR product
    1.1. Sample cleanup
    1.2. Determination of DNA template concentration

2. Sequencing reaction - cyclic
    sequencing primer (1 only)!
    polymerase
    buffer (+Mg^{2+})
    dNTPs
    labelled ddNTP

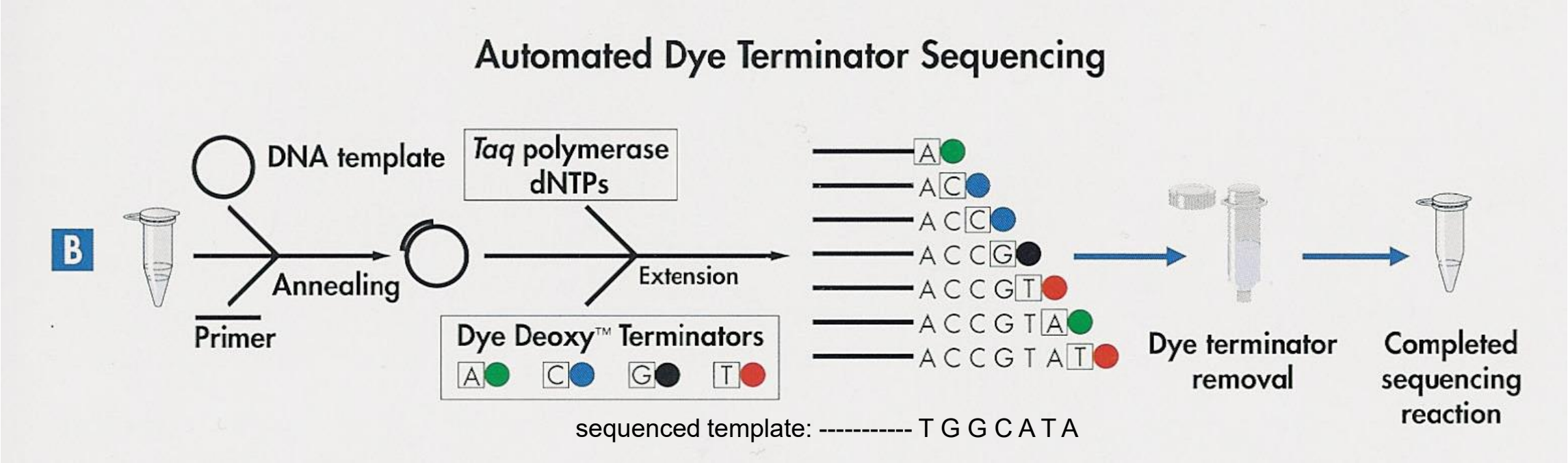
3. Purification of the sequencing mixture – removal of free ddNTPs

4. Electrophoresis (capillary) on the sequencer5. (Automatic) evaluation - determining the sequence

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

Four-color terminator sequencing

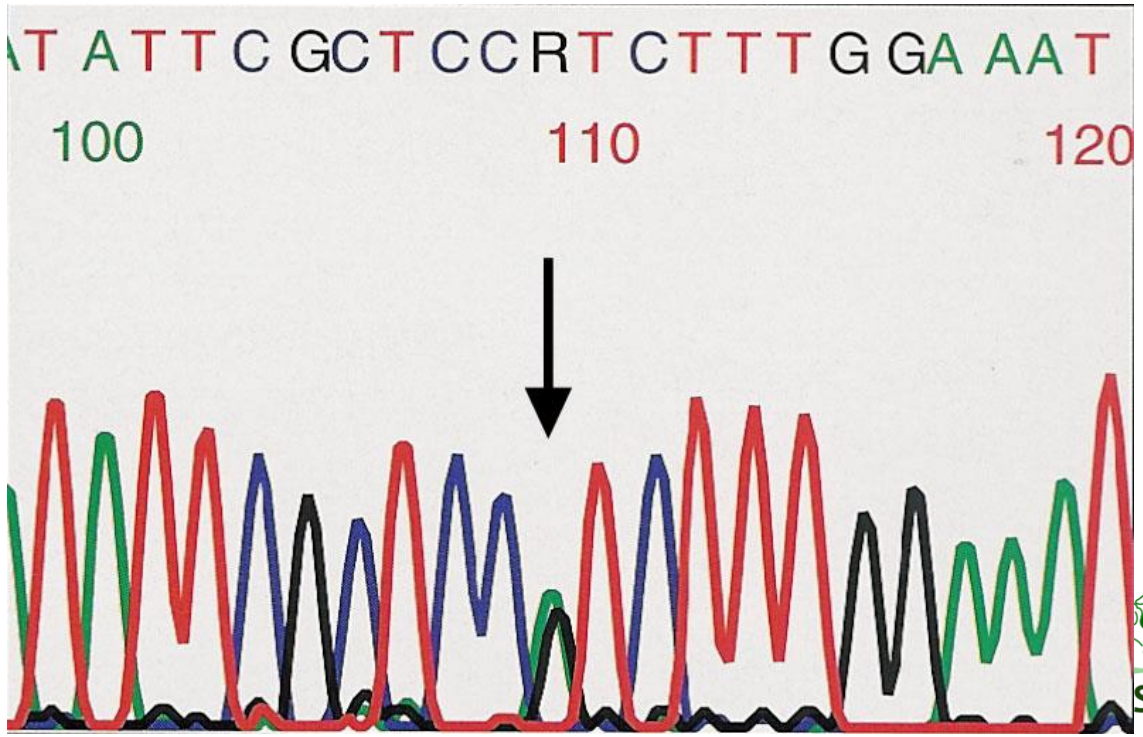


labeled terminators, everything takes place in 1 reaction, including electrophoresis, current variant

Sequencer

By capillary electrophoresis, single strands of different lengths are separated by size and the corresponding fluorescent color is read, which indicates the base present at that DNA site.

E.g., the 100 nucleotides long chain emits green light, i.e., there was an adenine at position 100 of the template



Co-funded by the European Union

Advantages and disadvantages of Sanger sequencing

- long reads (up to 800 bp)
- high accuracy, reliability
- low overall price
- low performance, high price per base sequenced
- suitable for sequencing individual genes and detecting mutations in known sequences

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

New generation sequencers

so-called "New Generation Sequencing", NGS

second generation sequencing

- used for whole-genome sequencing

- they sequence in small pieces, but in a huge number of these fragments at the same time (massively parallel sequencing)

- enables a high capacity of sequencers and a significant reduction in the price per sequenced base

- e.g. Illumina; Life Technologies SOLiD 3 and more

ISAGREED

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

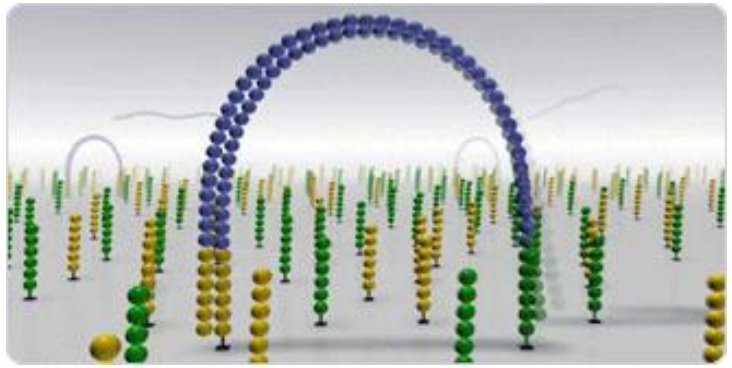
NGS Illumina

Illumina: NovSeq : aprox. 8 TB

NextSeq: up to 120 GB (formerly HiSeq 600GB)

MiSeq: up to 15 GB

MiniSeq: 7,5 GB



*sequencing by synthesis

Principle: bridge PCR, 4 SBS* fluorescence sequencing, read 150 to 300 bp, sequencing in clusters of the same sequence– the corresponding spot lights up in color according to NTP.

NextSeq

MiSeq

Co-funded by the European Union

NGS summary

- the size of mammalian genomes is around 3 billion bp
- covers the entire genetic information of an individual
- allows finding new or identifying existing polymorphisms and mutations
- the problem is a large amount of data and its interpretation
- results stored in genomic databases (ncbi etc.)
- possibility of individual sequencing (personal genomics)

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

Third-generation sequencing

Single Molecule Technologies

- sequencing of individual molecules
- there are no amplification errors
- long reads (de novo sequencing) – up to 7kb!!! (3.5-8)
- a small amount of template
- high reading error rate so far (that's why the sequence must be read repeatedly)

example:

SMRT – Single Molecule Real Time Sequencing (fa. Pacific Bioscience)

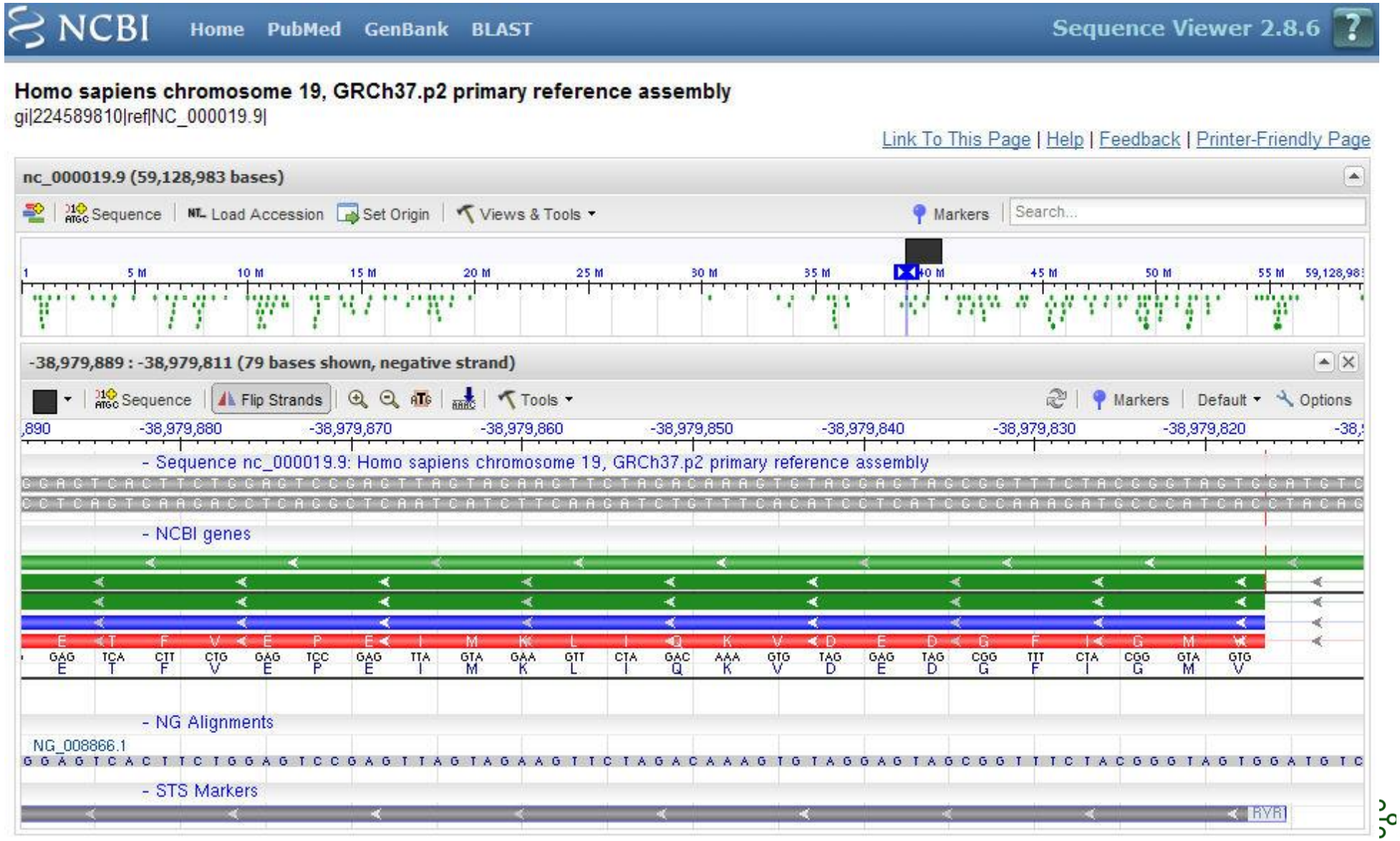
nanopore technology (Oxford Nanopore)

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

Physical map from genome sequencing

the NCBI Internet database allows viewing of genes down to the nucleotide sequence level



Co-funded by the European Union

Species identification using sequencing

Molecular taxonomy, **DNA barcoding**

makes it possible to identify a species where it is not possible with the classical method, e.g. in the case of insect larvae

a) mtDNA (COI)

b) nuclear genes

Sequencing of a specific fragment, comparison in Bold or Blast databases.

The advantage of mtDNA is a multicopy, i.e. greater amounts of DNA per amount of biol. material and higher stability. The disadvantage is the possibility of contamination with a bacterial genome (Wolbachia, etc.).

Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

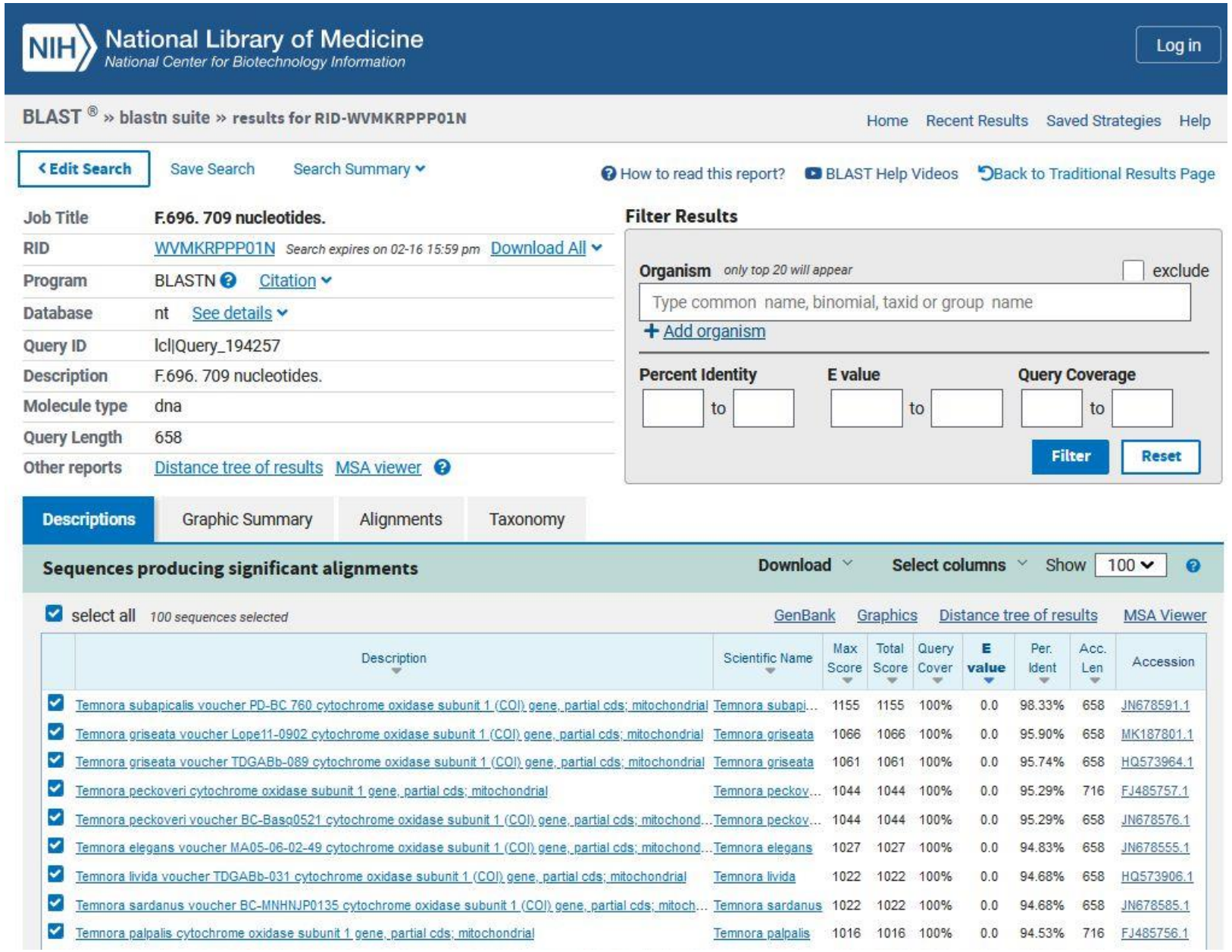
DNA barcoding

Example of Bar mtDNA sequence of a museum butterfly specimen

```
>F.696.  709 nucleotides.
AACATTATATTTTATTTTTGGAATTTGAGCAGGTATAGTAGGAACTTCATTAAGA
TTACTAATTCGAGCGGAATTAGGAACCCCAGGATCTTTAATTGGAGATGATCAAATTTATAATACAATTGTAACAGCCCA
TGCATTTATTATAATTTTTTTCATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTCCCTTTAATGTTAG
GAGCCCCTGATATAGCATTCCCACGAATAAATAATATAAGATTTTGACTTTTACCACCATCTTTAACTCTTTTAATTTCC
AGAAGTATCGTAGAAAATGGAGCAGGAACAGGATGAACAGTTTATCCCCCCCTATCTTCTAACATTGCCCATAGAGGTAG
TTCAGTAGATCTTGCAATTTTTTCTTTACATTTAGCTGGAATTTCATCTATTATAGGAGCAGTTAATTTTATTACCACAA
TTATTAATATACGAATTAATAATATATCATTCGATCAAATACCATTATTTATTTGAGCTGTGGGAATTACAGCATTTTTA
TTACTTTTATCTTTGCCAGTATTAGCTGGAGCTATTACTATACTTTTAACCGATCGTAATCTTAATACATCATTTTTTGA
TCCAGCTGGAGGAGGTGATCCTATTTTATACCAACATTTATTT
```


Co-funded by the European Union

Species identification (Blast)



nucleotide Blast:
https://blast.ncbi.nlm.nih.gov/Blast.cgi

highest homology:

= *Temnora subapicalis*

Co-funded by the European Union

ISAGREED

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.*

Aleš Knoll

knoll@mendelu.cz

Co-funded by the European Union