ISAGREED

# Studying the genetic diversity and structure of populations using mtDNA and nuclear microsatellite markers in honey bee


## Lecture

**Modul no. 1: Animal Genetics**

Tomáš Urban, Aleš Knoll

Mendel University in Brno, Faculty of AgriScience

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

ISAGREED

Co-funded by the European Union

# Contents

- Obtaining genetic data
- Using mtDNA sequences to assess genetic variability
- Assessment of genetic variability using nuclear STR markers (microsatellites)



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

ISAGREED
Co-funded by the European Union

# Why is genetic diversity important?

- Genetic diversity for population health, essential for tolerance and resistance to current and future diseases, pathogens and predators.
- Morphometric traits (wing parameters, pigmentation...) were used to assess bee diversity.
- 31 subspecies (breeds, races) of honey bees are described
- Using DNA analyses (mainly mitochondrial), evolutionary lineages have been described: the Western Mediterranean **M**, the Northern Mediterranean **C**, the African **A**, the Oriental **O**...

ISAGREED

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

ISAGREED
Co-funded by the European Union

# *Apis mellifera* genome

| Chromosome | GenBank | RefSeq | Size (bp) | GC content (%) |
|---|---|---|---|---|
| LG1 | CM000054.5 | NC_007070.3 | 29 893 408 | n/a |
| LG2 | CM000055.5 | NC_007071.3 | 15 549 267 | n/a |
| LG3 | CM000059.5 | NC_007072.3 | 13 234 341 | n/a |
| LG4 | CM000060.5 | NC_007073.3 | 12 718 334 | n/a |
| LG5 | CM000058.5 | NC_007074.3 | 14 363 272 | n/a |
| LG6 | CM000056.5 | NC_007075.3 | 18 472 937 | n/a |
| LG7 | CM000063.5 | NC_007076.3 | 13 219 345 | n/a |
| LG8 | CM000061.5 | NC_007077.3 | 13 546 544 | n/a |
| LG9 | CM000062.5 | NC_007078.3 | 11 120 453 | n/a |
| LG10 | CM000068.5 | NC_007079.3 | 12 965 953 | n/a |
| LG11 | CM000057.5 | NC_007080.3 | 14 726 556 | n/a |
| LG12 | CM000067.5 | NC_007081.3 | 11 902 654 | n/a |
| LG13 | CM000064.5 | NC_007082.3 | 10 288 499 | n/a |
| LG14 | CM000065.5 | NC_007083.3 | 10 253 655 | n/a |
| LG15 | CM000066.5 | NC_007084.3 | 10 167 229 | n/a |
| LG16 | CM000069.5 | NC_007085.3 | 7 207 165 | n/a |
| MT | na | NC_001566.1 | 16 343 | 15 |

Total number of genes

12 398

Protein-coding

9 935

Non-coding genes

2 421

**CHROMOSOMES OF HONEY BEES**

ISAGREED

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

ISAGREED

Co-funded by the European Union

ISAGREED

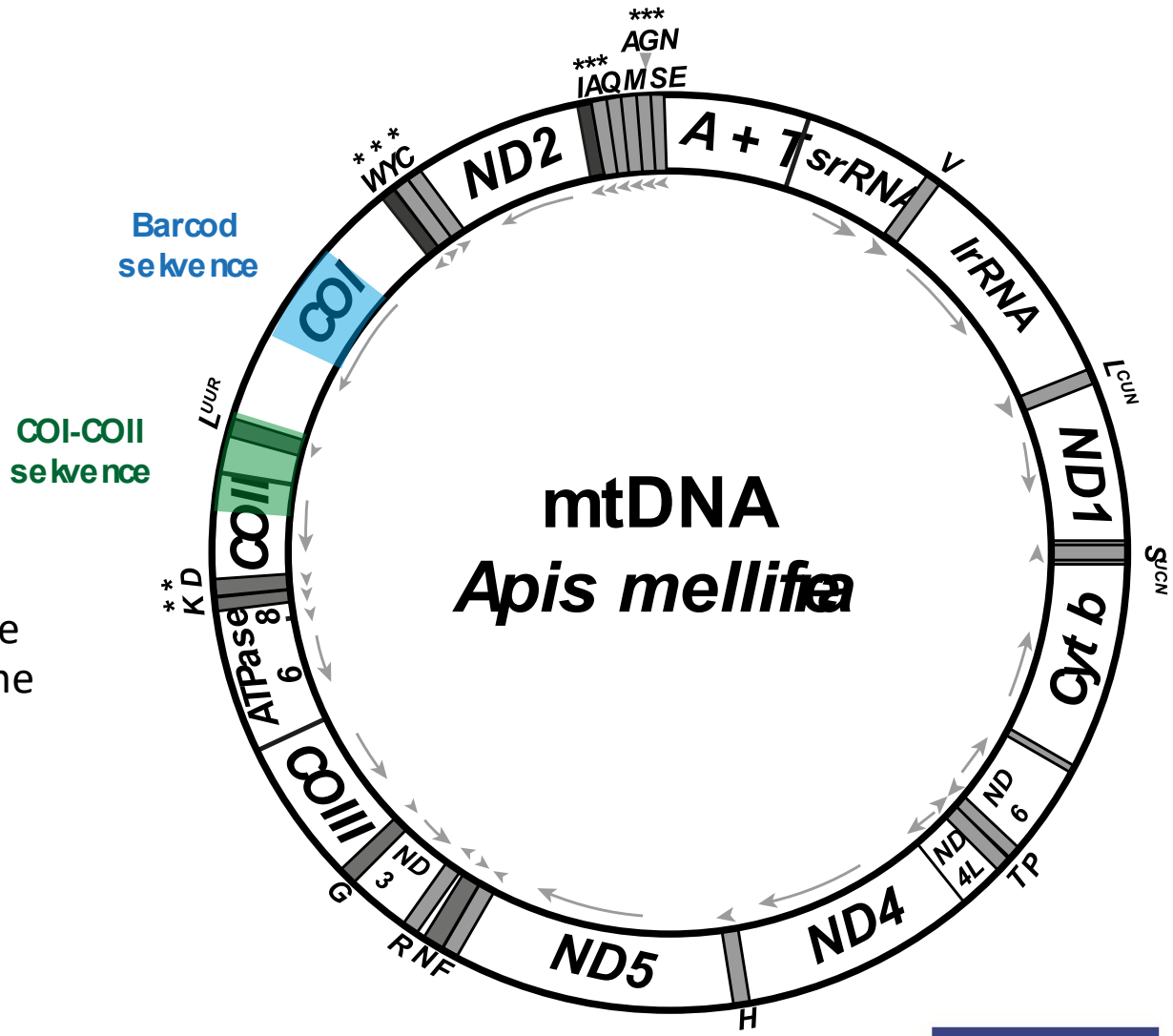
- The mitochondrial DNA (mtDNA) of most animal species is between 16 and 20 kb. The mitochondrial genome of the western honeybee is approximately 16.5-17 kb.
- The reference genome NC_001566 is 16,343 bp.
- The complete mtDNA genome in Carniolan race (MN250878) is 16,358 bp.

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

Co-funded by the European Union

# mtDNA

- **13 protein-coding genes**
- 22 genes tRNA
- genes rRNA (s and l-rRNA)




- R H Crozier, Y C Crozier, The mitochondrial genome of the honeybee Apis mellifera: complete sequence and genome organization., *Genetics*, Volume 133, Issue 1, 1 January **1993**, Pages 97–117, https://doi.org/10.1093/genetics/133.1.97

**mtDNA is used for phylogeography and genetic studies , no or low recombination, advantageous for the study of evolution, conserved structure and uniparental inheritance.**

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

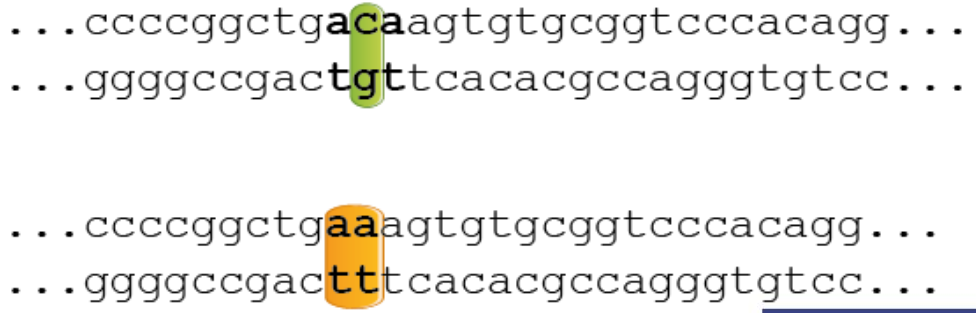
Co-funded by the European Union

# What types of variability can be observed with the use of molecular genetic markers?

- bi-allelic single nucleotide polymorphism (**SNP**) in coding or, more commonly, non-coding intron or intergenic regions


- inzertion-deletion of nucleotide (**indel**)

Deletion C in region ACA.




ISAGREED

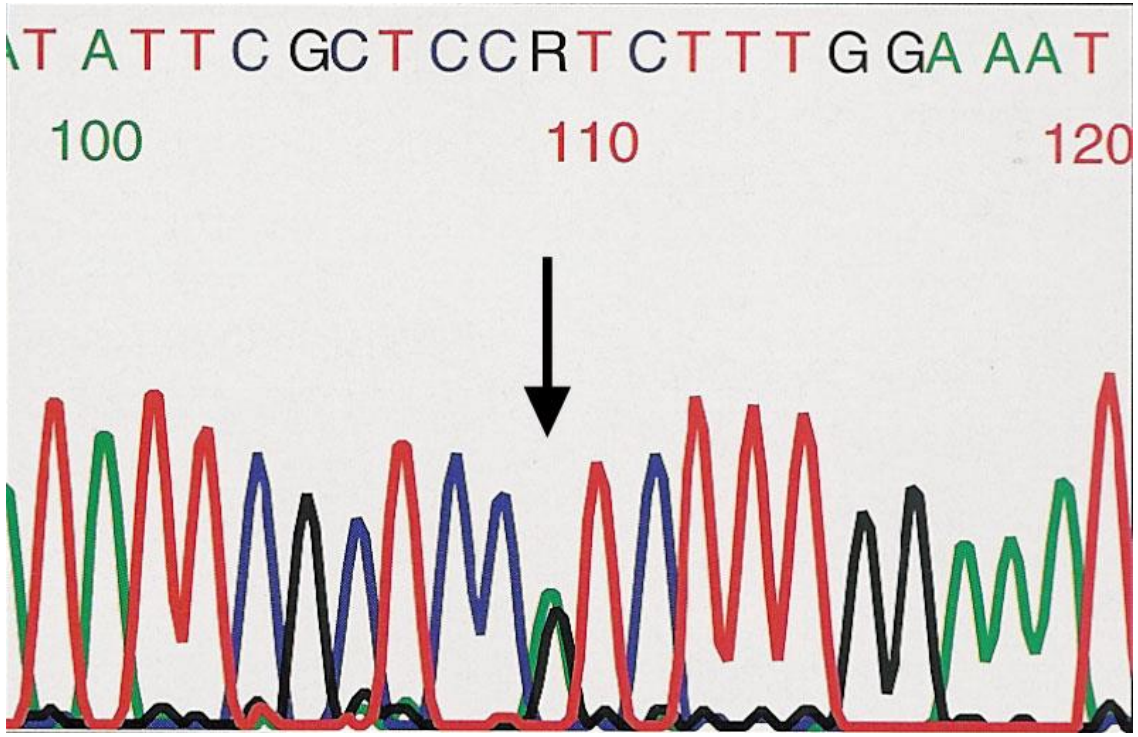
Co-funded by the European Union

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

# 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.

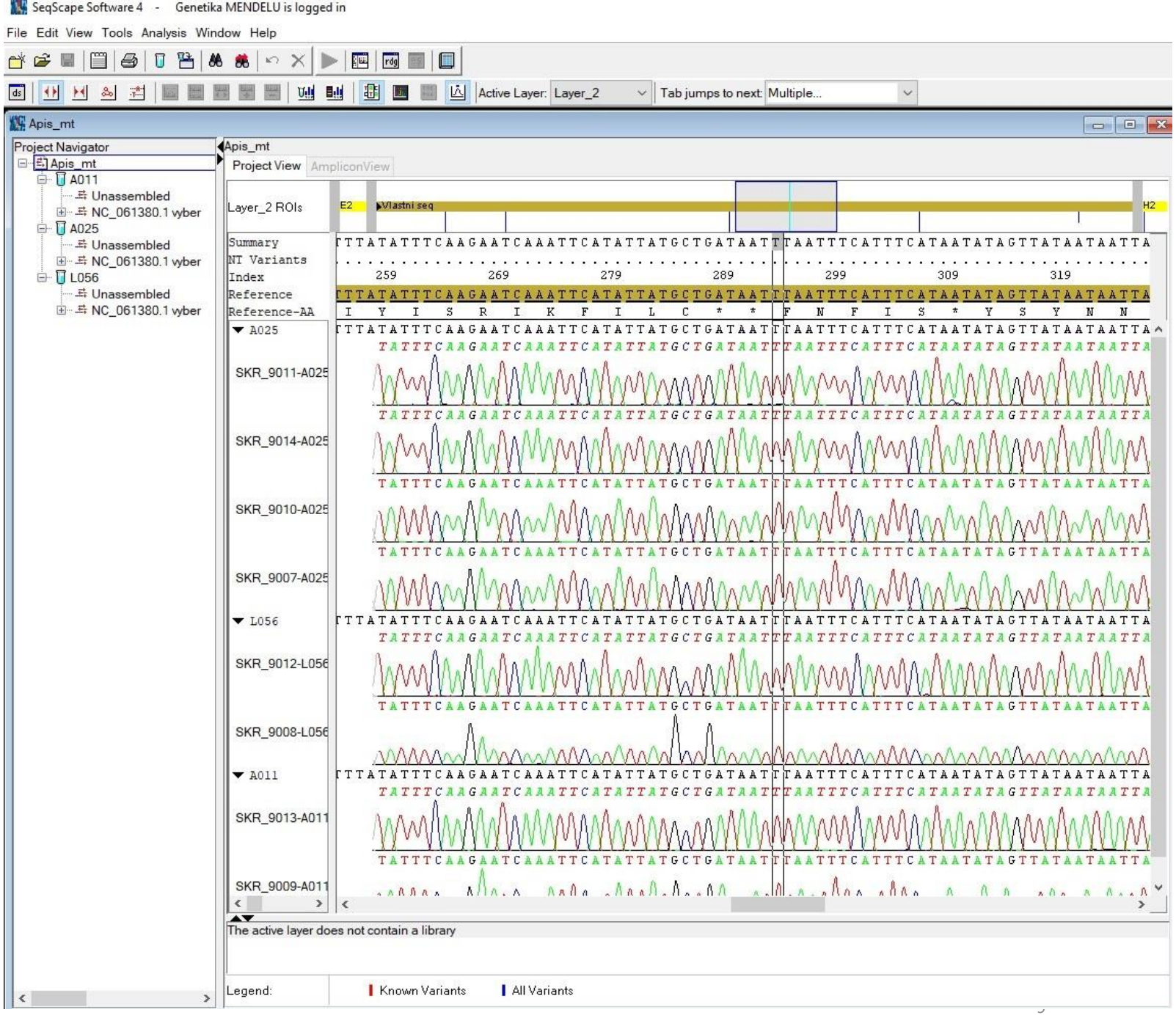
E.g., the 100 nucleotides long chain emits green light, i.e., there was an adenine at position 100 of the template







ISAGREED
Co-funded by the European Union





mt DNA - sekvenování



ISAGREED
Co-funded by the European Union

# Identification of sublines (mitotypes) in line C (carnica)

In particular, the following two mtDNA sequences are used:

- Sequencing of intergenic region COI-COII (***tRNA$^{leu}$-cox2***)
  - consists of two mitochondrial genes: transfer RNA (tRNA)-Leu, and cytochrome oxidase subunit II (*cox2*)
  - This fragment contains a high level of AT and shows significant differences in nucleotide length and composition between honeybee populations. This amplicon is cleaved by the *Dra*I restriction endonuclease, which specifically recognizes the TTTAAA sequence, to determine the lineage.
- Sequencing of the barcoding region in the COI gene(***cox1***)
  - This is compared to sequences stored in databases like the BoldSystem and GenBank. This DNA fragment is highly conserved within a taxon and is often used to distinguish between taxa.

ISAGREED

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

Co-funded by the European Union

# Evolutionary lineage by region *tRNA$^{leu}$-cox2* in *Apis meliffera*

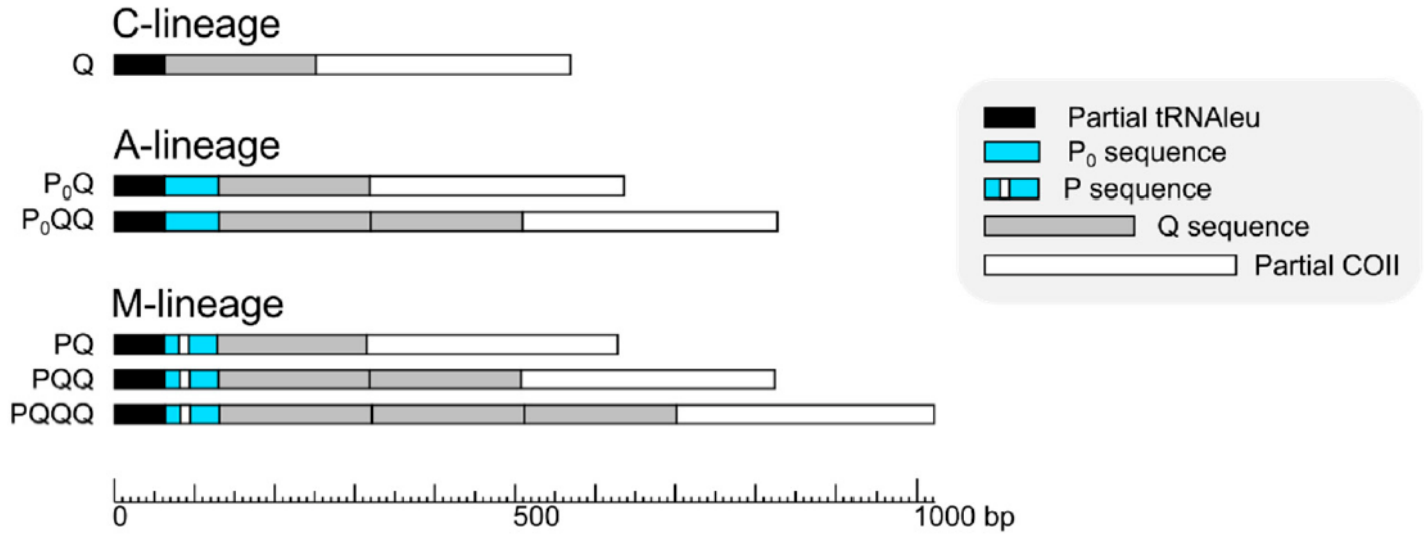


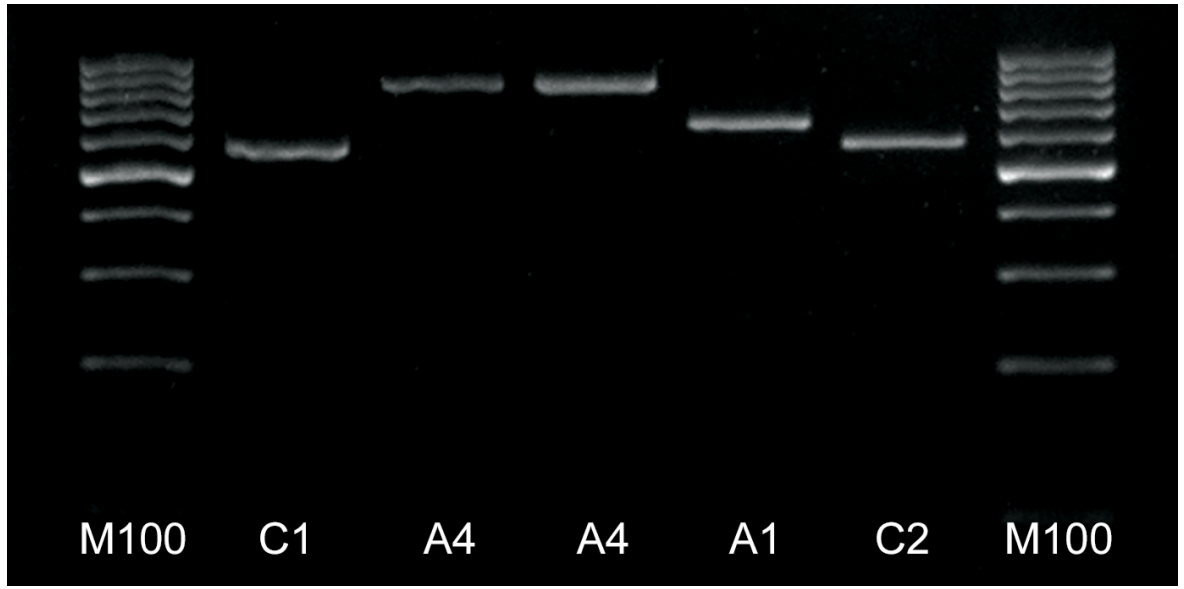

**Figure 2.** Schematic representation of COI-COII intergenic region of the mtDNA examined in this study. It is composed of partial sequences of tRNAleu and COII genes, and a variable number of non-coding sequences P and Q. Bees of C-lineage have only one Q-sequence and no P-sequences, while A- and M-lineages have up to three Q-sequences and one P-sequence, which occurs in two variants, i.e., $P_0$ in A-lineage, and P in M-lineage.

Oleksa et al. 2021



Co-funded by the European Union

# Cleavage *tRNA$^{leu}$-cox2* sequence by restriction enzyme




**Any deviation (due to mutation) will not be recognized -> the site will not be cleaved**
**The result is the formation of several fragments, the number and position of which are used to determine the evolutionary lineage (M, C, A, O…)**

- PCR product of the *tRNA$^{leu}$-cox2* region of mitochondrial DNA by agarose gel electrophoresis. M: 100 bp DNA ladder, C1 and C2 (Q), A4 ($P_0QQ$) and A1 ($P_0Q$) lineages

ISAGREED

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

Co-funded by the European Union

ISAGREED

```
>A054.  709 nucleotides.
----------------------------------------------------GATCTGGAaTACTAGGATCATCAATGAGA
CTTATTATTCGAATAGAATTAAGATCCCCAGGATCATGAATTAACAATGATCAAATTTATAATACAATTGTTACTAGTCA
TGCATTCCTAATAATTTTTTTTTATAGTTATACCATTTTTAATTGGAGGATTTGGAAATTGGCTTATTCCTTTAATACTAG
GATCACCTGATATAGCATTCCCCCGAATAAATAATATTAGATTTTGATTACTTCCTCCCTCATTATTTATACTTTTATTA
AGAAATTTATTTTATCCAAGACCAGGAACTGGATGAACAGTATATCCACCATTATCAGCATATTTATATCATTCTTCACC
TTCAGTAGATTTTGCAATTTTTTCTCTTCATATATCAGGAATTTCCTCAATTATAGGATCATTAAACTTAATAGTTACAA
TTATAATAATAAAAAATTTTTCTATAAATTATGACCAAATTTCATTATTTCCATGATCAGTTTTTATTACAGCAATTTTA
TTAATTATATCATTACCTGTATTAGCTGGAGCAATTACTATACTATTATTTGATCGAAATTTTAATACATCATTTTTCGA
TCCTATAGGAGGTGGAGATCCAATTCTTTATCAACATTTATTT-------------------------
>A071.  709 nucleotides.
-------------------------GATCTTGTATATTATTCTAGCTTTATGATCTGGAATACTAGGATCATCAATGAGA
CTTATTATTCGAATAGAATTAAGATCCCCAGGATCATGAATTAACAATGATCAAATTTATAATACAATTGTTACTAGTCA
TGCATTCCTAATAATTTTTTTTTATAGTTATACCATTTTTAATTGGAGGATTTGGAAATTGGCTTATTCCTTTAATACTAG
GATCACCTGATATAGCATTCCCCCGAATAAATAATATTAGATTTTGATTACTTCCTCCCTCATTATTTATACTTTTATTA
AGAAATTTATTTTATCCAAGACCAGGAACTGGATGAACAGTATATCCACCATTATCAGCATATTTATATCATTCTTCACC
TTCAGTAGATTTTGCAATTTTTTCTCTTCATATATCAGGAATTTCCTCAATTATAGGATCATTAAACTTAATAGTTACAA
TTATAATAATAAAAAATTTTTCTATAAATTATGACCAAATTTCATTATTTCCATGATCAGTTTTTATTACAGCAATTTTA
TTAATTATATCATTACCTGTATTAGCTGGAGCAATTACTATACTATTATTTGATCGAAATTTTAATACATCATTTTTCGA
TCCTATAGGAGGTGGAGATCCAATTCTTTATCAACATTTATTT-------------------------
>A1026.  709 nucleotides.
-------------------------GATCTTGTATATTATTCTAGCTTTATGATCTGGAATACTAGGATCATCAATGAGA
CTTATTATTCGAATAGAATTAAGATCCCCAGGATCATGAATTAACAATGATCAAATTTATAATACAATTGTTACTAGTCA
TGCATTCCTAATAATTTTTTTTTATAGTTATACCATTTTTAATTGGAGGATTTGGAAATTGGCTTATTCCTTTAATACTAG
GATCACCTGATATAGCATTCCCCCGAATAAATAATATTAGATTTTGATTACTTCCTCCCTCATTATTTATACTTTTATTA
AGAAATTTATTTTATCCAAGACCAGGAACTGGATGAACAGTATATCCACCATTATCAGCATATTTATATCATTCTTCACC
TTCAGTAGATTTTGCAATTTTTTCTCTTCATATATCAGGAATTTCCTCAATTATAGGATCATTAAACTTAATAGTTACAA
TTATAATAATAAAAAATTTTTCTATAAATTATGACCAAATTTCATTATTTCCATGATCAGTTTTTATTACAGCAATTTTA
TTAATTATATCATTACCTGTATTAGCTGGAGCAATTACTATACTATTATTTGATCGAAATTTTAATACATCATTTTTCGA
TCCTATAGGAGGTGGAGATCCAATTCTTTATCAACATTTATTT-------------------------
>A1065.  709 nucleotides.
-------------------------GATCTTGTATATTATTCTAGCTTTATGATCTGGAATACTAGGATCATCAATGAGA
CTTATTATTCGAATAGAATTAAGATCCCCAGGATCATGAATTAGCAATGATCAAATTTATAATACAATTGTTACTAGTCA
TGCATTCCTAATAATTTTTTTTTATAGTTATACCATTTTTAATTGGAGGATTTGGAAATTGGCTTATTCCTTTAATACTAG
GATCACCTGATATAGCATTCCCCCGAATAAATAATATTAGATTTTGATTACTTCCTCCATCATTATTTATACTTTTATTA
AGAAATTTATTTTATCCAAGACCAGGAACTGGATGAACAGTATATCCACCATTATCAGCATATTTATATCATTCTTCACC
TTCAGTAGATTTTGCAATTTTTTCTCTTCATATATCAGGAATTTCCTCAATTATAGGATCATTAAACTTAATAGTTACAA
TTATAATAATAAAAAATTTTTCTATAAATTATGACCAAATTTCATTATTTCCATGATCAGTTTTTATTACAGCAATTTTA
TTAATTATATCATTACCTGTATTAGCTGGAGCAATTACTATACTATTATTTGATCGAAATTTTAATACATCATTTTTCGA
TCCTATAGGAGGTGGAGATCCAATTCTTTATCAACATTTATTT-------------------------
```

# Individual sequences of *tRNA$^{leu}$-cox2*



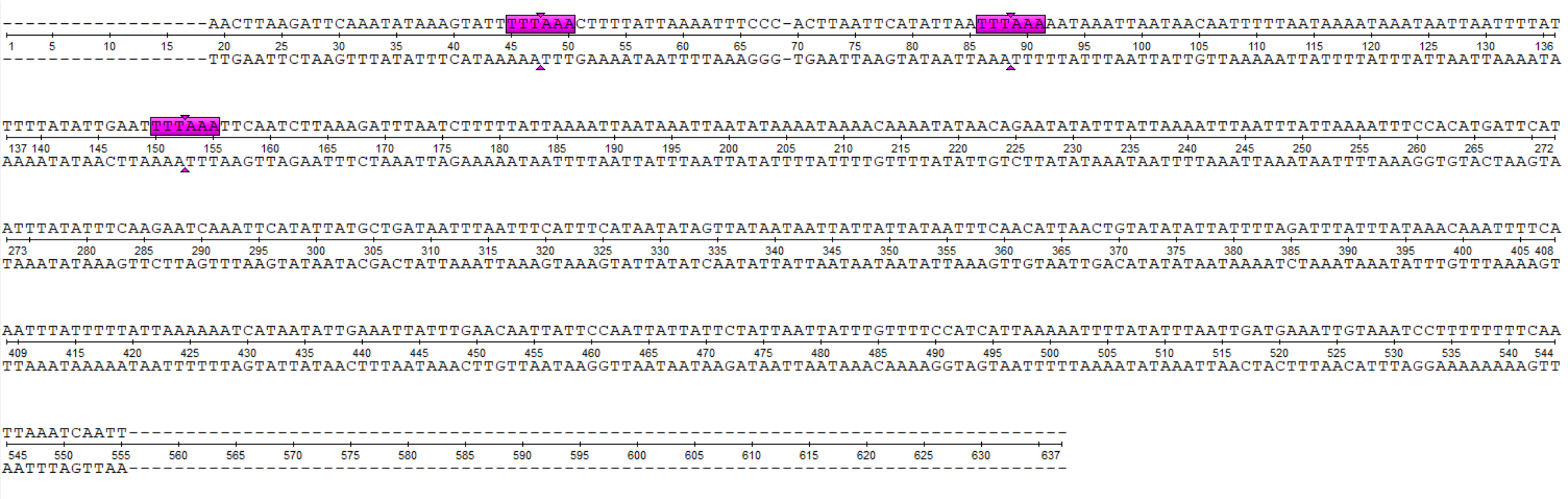
ISAGREED

Co-funded by the European Union

# *Dra*I test for area *tRNA$^{leu}$-cox2*

## lineage C

The *Dra*I restriction spectra of the *tRNA$^{leu}$-cox2* fragment sequences were performed in silico using the **Unipro UGENE**



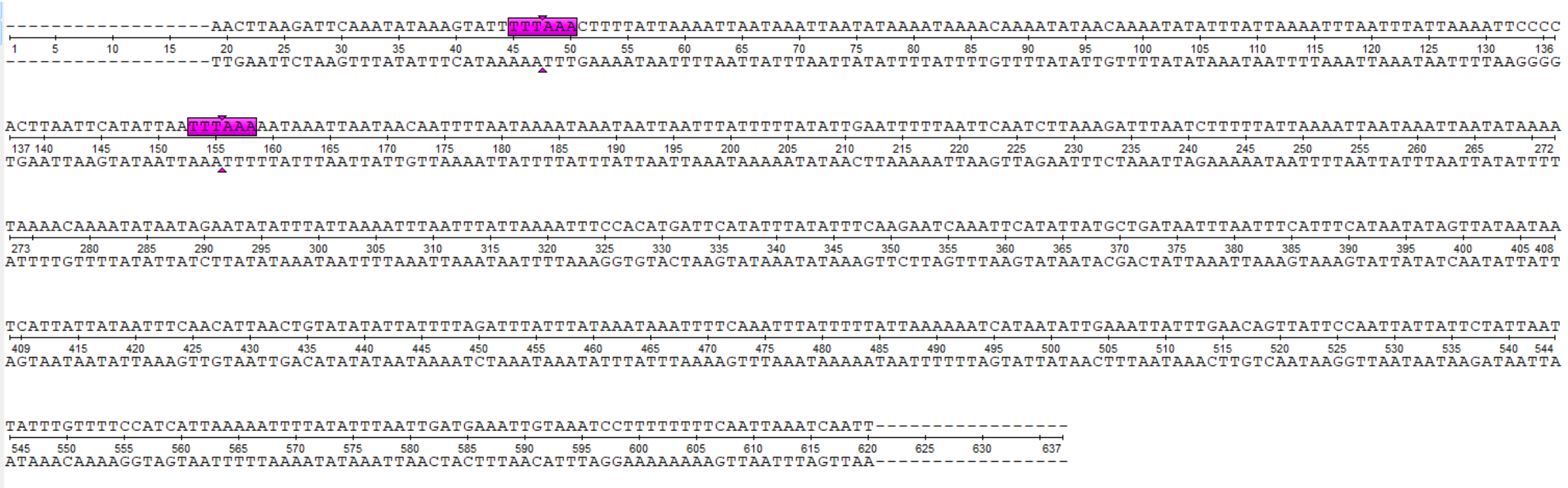

| | | |
|---|---|---|
| DraI | Restriction Site | 45..50 |
| DraI | Restriction Site | 86..91 |
| DraI | Restriction Site | 149..154 |

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

ISAGREED

Co-funded by the European Union

# lineage A




| | | |
|---|---|---|
| DraI | Restriction Site | 45..50 |
| DraI | Restriction Site | 153..158 |

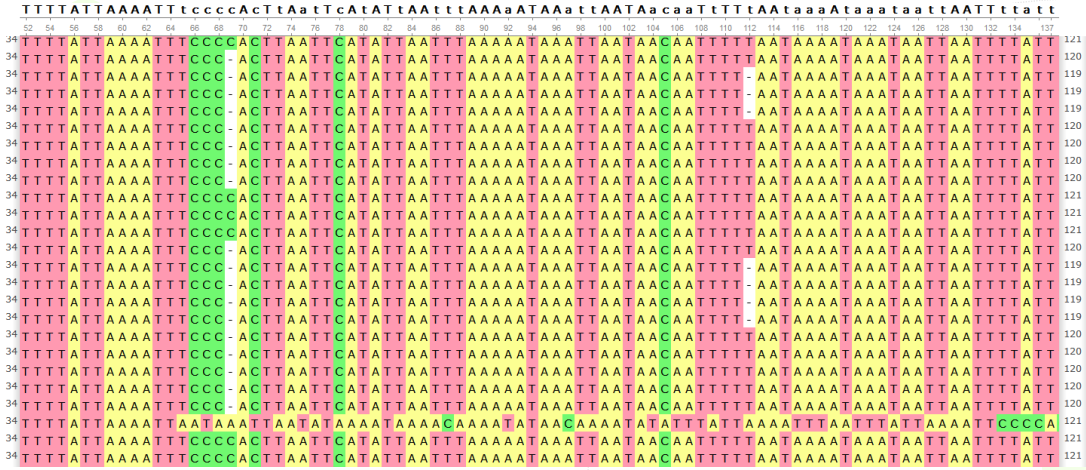
Erasmus+ project 2021-1-SK01-KA220-HED-000032068

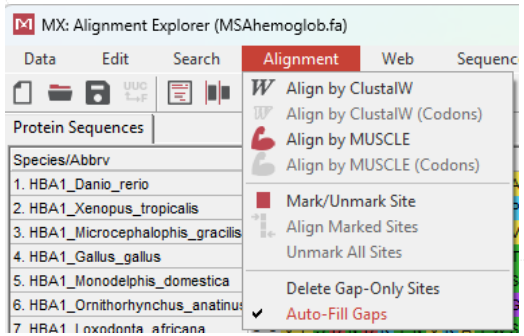
ISAGREED

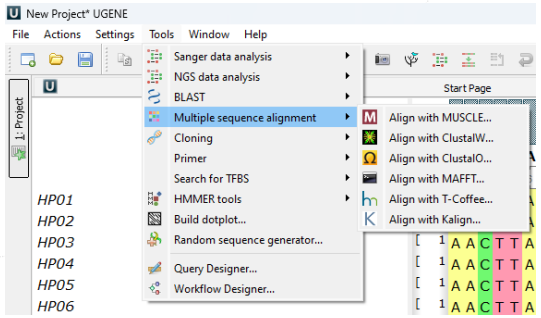
Co-funded by the European Union

# Multiple Sequence Alignment Tools - MSA

- The haplotypes identified in the two regions were characterized by means of multiple sequence alignment using the **Kalign** www.ebi.ac.uk/jdispatcher/msa/kalign
- **Other suitable MSA tools suitable for medium-large alignments :**
  - **Clustal Omega** (that uses seeded guide trees and HMM profile-profile techniques to generate alignments)
  - **MAFT** (that uses Fast Fourier Transforms).
  - www.ebi.ac.uk/jdispatcher/msa
- **Programs MEGA, Unipro UGENE**

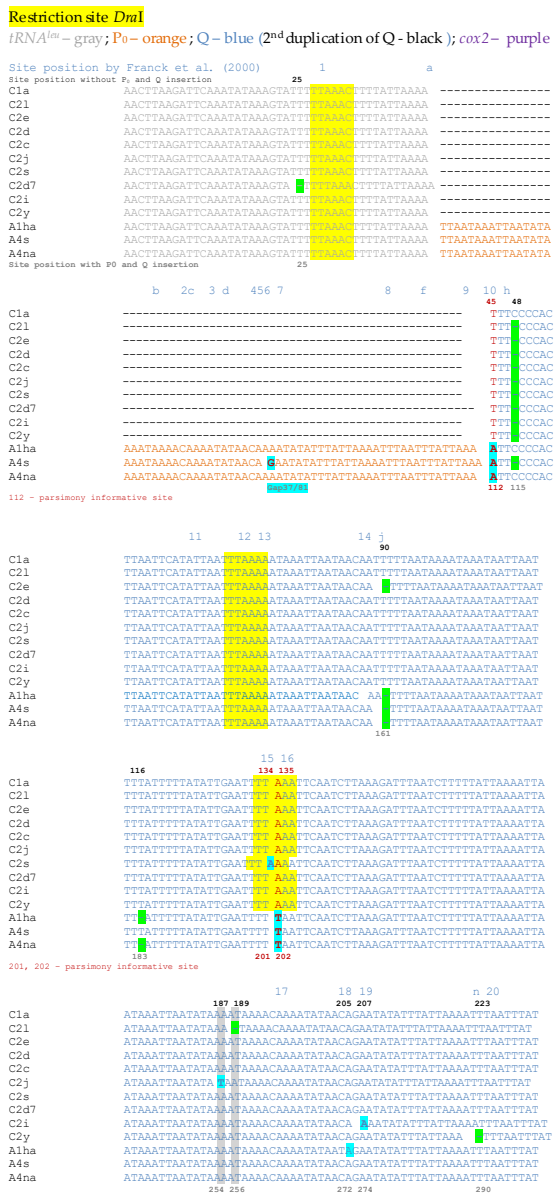







Co-funded by the European Union






# Multiple sequencing alignment by **Kalign** and **manual processing**

*tRNA$^{leu}$-cox2*

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

Co-funded by the European Union

- For detection DNA polymorphisms and haplotypes in both mtDNA regions using all sequences from a multiple sequence alignment (in FASTA format), **DnaSP** was used.

DnaSP DNA Sequence Polymorphism

- The nucleotide substitutions and insertions/deletions for each haplotype were determined by comparison of positions with primer-free sequences and the NC_001566 genome reference (www.ncbi.nlm.nih.gov/nuccore/NC_001566 )

- To identify specific *tRNA$^{leu}$-cox2* haplotypes of C and A lineages, reference sequences with 100% identity were searched (using **BLAST**, blast.ncibi.nlm.nih.gov/Blast.cgi) in the National Center for Biotechnology Information (NCBI) nucleotide database (GenBank).
- BLAST was used to verify the *cox1* haplotypes. The *cox1* sequence was also verified using the **BOLD database** (boldsystems.org/index.php/IDS_IdentificationRequest).

ISAGREED

ISAGREED

Co-funded by the European Union

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

# Frequencies of detected *tRNA$^{leu}$-cox2* haplotypes

| L | G | Haplotype | NCBI accession | DraI spectrum (bp) | Seq. length (bp) | NCBI reference | |
|---|---|---|---|---|---|---|---|
| | | | | | | accession number (% similarity) | designated as |
| C | Q | C1a | PP350804 | 47, 41, 64, 420 | 572 | JQ977699.1 (100) | carnica |
| C | Q | C2l | PP350810 | 47, 40, 64, 419 | 570 | OR761864.1 (100) | carnica |
| C | Q | C2e | PP350812 | 47, 40, 63, 420 | 570 | JQ977702.1 (100) | carnica |
| C | Q | C2d | PP350805 | 47, 40, 64, 420 | 571 | JF723977.1 (100) | carnica |
| C | Q | C2c | PP350806 | 47, 40, 64, 420 | 571 | JF723976.1 (100) | carnica |
| C | Q | C2j | PP350807 | 47, 40, 64, 420 | 571 | KX463941.1 (100) | carnica |
| C | Q | C2s | PP350808 | 47, 40, 63, 421 | 571 | JF723979.1 (100) | carnica |
| C | Q | C2d7* | PP397042 | 46, 40, 64, 420 | 570 | JF723977.1 (99.81) | carnica |
| C | Q | C2i | PP350809 | 47, 40, 64, 420 | 571 | JQ977703.1 (100) | carnica |
| C | Q | C2y | PP350811 | 47, 40, 64, 420 | 571 | JQ754650.1 (100) | carnica |
| A | $P_0Q$ | A1ha* | PP430326 | 47, 108, 482 | 637 | KX463739.1 (99.83) | iberiensis |
| A | $P_0QQ$ | A4s | PP430327 | 47, 107, 192, 483 | 829 | MW939597.1 (100) | n.a. |
| A | $P_0QQ$ | A4na* | PP430328 | 47, 108, 190, 482 | 827 | KX463793.1 (99.75) | iberiensis |



Co-funded by the European Union

# C haplotypes of *tRNA$^{leu}$-cox2* intergenic region

| Pozice/ Haplotyp | T/del 29 | C/del 50:1 (inz) | T/del 92:1 (inz) | T/A 132 | A/T 185 | A/del 187 | G/A 225 | A/del 221 | T/C 250 | C/T 385 |
|---|---|---|---|---|---|---|---|---|---|---|
| C1a | T | C | T | T | A | A | G | A | T | C |
| C2l | T | del | T | T | A | del | G | A | T | C |
| C2e | T | del | del | T | A | A | G | A | C | T |
| C2d | T | del | T | T | A | A | G | A | C | T |
| C2c | T | del | T | T | A | A | G | A | T | C |
| C2j | T | del | T | T | T | A | G | A | C | T |
| C2s | T | del | T | A | A | A | G | A | C | T |
| C2d7* | del | del | T | T | A | A | G | A | C | T |
| C2i | T | del | T | T | A | A | A | A | C | T |
| C2y | T | del | T | T | A | A | G | del | T | C |

ISAGREED

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

Co-funded by the European Union

# Haplotypes in sequence for barcoding (in *cox1*)

| Pozice/ Haplotyp | G/A 1 | G/A 52 | A/G 99 | C/T 142 | G/A 196 | T/C 259 | T/A 271 | C/A 274 | C/A 421 | A/G 448 | T/C 571 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HpB_1 | G | G | A | C | G | T | T | C | C | A | T |
| HpB_2 | G | G | G | C | G | T | T | A | C | A | T |
| HpB_3 | G | G | G | C | G | T | T | C | C | A | T |
| HpB_4 | G | G | G | C | G | T | T | C | C | G | T |
| HpB_5 | G | G | A | C | G | T | T | C | A | A | C |
| HpB_6 | G | G | G | T | G | T | T | C | C | A | T |
| HpB_7 | A | G | A | C | G | T | T | C | C | A | T |
| HpB_8 | G | A | A | C | G | T | T | C | C | A | T |
| HpB_9 | G | G | A | C | G | C | T | C | C | G | T |
| HpB_10 | G | G | G | C | A | T | T | C | C | A | T |
| HpB_11 | G | G | G | C | G | T | A | C | C | A | T |
| HpB_12 | A | G | G | C | G | T | T | C | C | G | T |
| HpB_13 | G | G | A | C | A | T | T | C | C | A | T |



Co-funded by the European Union

# Frequencies of detected haplotypes in A. m. in Czech Republic

*tRNA$^{leu}$-cox2*

| Haplotype | N | Frequency |
|---|---|---|
| C1a | 155 | 0.5032 |
| C2l | 25 | 0.0812 |
| C2e | 58 | 0.1883 |
| C2d | 12 | 0.0390 |
| C2c | 47 | 0.1526 |
| C2j | 1 | 0.0032 |
| C2s | 4 | 0.0130 |
| C2d7 | 1 | 0.0032 |
| C2i | 1 | 0.0032 |
| C2y | 1 | 0.0032 |
| A1ha | 1 | 0.0032 |
| A4s | 1 | 0.0032 |
| A4na | 1 | 0.0032 |
| Total | 308 | 1.0000 |

*cox1*

| Haplotype | N | Frequency |
|---|---|---|
| HpB01 | 64 | 0.2078 |
| HpB02 | 134 | 0.4351 |
| HpB03 | 65 | 0.2110 |
| HpB04 | 24 | 0.0779 |
| HpB05 | 4 | 0.0130 |
| HpB06 | 4 | 0.0130 |
| HpB07 | 4 | 0.0130 |
| HpB08 | 1 | 0.0032 |
| HpB09 | 1 | 0.0032 |
| HpB10 | 1 | 0.0032 |
| HpB11 | 1 | 0.0032 |
| HpB12 | 1 | 0.0032 |
| HpB13 | 1 | 0.0032 |
| HpB14 | 1 | 0.0032 |
| HpB15 | 1 | 0.0032 |
| HpB16 | 1 | 0.0032 |
| Total | 308 | 1.0000 |

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

Co-funded by the European Union

# Haplotype diversity parameters of *tRNA$^{leu}$-cox2* and *cox1*

Haplotype diversity Hd
Molecular diversity π
Tajima´s D

| Haplotype | N | H | Hd | π | D |
|---|---|---|---|---|---|
| tRNA$^{leu}$-cox2 (796 bp) | 308 | 13 | 0.68186 | 0.00172 (1.38726-06) | -3.02386 (0.00249) |
| cox1 (658 bp) | 308 | 16 | 0.71866 | 0.00203 (1.96124e-06) | -1.60658 (0.10815) |

Genetic diversity indices, including nucleotide diversity π, haplotype diversity Hd, and Tajima's D for testing the neutral mutation hypothesis, were assessed using **pegas package**, in the R program.
It can be used also **DnaSP**, **MEGA, Arlequin**



Co-funded by the European Union

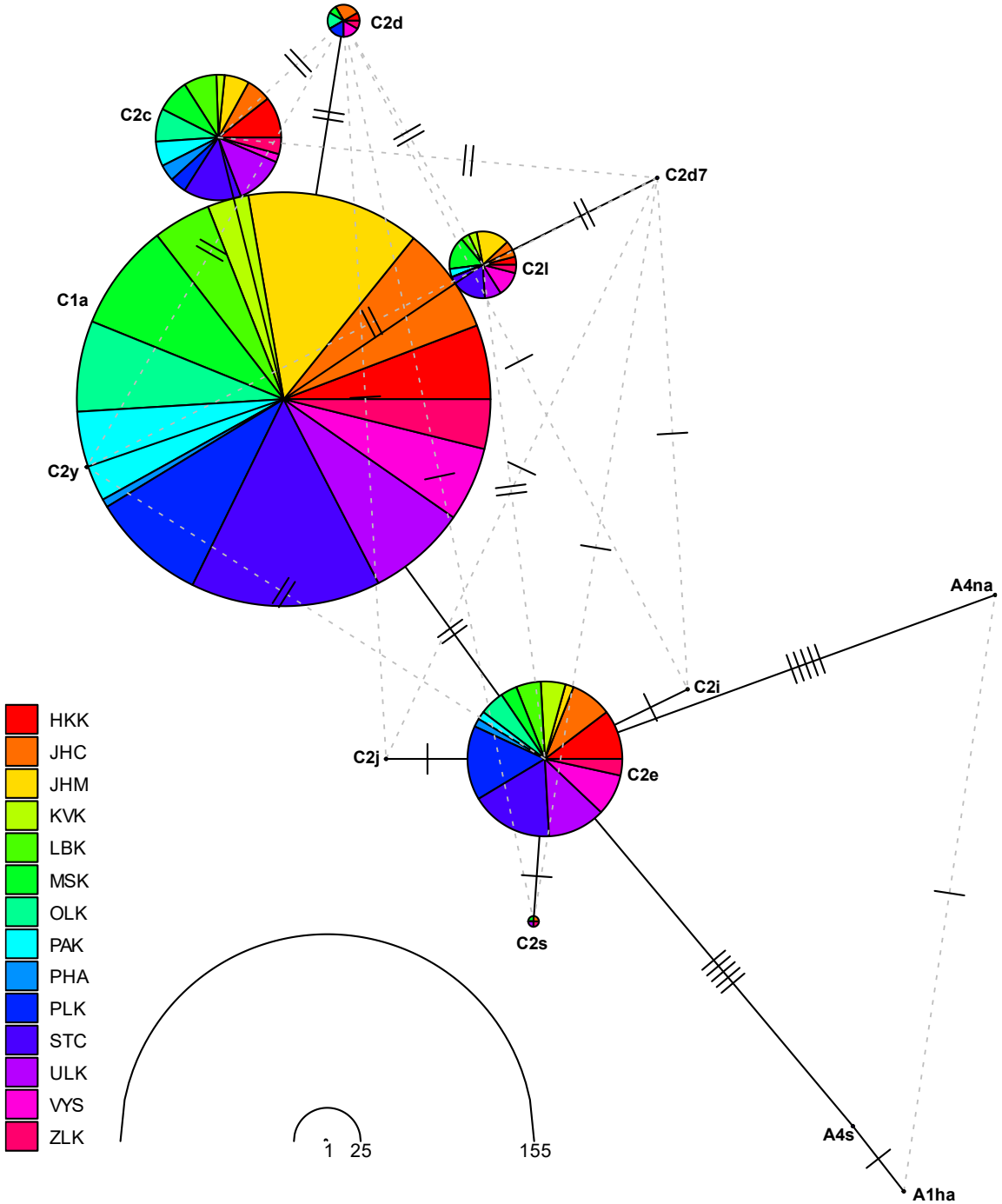
# Haplotype networks

- RMST (the randomized minimum spanning tree) haplotype networks illustrating the frequencies and relationships of the *tRNA$^{leu}$-cox2* and *cox1* haplotypes found in the Czech Republic
- The phylogeographic networks were generated using the RMST method with 1000 iterations by rmst function in the **pegas package** in R.
- It can be use also:
  - PopART (Leigh, Bryant, 2015)

ISAGREED

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

ISAGREED
Co-funded by the European Union

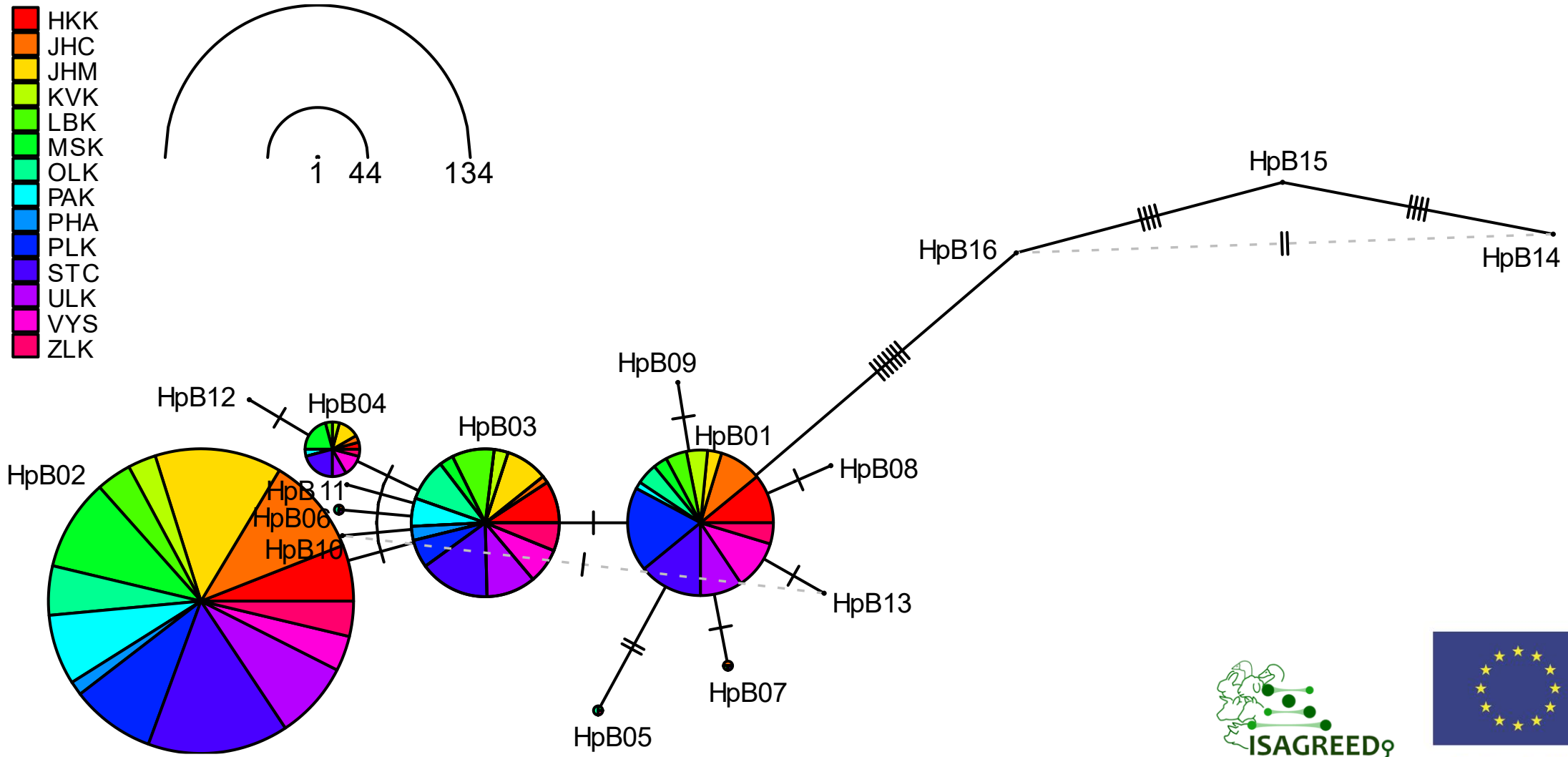



- RMST haplotype network analysis among studied 308 *tRNA$^{leu}$-cox2* sequencies
- The size of the circles is proportional to the number of the individuals and the different colors represent different regions.

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

Co-funded by the European Union

- RMST haplotype network analysis among studied 308 *cox1* sequencies






Co-funded by the European Union

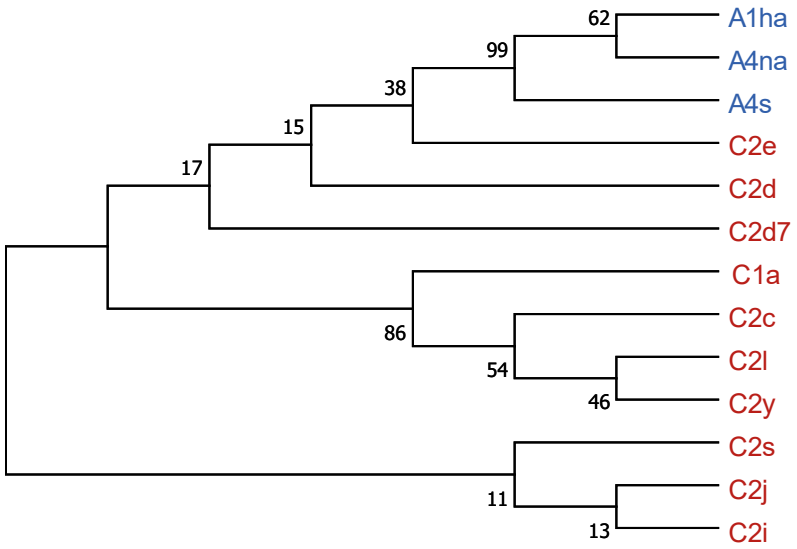
# Phylogenetic analysis based on sequencies

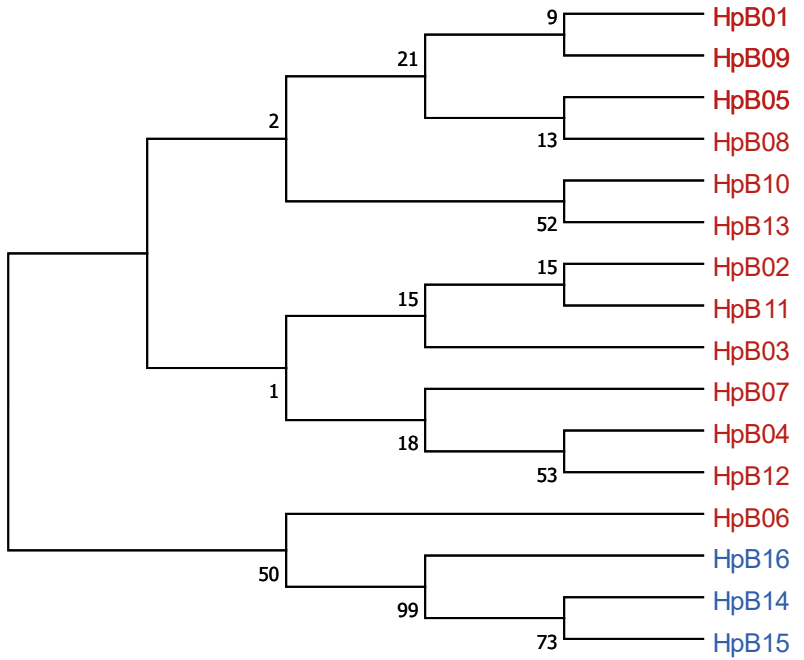
1. MSA
2. Phylogenetic analysis
   - Construct phylogenetic trees to determine the evolutionary relationships among *A. melifera* haplotypes (*tRNA$^{leu}$-cox2* and *cox1*) was inferred by using Maximum Likelihood method and Tamura-Nei model in program **Mega X**.
   - The bootstrap consensus tree was derived from 10 000 replicates. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. The percentage of replicate trees in which the associated haplotypes clustered together in the bootstrap test (10 000 replicates) are shown next to the branches.
   - Initial tree for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with the highest-level log likelihood value.



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

ISAGREED
Co-funded by the European Union




C lineage are in red and A lineage in blue.

- Bootstrap consensus tree of ***tRNA$^{leu}$-cox2*** haplotypes in the Czech Republic. The analysis included 13 nucleotide sequences. A total of 796 positions were included in the final dataset.

- The bootstrap consensus tree ***cox1*** haplotypes in the Czech Republic. The analysis included 16 nucleotide sequences. A total of 658 positions were included in the final dataset.

**Bootstrap consensus tree** by using the Maximum Likelihood method and Tamura-Nei model. Beside the branches is the percentage of replicates where the corresponding haplotypes were found to cluster (the bootstrap test - 10 000 replicates)

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

Co-funded by the European Union

# Mikrosatelites

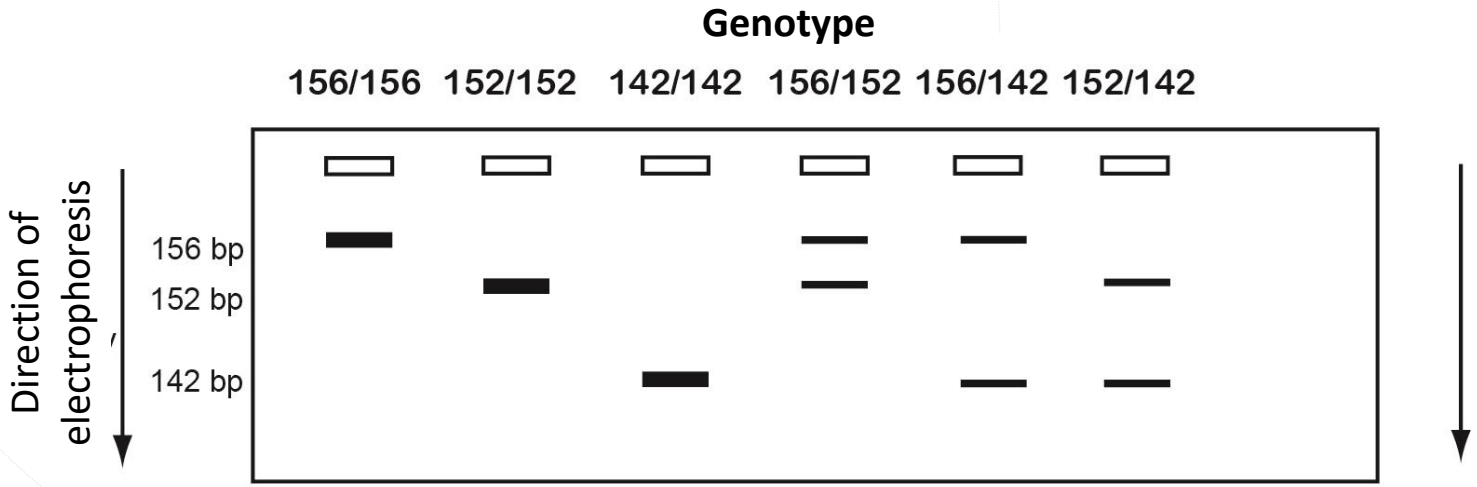
$(GC)_n$
size of PCR

↓ primer direct — mikrosatellite — reverse primer ↓

5´CATTGAATCGGTATCAT$(N)_{60}$AGCGCGCGCGCGCGCGCT$(N)_{40}$ACGTTAATTCATTCGTG 3´ — n=8, 152 bp

5´CATTGAATCGGTATCAT$(N)_{60}$AGCGCGCT$(N)_{40}$ACGTTAATTCATTCGTG 3´ — n=3, 142 bp

5´CATTGAATCGGTATCAT$(N)_{60}$AGCGCGCGCGCGCGCGCGCGCT$(N)_{40}$ACGTTAATTCATTCGTG 3´ — n=10, 156 bp

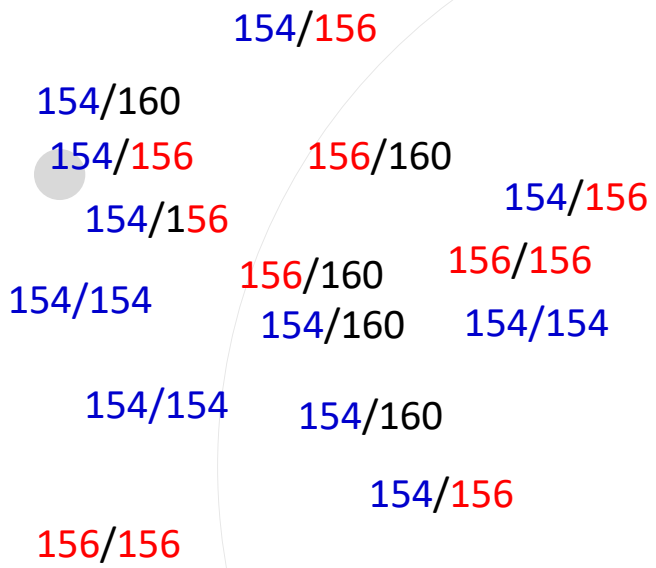



Different combinations of genotypes with the 3 microsatellite alleles (3 homozygous, 3 heterozygous).

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

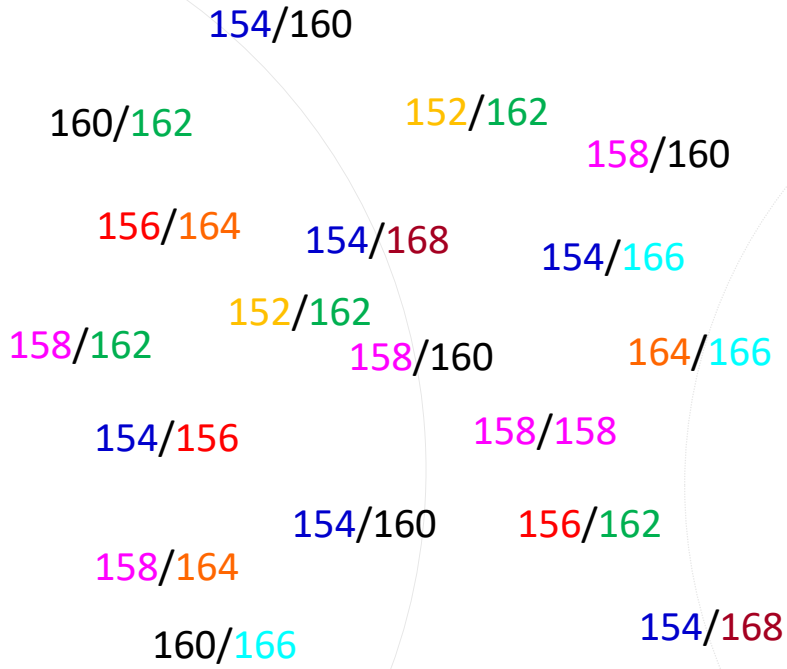
Co-funded by the European Union

# Example of MS variability in two populations




low variability

(3 alleles, low heterozygosity)




highly variable

(9 alleles, high heterozygosity)



Co-funded by the European Union

ISAGREED

- Microsatellites remain an affordable genetic marker with the capacity to capture multilocus genotype information, which can then be employed to estimate genetic diversity, population structure, with application in conservation genetics and breeding of a wide range of organisms

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

Co-funded by the European Union

# Fragmentation analysis

Capillary electrophoresis + fluorescence detection (for accurate determination of fragment size - alleles)

Fragments are separated according to their size in the capillary. The sensor detects the passage of the molecule - its colour and signal intensity over time - smaller fragments reach the sensor earlier than larger ones. A size standard goes with the samples, the software draws a calibration curve and uses this to determine the size of the fragments to be tested.

Instrument: **genetic analyser**

Fragment analysis was performed using ABI PRISM 3500 Genetic Analyser (Applied Biosystems) based on standard conditions (POP-7 polymer, G5 matrix). Fragment size was accurately determined with GeneScan™ 600 LIZ™ Size Standard and evaluated using GeneMapper v 6.0 software (Applied Biosystems).

ISAGREED

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

Co-funded by the European Union

# Microsatellite loci and multiplex primer group

| Locus | Primer sequence | Length (nt) | Range (bp) | Fluorescent label | Motif | Author |
|---|---|---|---|---|---|---|
| A079 | CGAAGGTTGCGGAGTCCTC<br>GTCGTCGGACCGATGCG | 19<br>17 | 89-124 | 6-FAM | $(CT)_{14}$ | Estoup et al. 1994 |
| Ac306 | GAATATGCCGCTGCCACC<br>TTTCGTTGCATCCGAGCG | 18<br>18 | 161-187 | 6-FAM | $(CT)_{11}$ | Solignac et al., 2003 |
| Ap226 | AACGGTGTTCGCGAAACG<br>AGCCAACTCGTGCGGTCA | 18<br>18 | 228-248 | 6-FAM | $(CT)_8$ | Solignac et al., 2003 |
| A007 | CCCTTCCTCTTTCATCTTCC<br>GTTAGTGCCCTCCTCTTGC | 20<br>19 | 94-136 | VIC | $(CT)_3(T)_7CTTCG(CT)_{24}$ | Estoup et al. 1994 |
| HB-C16-01 | AAAATGCGATTCTAATCTGG<br>TTGCCTAAAATGCTTGCTAT | 20<br>20 | 249-309 | VIC | $(GA)_{35}$ | Shaibi et al.2008 |
| Ap068 | TGTCTGCCCTCCTCTCTGTT<br>CACATCGAGCGAGAAGGC | 20<br>18 | 147-169 | NED | $(CT)_{12}(TA)_8$ | Solignac et al., 2003 |
| A014 | GTGTCGCAATCGACGTAACC<br>GTCGATTACCGATCGTGACG | 20<br>20 | 214-247 | NED | $(CT)_{13}(GGT)_9$ | Estoup et al. 1994 |
| Ap223 | TCGTACAACGTCGCGCAA<br>GCCGCTCGCCTGTATCTG | 18<br>18 | 147-182 | PET | $(T)_5(C)_4A(T)6(C_5)$ | Solignac et al., 2003 |

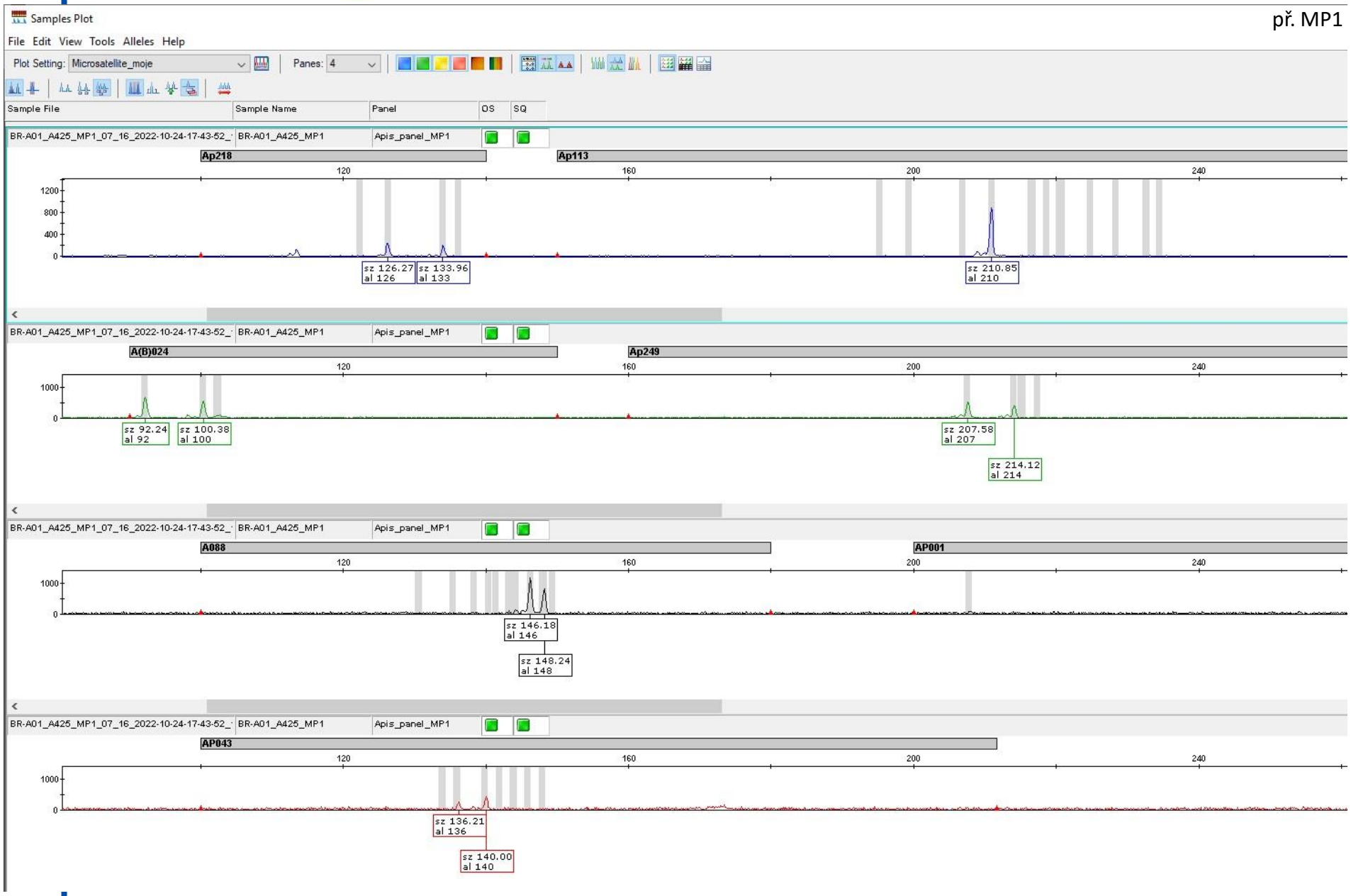
The cycling conditions were as follows: 95° C (2 min); 30 cycles of a 20 s denaturation at 95 ° C, a 20 s annealing at 57 ° C, a 30 s elongation at 72 ° C; and a final extension step at 72 ° C for 60 min.



Co-funded by the European Union

# Result of microsatellites

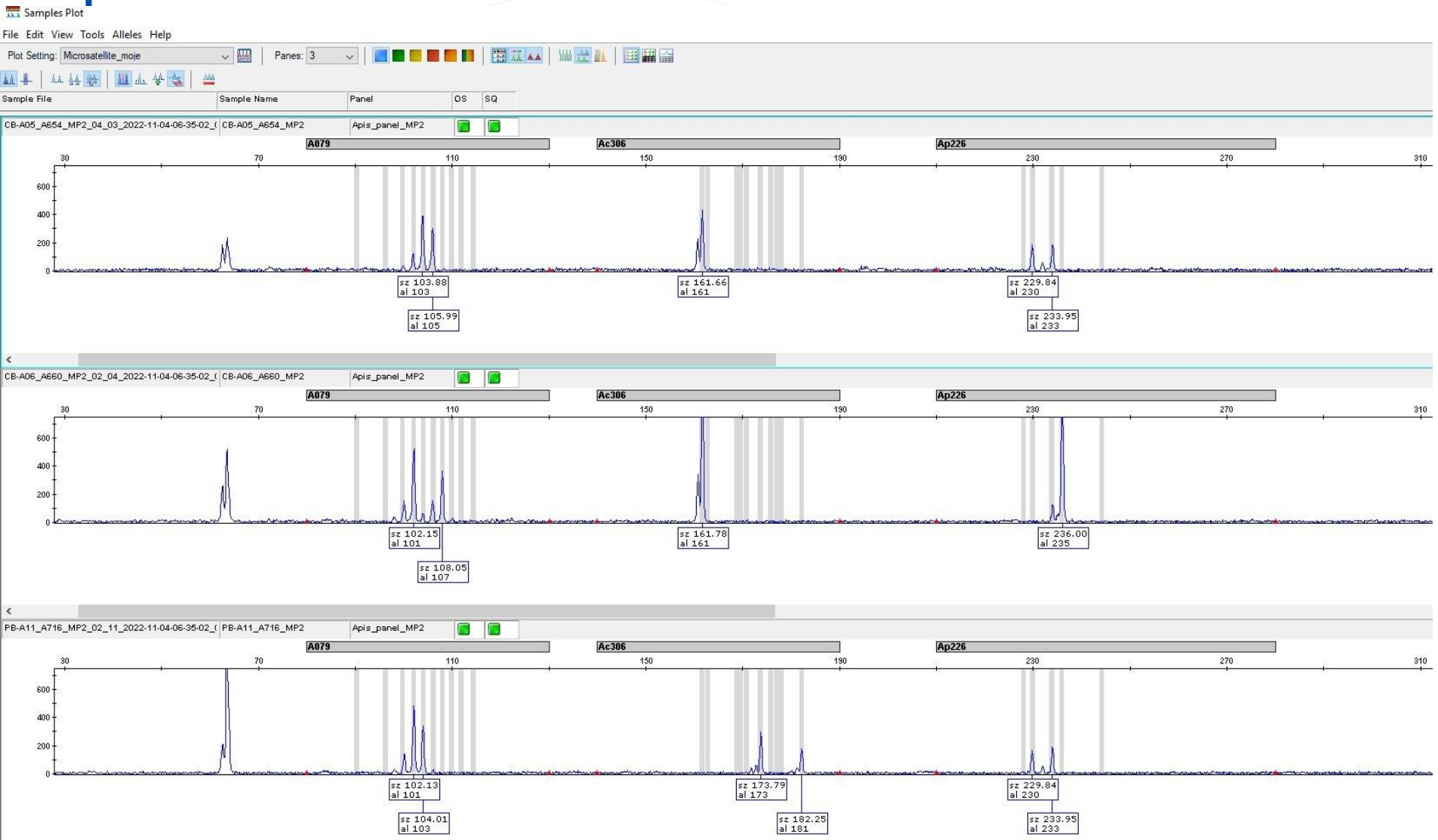
př. MP1






ISAGREED
Co-funded by the European Union

# Variability - microsatellites




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

ISAGREED
Co-funded by the European Union

| Locus | N | Na | Ne | I | Ho | He | uHe | F | HWE |
|---|---|---|---|---|---|---|---|---|---|
| A(B)024 | 545 | 4 | 2.605 | 1.018 | 0.591 | 0.616 | 0.617 | 0.041 | ns |
| A088 | 546 | 10 | 2.296 | 1.174 | 0.551 | 0.564 | 0.565 | 0.023 | ns |
| Ap043 | 543 | 10 | 3.984 | 1.593 | 0.729 | 0.749 | 0.750 | 0.026 | ns |
| Ap113 | 550 | 16 | 1.770 | 0.934 | 0.438 | 0.435 | 0.435 | -0.007 | *** |
| Ap218 | 542 | 5 | 2.700 | 1.060 | 0.546 | 0.630 | 0.630 | 0.133 | *** |
| Ap249 | 550 | 8 | 1.847 | 0.832 | 0.467 | 0.459 | 0.459 | -0.019 | ns |
| A007 | 548 | 15 | 4.374 | 1.765 | 0.765 | 0.771 | 0.772 | 0.009 | ns |
| A014 | 548 | 22 | 2.023 | 1.298 | 0.500 | 0.506 | 0.506 | 0.011 | ns |
| A079 | 548 | 13 | 6.169 | 1.992 | 0.816 | 0.838 | 0.839 | 0.026 | ns |
| Ac306 | 548 | 11 | 1.975 | 1.076 | 0.484 | 0.494 | 0.494 | 0.020 | ns |
| Ap068 | 547 | 11 | 3.589 | 1.514 | 0.702 | 0.721 | 0.722 | 0.027 | ns |
| Ap223 | 549 | 5 | 1.744 | 0.796 | 0.439 | 0.427 | 0.427 | -0.029 | ns |
| Ap226 | 549 | 9 | 3.093 | 1.310 | 0.676 | 0.677 | 0.677 | 0.001 | *** |
| HB-C16-01 | 511 | 20 | 10.989 | 2.582 | 0.738 | 0.909 | 0.910 | 0.188 | *** |
| A(B)124 | 532 | 17 | 4.283 | 1.844 | 0.774 | 0.767 | 0.767 | -0.010 | ns |
| AP019 | 541 | 7 | 2.815 | 1.290 | 0.647 | 0.645 | 0.645 | -0.003 | ns |
| Ap273 | 551 | 4 | 1.466 | 0.624 | 0.318 | 0.318 | 0.318 | 0.001 | ns |
| Ap289 | 552 | 24 | 1.942 | 1.317 | 0.476 | 0.485 | 0.485 | 0.018 | ns |
| HB-C16-05 | 546 | 17 | 2.305 | 1.379 | 0.535 | 0.566 | 0.567 | 0.055 | ns |
| A043 | 541 | 8 | 2.250 | 1.162 | 0.505 | 0.556 | 0.556 | 0.092 | *** |
| Ap049 | 541 | 6 | 1.767 | 0.926 | 0.370 | 0.434 | 0.434 | 0.148 | *** |
| Ap288 | 541 | 4 | 1.218 | 0.382 | 0.174 | 0.179 | 0.179 | 0.031 | ns |
| **Mean** | 544.04 | **11.18** | **3.055** | **1.267** | **0.556** | **0.579** | **0.580** | **0.036** | |
| SE | 1.86 | 1.29 | 0.454 | 0.104 | 0.035 | 0.037 | 0.037 | 0.012 | |

- The calculations were performed in the **GenAlEx** version 6.5 environment.
- The following parameters were calculated: the number of alleles (*Na*), the effective number of alleles (*Ne*), the Shannon information index (*I*), the observed (*Ho*), expected (*He*) and unbiased expected heterozygosity (*uHe*), and Fixation index (*F*).
- Other program: **diveRsity** package in R



ISAGREED
Co-funded by the European Union

- Wright's F statistics ($F_{ST}$, $F_{IS}$ and $F_{IT}$) as proposed by Weir and Cockerham and the distribution of genetic diversity were analyzed using analysis of molecular variance (AMOVA).

| Locus | $F_{IS}$ | $F_{IT}$ | $F_{ST}$ |
|---|---|---|---|
| A(B)024 | -0.041 | 0.047 | 0.085 |
| A088 | -0.096 | 0.000 | 0.087 |
| AP043 | -0.067 | 0.021 | 0.083 |
| Ap113 | -0.092 | -0.005 | 0.080 |
| Ap218 | 0.059 | 0.150 | 0.097 |
| Ap249 | -0.100 | -0.011 | 0.081 |
| A007 | -0.067 | 0.017 | 0.079 |
| A014 | -0.085 | 0.004 | 0.082 |
| A079 | -0.061 | 0.030 | 0.086 |
| Ac306 | -0.060 | 0.024 | 0.079 |
| Ap068 | -0.063 | 0.028 | 0.086 |
| Ap223 | -0.124 | -0.031 | 0.082 |
| Ap226 | -0.110 | -0.008 | 0.092 |
| HB-C16-01 | 0.082 | 0.186 | 0.113 |
| A(B)124 | -0.108 | -0.018 | 0.080 |
| AP019 | -0.084 | 0.005 | 0.083 |
| Ap273 | -0.072 | -0.007 | 0.060 |
| Ap289 | -0.064 | 0.019 | 0.078 |
| HB-C16-05 | -0.030 | 0.058 | 0.085 |
| A043 | 0.018 | 0.110 | 0.093 |
| Ap049 | 0.039 | 0.155 | 0.121 |
| Ap288 | -0.066 | 0.015 | 0.076 |
| **Mean** | **-0.054** | **0.036** | **0.086** |
| SE | 0.012 | 0.013 | 0.003 |

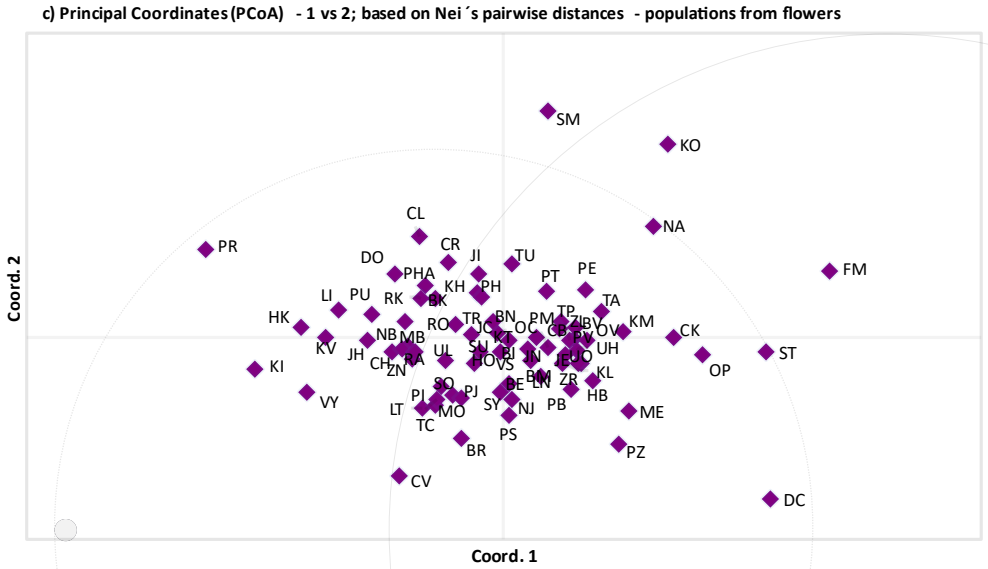
## GenAlEx

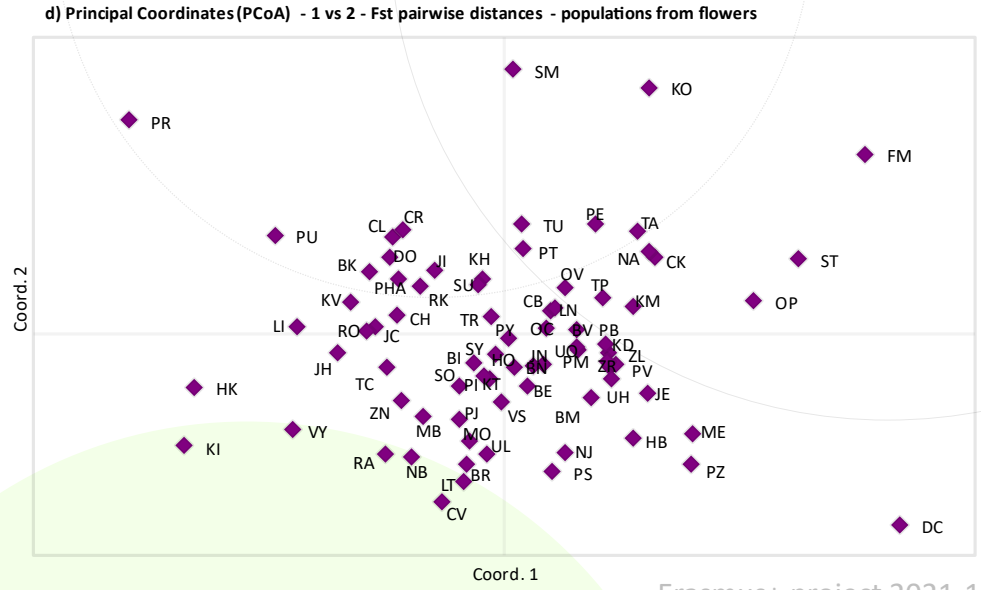
| Source | df | Sum of squares | Mean of squares | Estimated variance | Percentage of variation (%) |
|---|---|---|---|---|---|
| Between regional populations | 77 | 619.940 | 8.157 | 0.095 | **1** |
| Among Individuals | 476 | 3231.345 | 6.789 | 0.388 | **6** |
| Within Individuals | 553 | 3325.000 | 6.013 | 6.013 | **93** |
| Total | 1105 | 7176.285 | | 6.496 | **100** |

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

Co-funded by the European Union

- Pairwise Nei's unbiased and pairwise $F_{ST}$ genetic distances between populations were calculated and used for principal component analysis (PCA).



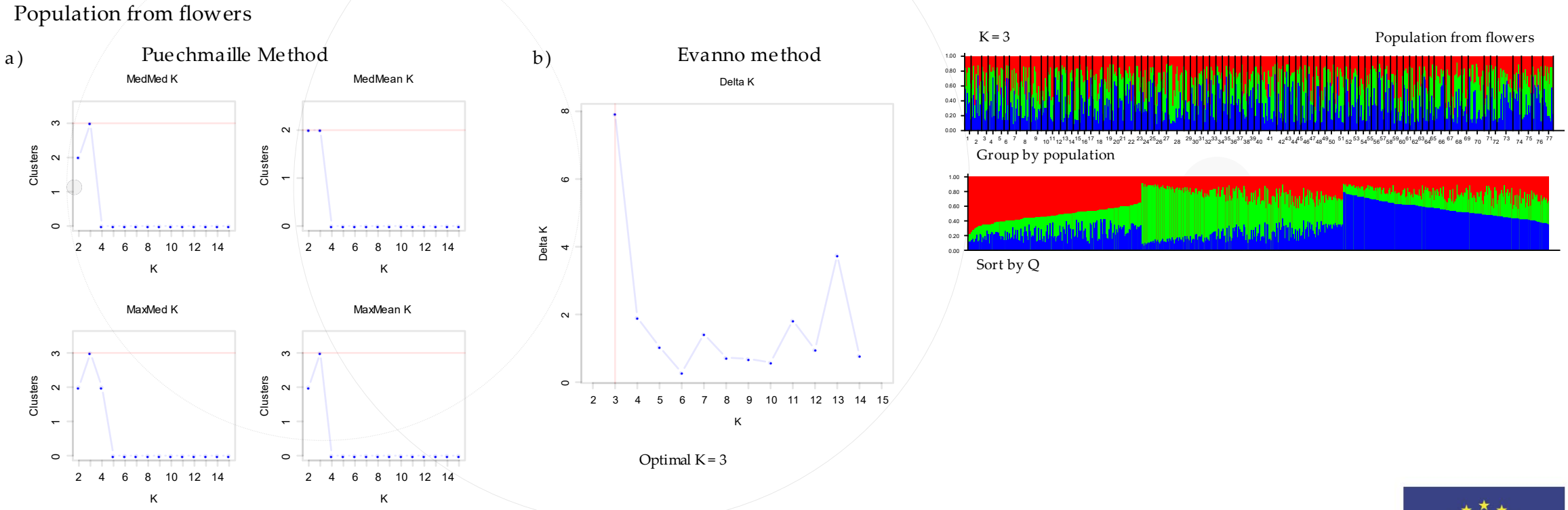




# GenAlEx

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

Co-funded by the European Union

- The Bayesian clustering method of **STRUCTURE ver. 2.3.4** was used to analyze the genetic diversity and degree of admixture of honey bee populations. Ten independent simulations were run, each including 10,000 burn-in steps followed by 100,000 Markov chain Monte Carlo (MCMC) iterations. Subsequently, we employed Clumpak and Structure Selector, which implement Evanno method and Puechmaille method to ascertain the **optimal number of clusters** (**K**) that best fit the data (**ΔK**, **MedMeaK, MaxMeaK, MedMedK, and MaxMedK**).

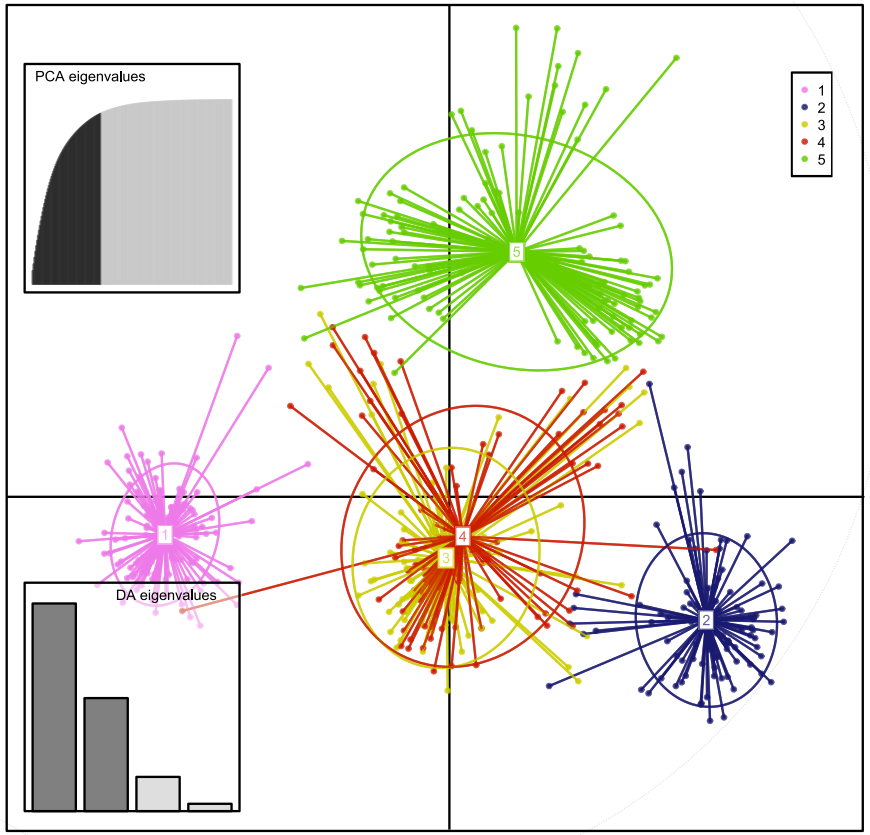


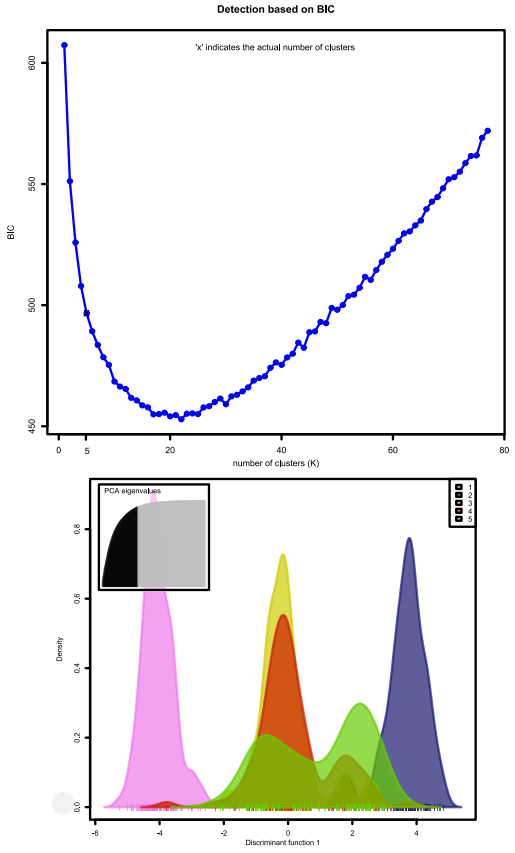

Optimal K after removing spurious clusters (indicated gy red lines) are for MedMeaK, MaxMeaK, MedMedK, MaxMedK (Threshold=0.5) = 3



Co-funded by the European Union

- To determine genetic structure and infer genetic admixture, a **discriminant analysis of principal components (DAPC)** was conducted using the **adegenet R package** (within the R statistical computing environment)






ISAGREED
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.*

Tomáš Urban

urban@mendelu.cz



Co-funded by the European Union