ISAGREED

# Methodological approaches to estimate effective population size

**Modul no.3: Animal Breeding**

Luboš Vostrý

Česká zemědělská univerzita v Praze, Fakulta agrobiologie, potravinových a přírodních zdrojů

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

Co-funded by the European Union

ISAGREED

# Effective population size

ISAGREED

Co-funded by
the European Union

# The effective population size Ne

## The most important parameter in conservation genetics

**The effective population size (Ne)** of a **real population** X is the size of a **hypothetical ideal population** (Wright-Fisher) that will result in the **- same amount of genetic drift -** as in the **real (actual) population considered**.

By definition $N_{eLD}$, $N_{eV}$, $N_{eI}$ and $N_{eE}$ are not the always equal (same) !

Co-funded by the European Union

# Effective population size (Ne)

An ideal Wright-Fisher population (Fisher 1930; Wright 1931) is;

- discrete generations,
- diploid individuals,
- sexual reproduction,
- the population size is constant across generations,
- no migration,
- mating is random,
- no mutations,
- the sex ratio is 1:1,
- no selection,
- the average number of recruits produced by each individual is Poisson distributed with a mean and variance of two

Co-funded by the European Union

# In an ideal Wright-Fisher population on a timescale;

Infinite population size;

- No change in allele and genotype frequencies
- No change in inbreeding level
- No "random extinction" of alleles
- Alleles of non-linked genes are in gametic phase/linkage equilibrium

Finite population size –> genetic drift;

- Change in allele frequencies
- Change in inbreeding level
- Lost of genetic variation (fixation / loss of alleles)
- Change in gametic phase/linkage disequilibrium

ISAGREED
Co-funded by the European Union

# The effective population size Ne

## The most important parameter in conservation genetics

**The effective population size (Ne)** of a **real population** X is the size of a **hypothetical ideal population** (Wright-Fisher) that will result in the **- same amount of genetic drift -** as in the **real (actual) population considered**.

**Effective variance population size ($N_{eV}$):**

- same change in allele frequencies -

**Effective inbreeding population size ($N_{eI}$):**

- same change in inbreeding level -

**Effective eigenvalue population size ($N_{eE}$):**

- same long term rate at which genetic variants are lost -

**Effective linkage disequilibrium population size ($N_{eLD}$):**

- same change in gametic phase/linkage disequilibrium -

By definition $N_{eLD}$, $N_{eV}$, $N_{eI}$ and $N_{eE}$ are not the always equal (same) !

Co-funded by the European Union

# Effective linkage disequilibrium population size (NeLD)

The linkage disequilibrium effective population size $N_{eLD}$ of a real population X with an observed LD for a given interval length is the size of hypothetic ideal population that in an equilibrium state would display the same pattern of LD for the same interval length as observed in real finite population (H. Simianer, 2012; LCG)

Note that by definition $N_{eLD}$ is not the same as those defined for $N_{eV}$, $N_{eI}$ and $N_{eE}$ !

Co-funded by the European Union

# NeLD

In an ideal population of infinite size that has reached an equilibrium state, all loci are in linkage equilibrium.

⇕

In an ideal population of finite size that has reached an equilibrium state, loci are in linkage disequilibrium while the amount of LD is a function of the genetic distance of the considered loci and the size of population.

ISAGREED
Co-funded by
the European Union

# Widely used in calculation of the effective population size

| Author | Software | Ne type | Information |
|---|---|---|---|
| *Gutiérrez et al., 2008 GSE, 2009 JABG* | *Endog* | $N_{el}$ | *Pedigree* |
| *Barbato et al., 2015 FG* | *SNeP* | $N_{eLD}$ | *Genomic* |
| *Do et al., 2014 MER* | *NeEstimator V2* | $N_{eGD}$ | *Genomic* |
| *Santiago et al., 2020 MBE* | *GONE* | $N_{eGLD}$ | *Genomic* |
| *Santiago et al., 2020 MBE* | *CurrentNe* | $N_{eGD}$ | *Genomic* |
| *Browning et al., 2018 PLOS G.* | *IBDNe* | $N_{eIBD}$ | *Genomic* |
| *Fournier eta al., 2023 NC* | *HapNe* | $N_{eIBD}$ | *Genomic* |

SNeP estimated $N_e$LD in the current generation estimated by linear regression based on historical estimates as applied in Kukučkova et al., 2017

# Empirical population with informative pedigree

## The Old Kladruber horse (OKH)

### Informative pedigree

# Individuals in pedigree: 9288 (1729 – 2018)

Tracing up to 55 generations

The **Kladruber** (Czech *Starokladrubský kůň*) is the oldest Czech horse breed and one of the world's oldest horse breeds, bred for more than 400 years.

## Calculation based data set

### Genotyped animals (n = 215; 1994 - 2014)

Q1: **38996** autosomal SNPs

Q2: **60645** autosomal SNPs

# Individuals in pedigree: 3120 (1729 – 2014)

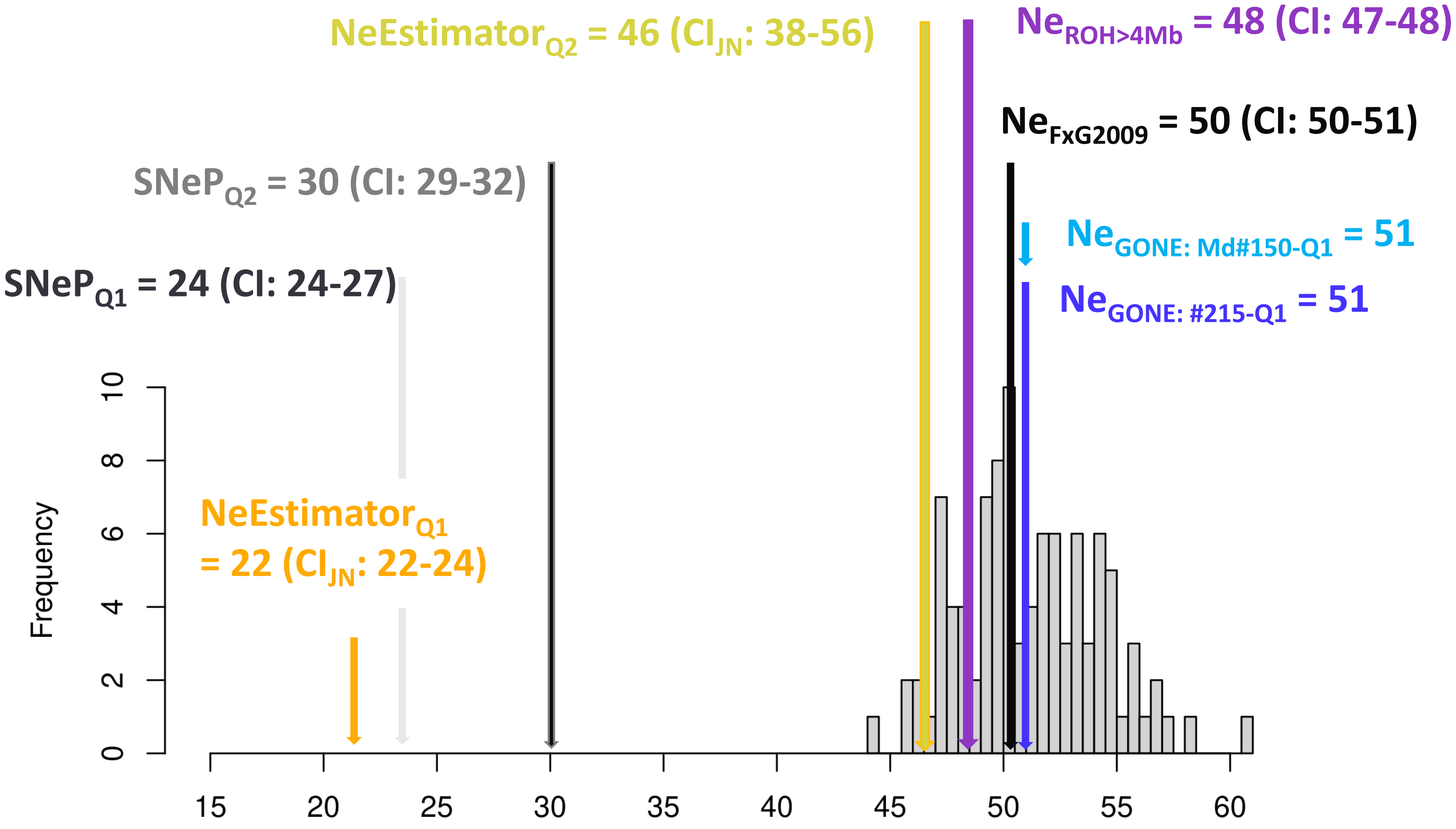
Tracing up to 49 generations

ECG: 15.9 (13.5 – 17.4)

ISAGREED
Co-funded by the European Union

# Results of the current effective population size: OKH

ISAGREED



NeEstimator$_{Q2}$ = 46 (CI$_{JN}$: 38-56)

Ne$_{ROH>4Mb}$ = 48 (CI: 47-48)

Ne$_{FxG2009}$ = 50 (CI: 50-51)

SNeP$_{Q2}$ = 30 (CI: 29-32)

Ne$_{GONE: Md\#150-Q1}$ = 51

SNeP$_{Q1}$ = 24 (CI: 24-27)

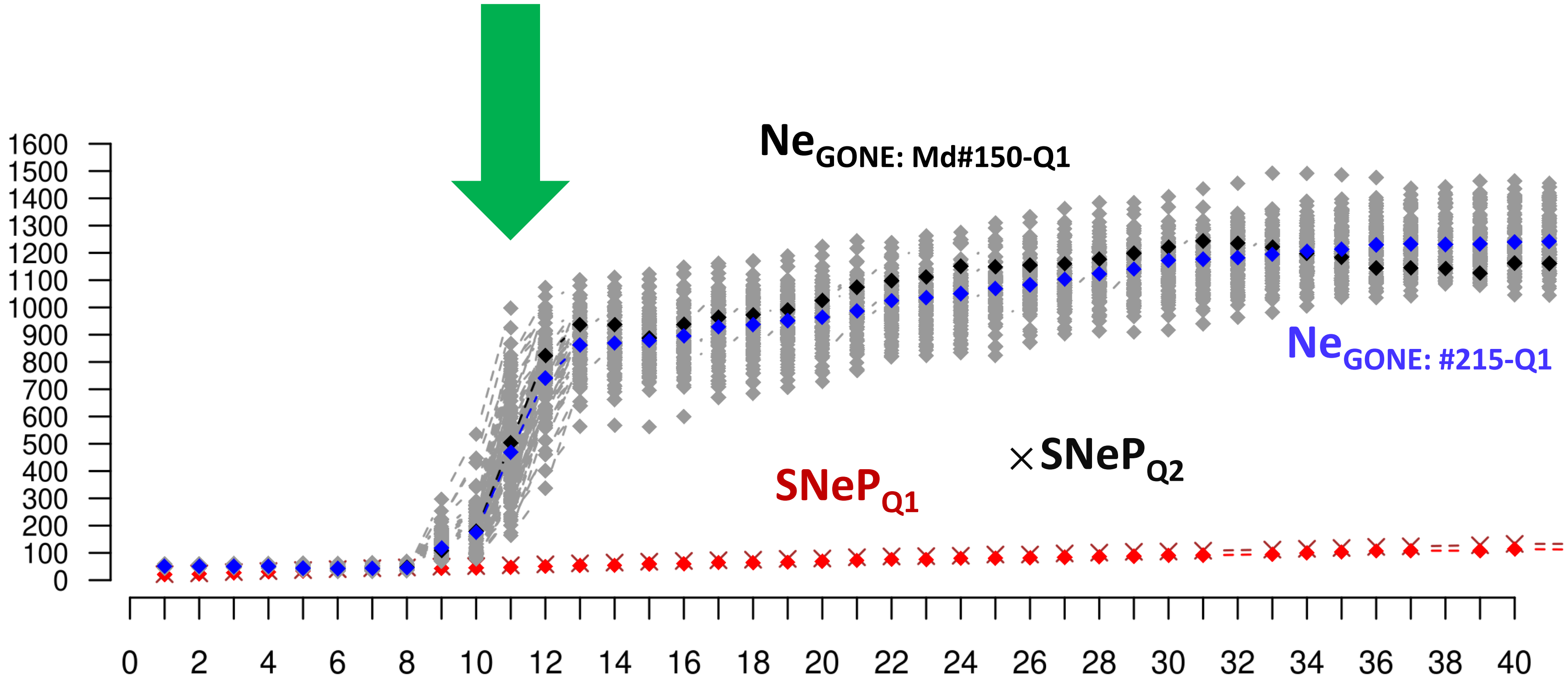
Ne$_{GONE: \#215-Q1}$ = 51

NeEstimator$_{Q1}$ = 22 (CI$_{JN}$: 22-24)

Co-funded by the European Union

# Results of the historical effective population size: OKH

≈ 100 years back / fall of Austro-Hungarian Monarchy / drop in the number of horses



# Conclusions

## Highly recommended

**Software GONE**

### Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium

Enrique Santiago,[*,1] Irene Novo,[2] Antonio F. Pardiñas,[3] María Saura,[4] Jinliang Wang,[5] and Armando Caballero[2]

*Mol. Biol. Evol.* 37(12):3642–3653 doi:10.1093/molbev/msaa169

Co-funded by
the European Union

ISAGREED

# Estimation of effective populations size with software GONE

## - First run

Co-funded by the European Union

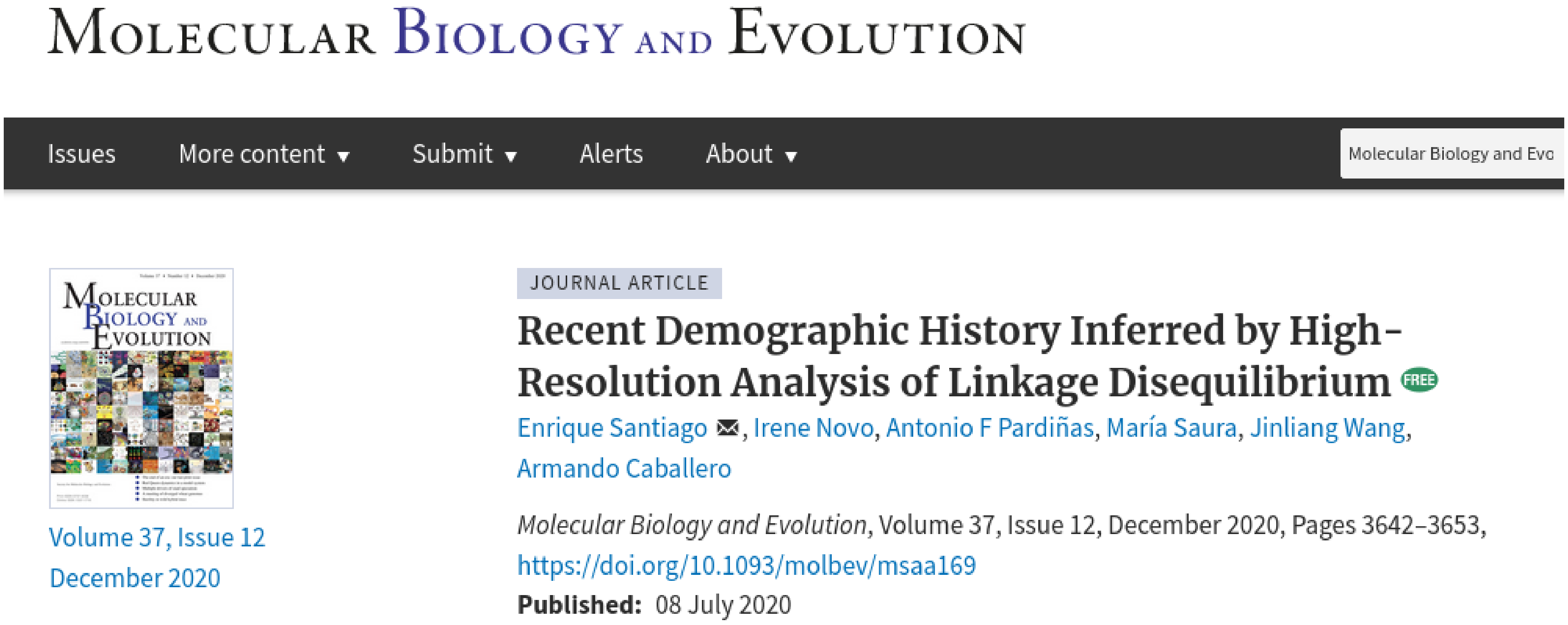
# Download

- https://github.com/esrud/GONE

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

Co-funded by the European Union

# Methodological approach



MOLECULAR BIOLOGY AND EVOLUTION

Volume 37, Issue 12
December 2020

JOURNAL ARTICLE

**Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium** FREE

Enrique Santiago ✉, Irene Novo, Antonio F Pardiñas, María Saura, Jinliang Wang, Armando Caballero

*Molecular Biology and Evolution*, Volume 37, Issue 12, December 2020, Pages 3642–3653, https://doi.org/10.1093/molbev/msaa169

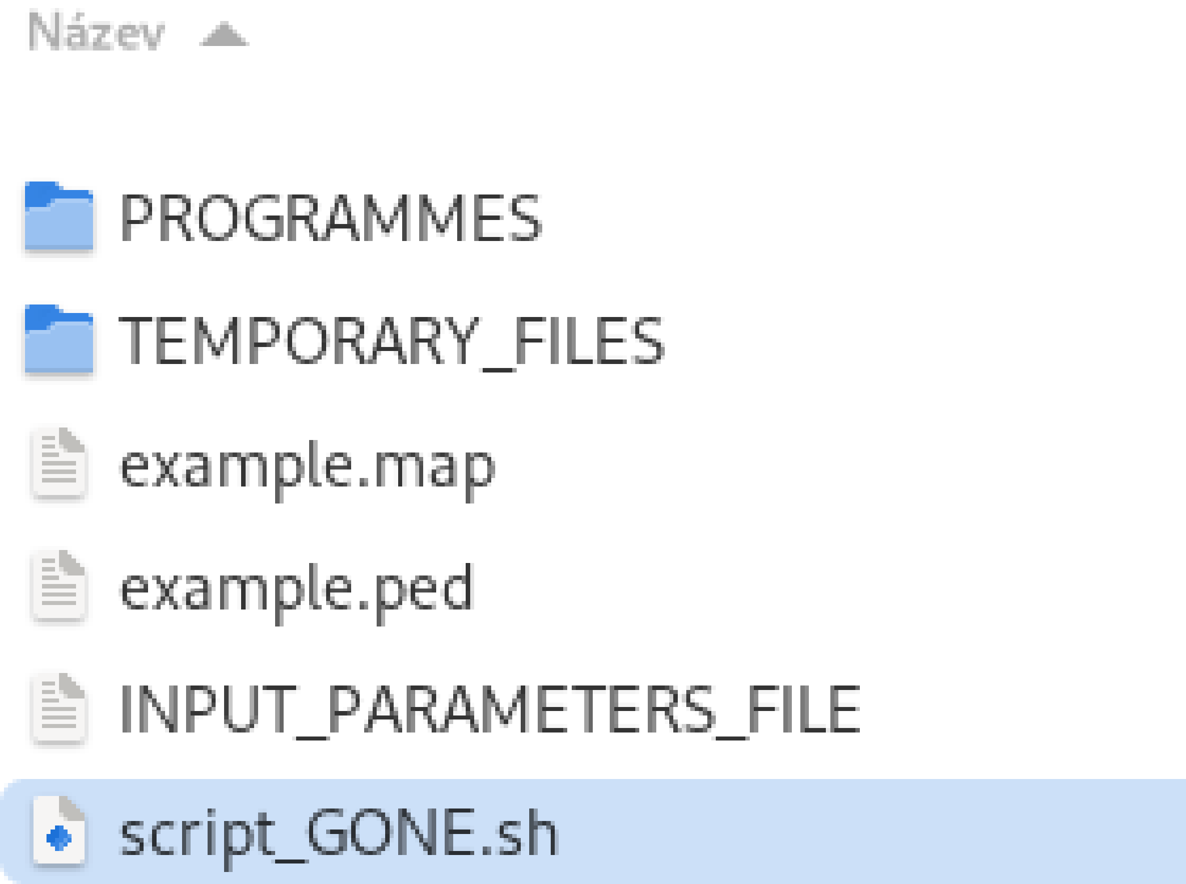
**Published:** 08 July 2020

https://doi.org/10.1093/molbev/msaa169

ISAGREED

Co-funded by the European Union

# After download



Název

PROGRAMMES
TEMPORARY_FILES
example.map
example.ped
INPUT_PARAMETERS_FILE
script_GONE.sh

Co-funded by
the European Union

# Input files

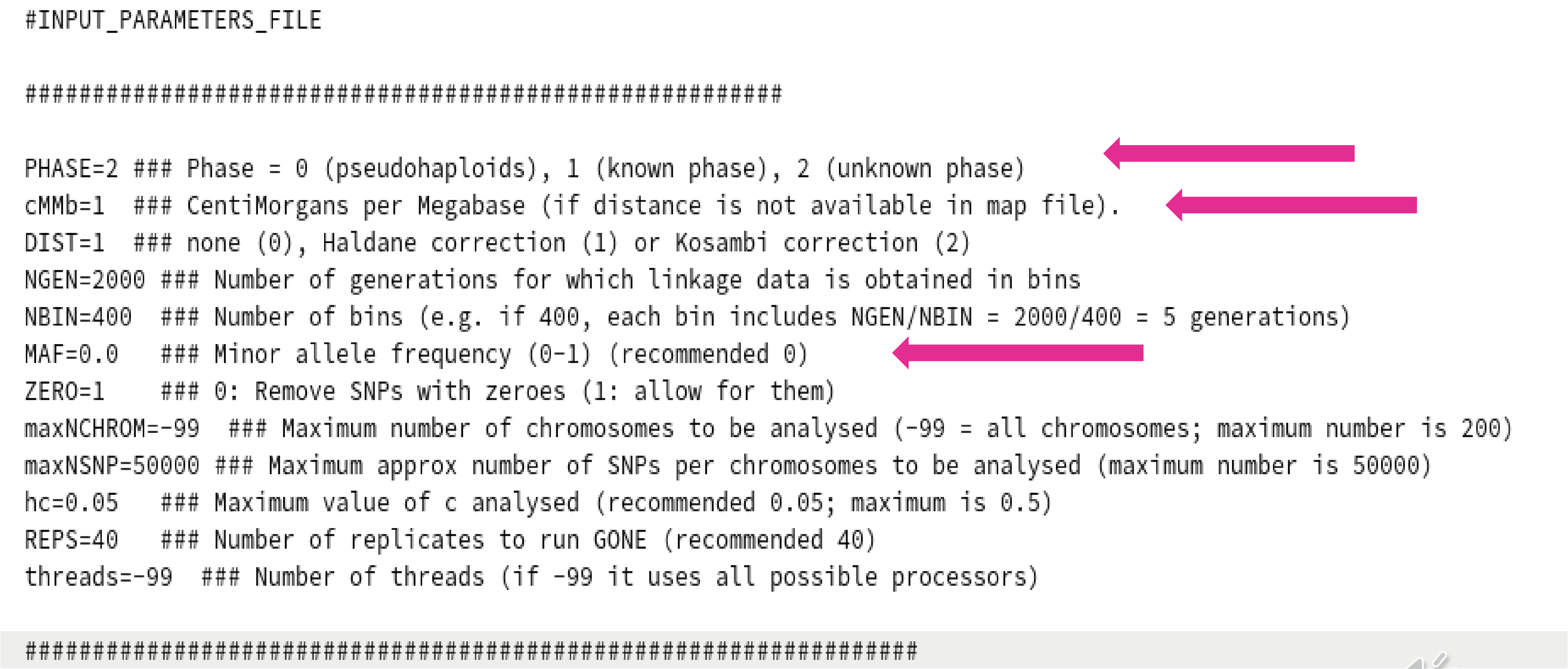
- Plink format
  - .map
    - Maximum number of SNPs per chromosome: 100,000
    - Maximum number of SNPs per genome: 1 million
  - .ped
    - 1 IND1 0 0 1 -9 T A A T ...
    - 1 IND2 0 0 1 -9 T T A T …

      or
    - 1 IND1 0 0 1 -9 2 1 1 2 ...
    - 1 IND2 0 0 1 -9 2 2 1 2 …

      Minimum number of individuals: 2
    - Maximum number of individuals: 1,800

Co-funded by the European Union

# INPUT_PARAMETERS_FILE



```
#INPUT_PARAMETERS_FILE

###########################################################

PHASE=2 ### Phase = 0 (pseudohaploids), 1 (known phase), 2 (unknown phase)
cMMb=1  ### CentiMorgans per Megabase (if distance is not available in map file).
DIST=1  ### none (0), Haldane correction (1) or Kosambi correction (2)
NGEN=2000 ### Number of generations for which linkage data is obtained in bins
NBIN=400  ### Number of bins (e.g. if 400, each bin includes NGEN/NBIN = 2000/400 = 5 generations)
MAF=0.0   ### Minor allele frequency (0-1) (recommended 0)
ZERO=1    ### 0: Remove SNPs with zeroes (1: allow for them)
maxNCHROM=-99  ### Maximum number of chromosomes to be analysed (-99 = all chromosomes; maximum number is 200)
maxNSNP=50000 ### Maximum approx number of SNPs per chromosomes to be analysed (maximum number is 50000)
hc=0.05   ### Maximum value of c analysed (recommended 0.05; maximum is 0.5)
REPS=40   ### Number of replicates to run GONE (recommended 40)
threads=-99  ### Number of threads (if -99 it uses all possible processors)

##################################################################
```

Co-funded by
the European Union

# To run the software

```
bash script_GONE.sh <FILE>
```

```
bash script_GONE.sh example
```

Co-funded by
the European Union

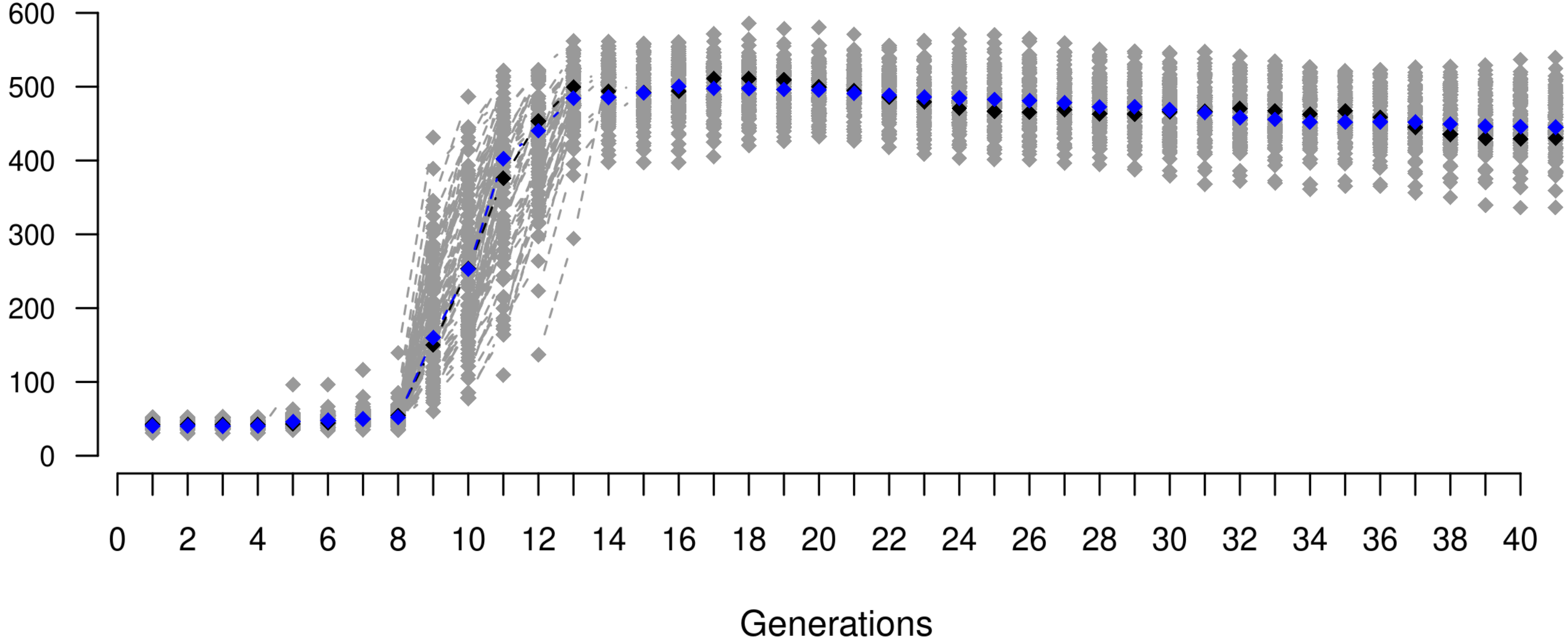
# Tips 1

- maxNSNP=50000 ### Approximated number of SNPs per chromosome to be analysed (maximum is 100,000; default 50,000). If the number of SNPs in the chromosome is **lower than this number, all SNPs of the chromosome will be analysed**.

- **If the value is larger, however, a random sample of 50,000 SNPs will be used. This is useful to avoid too lengthy estimations. In addition, different runs may include different random subsets of SNPs allowing for empirical errors of temporal Ne to be obtained. If you do so, do not sent all runs at the same time, as the initial random seed is taken from the computer clock time.**

Co-funded by the European Union

# Confidence intervla



Co-funded by the European Union

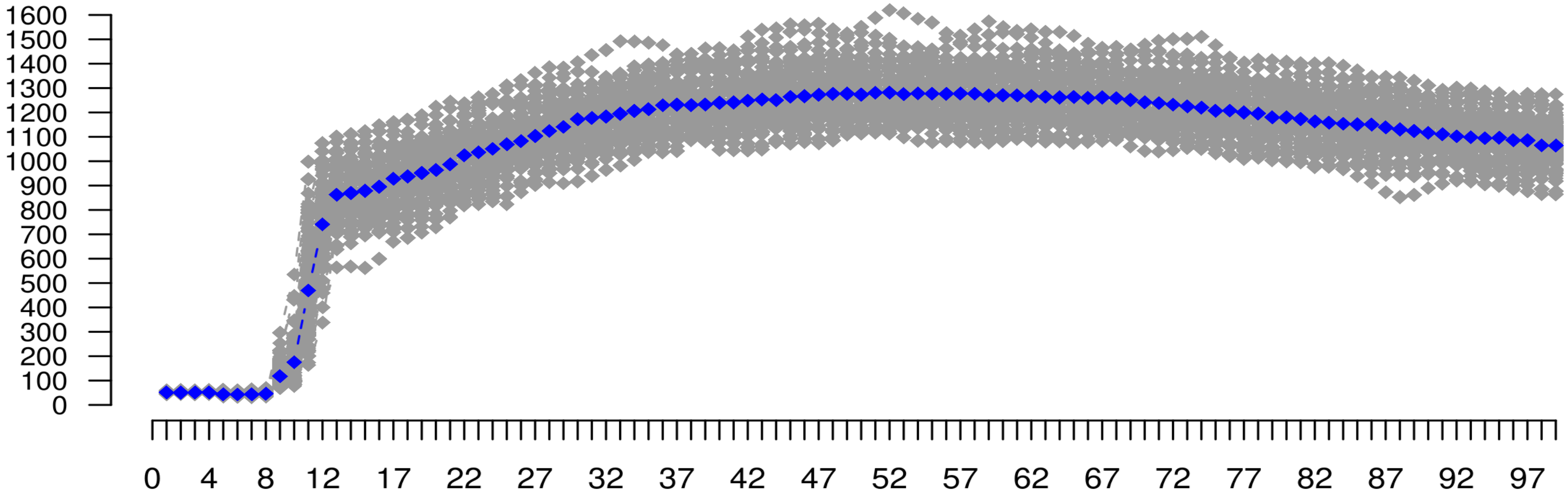
# Tips 2

- NGEN=2000 ### Number of generations
  - Realistic estimation on midle density SNPs are 100 generations
    - Questions???? 100 generation by livestock?

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

Co-funded by
the European Union

# 100 generations – Old Kladruber horse



Co-funded by the European Union

ISAGREED

# Case studie
# Estimation of effecitve population size

Co-funded by the European Union

# Aims of the study

Estimation of effective population size for local sheep breeds in the Danube region

1. **Estimation current effective population size**

2. **Estimation historical effective population size**

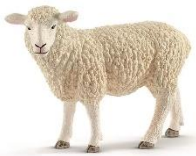
Co-funded by the European Union

# Analysed breeds

ISAGREED

**Czechia**
Improved Valachian
Original Valachian
umava sheep

**Slovakia**
East Friesian
Improved Valachian
Lacaune
Original Valachian
Texel
Tsihaja
Slovak Dairy
Crossbreed



**Czechia**
White shorthair
Brown shorthair
Landrace
Alpine

**Slovakia**

White shorthair

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

Co-funded by
the European Union

# Methods: Identification of selection signals

NeEstimator v.2 (Do et al., 2014)

- The leading software in the last decades
- STR/SSR data and SNP data
- In **small sample size** – **infinity** when pop. Size is large

- GONE (Santiago et al., 2020) - historical and current $N_e$

  - average over the most recent generations in the past
- CurrntNe (Santiago et al., 2020) - current $N_e$

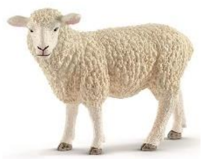
k = 0, monogamy was ignored, that is, the number of full siblings, k, of a random individual was set to 0

k = 2, corresponding to full lifetime monogamy,

Co-funded by
the European Union

# Results - sheep

|  | **Breed** | **N.** | **NeEstimator** | **GONE** | **CurrentNe k=2** |
|---|---|---|---|---|---|
| CZ | Improved Valachian | 39 | 352 (217-873) | 303 (211-435) | 542 (317-927) |
|  | Sumava sheep | 48 | 52 (37-78) | 73 (58-92) | 98 (78-123) |
|  | Valachian | 72 | 40 (30-55) | 65 (55-77) | 80 (76-104) |
| SK | East Friesian | 44 | ***30 (21-48)*** | 67 ( 53-84) | 52 (43-64) |
|  | Improved Valachian | 71 | **106 (73-177)** | 265 (210-331) | **169 (138-208)** |
|  | Lacaune | 104 | 81 (63-109) | 121 (105-140) | 111 (98-126) |
|  | Valachian | 94 | **34 (29-40)** | 75 (64-85) | 75 (67-84) |
|  | Tsihaja | 70 | 97 (58-233) | 552 (421-725) | 39 (33-45) |



Co-funded by
the European Union

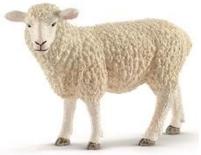
# Results - goat

|  | **Plemeno** | **N.** | **NeEstimator** | **GONE** | **CurrentNe k=2** |
|---|---|---|---|---|---|
| CZ | White shorthair | 36 | 105 (68-205) | 254 (179-356) | 181 (125-263) |
|  | Brown shorthair | 34 | 46 (31-82) | 127 (93-173) | 90 (67-122) |
|  | Landrase | 20 | 21 (13-29) | 116 (74-179) | 37 (26-52) |
|  | Alpine | 28 | 44 (28-87) | 127 (89-180) | 91 (63-130) |
| SK | White shorthair | 24 | **37 (23-78)** | 118 (79-173) | 74 (51-108) |

Co-funded by the European Union

# Results - sheep

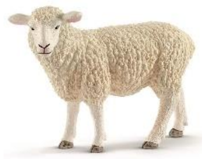
| | **Breeds** | **N.** | **NeEstimator** | **GONE** | **CurrentNe k=2** |
|---|---|---|---|---|---|
| SK | Crossbreed | 19 | 13 (4-77) | 57 (41-80) | 28 (20.-38) |
| | Slovak Dairy | 36 | 46 (30-62) | 94 (70-123) | 78 (60-103) |
| | Texel | 5 | Inf. | 138 (45-427) | 182 (27-1231) |

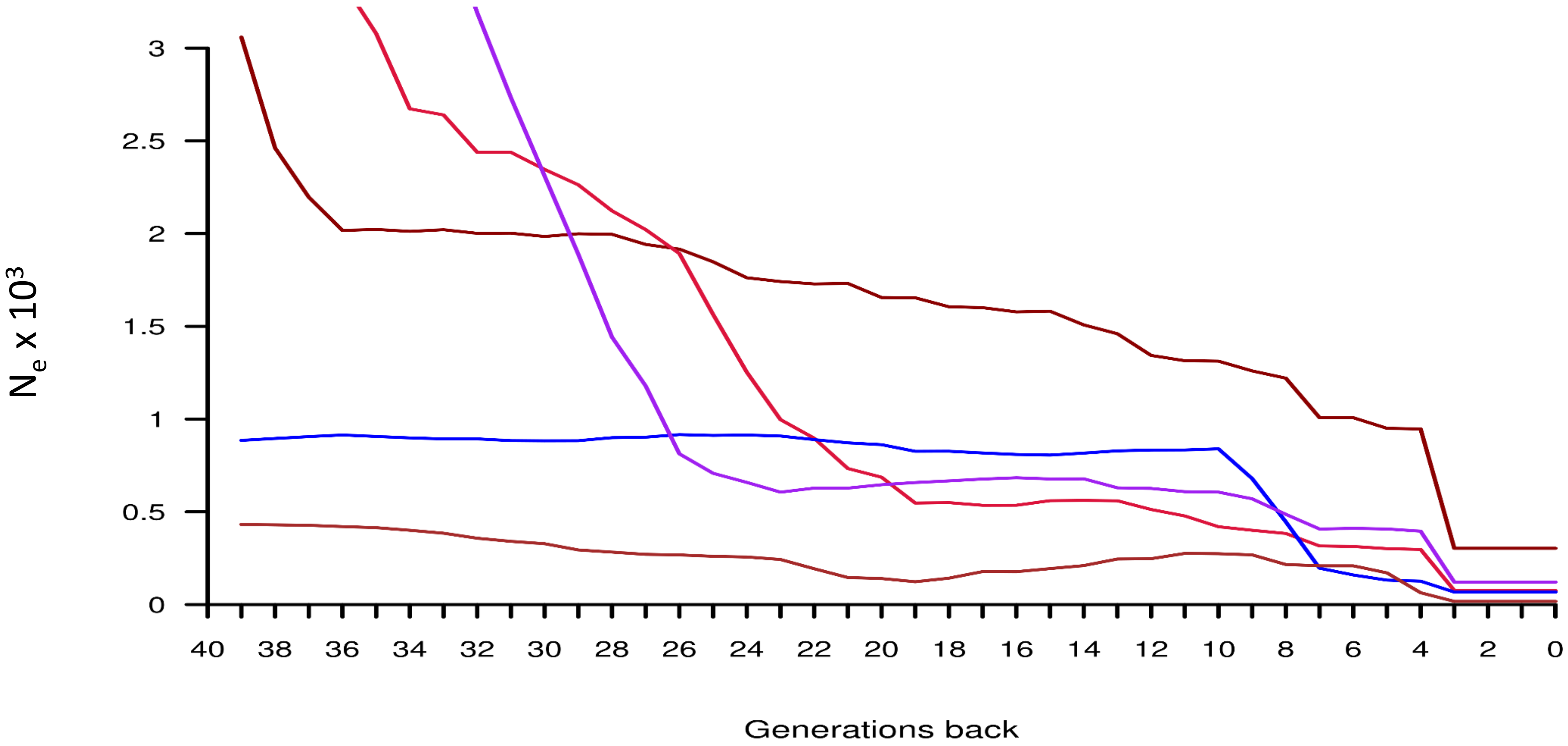


Erasmus+ project 2021-1-SK01-KA220-HED-000032068

Co-funded by the European Union

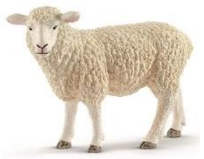
# Results - sheep

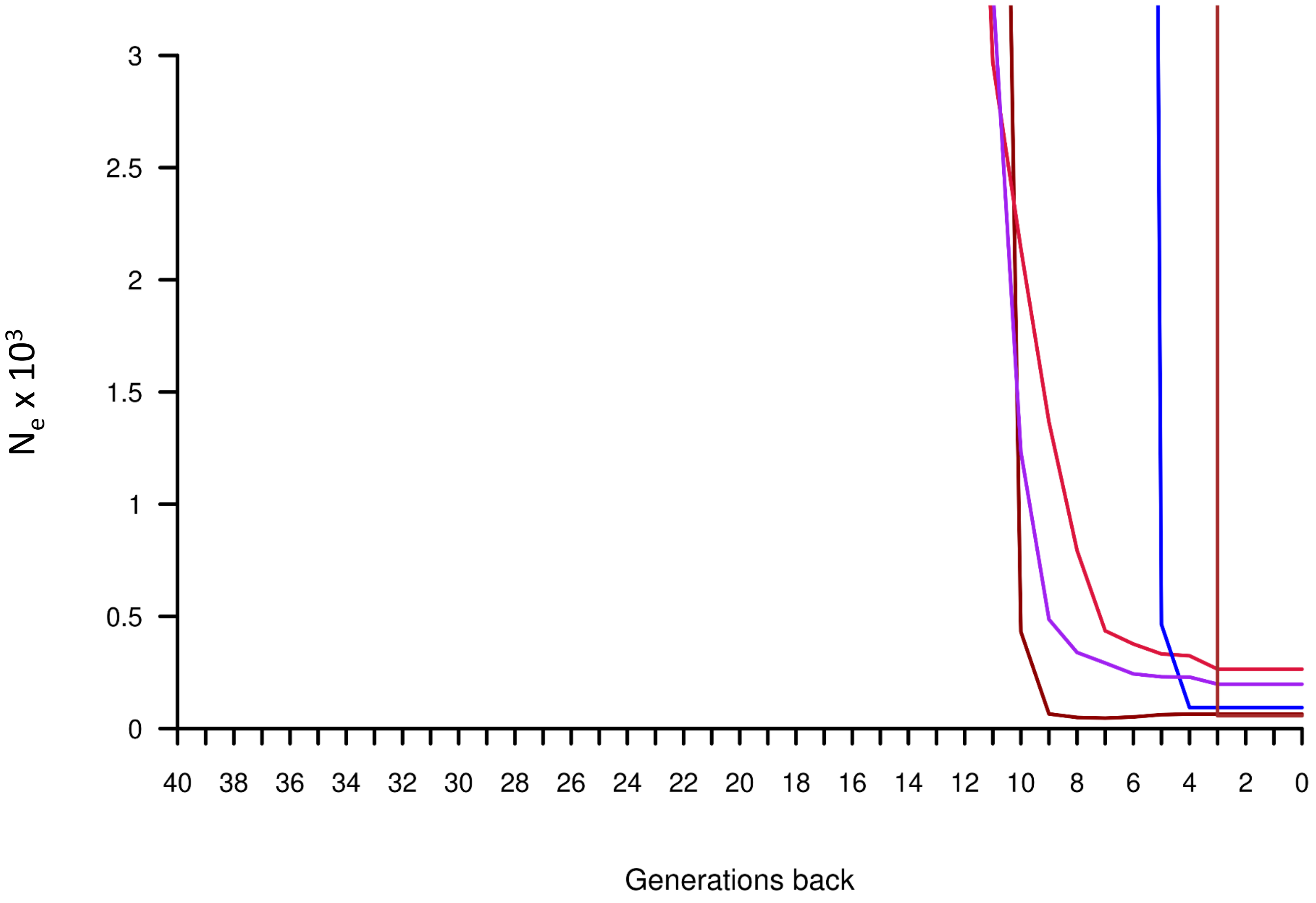




ISAGREED
Co-funded by the European Union

# Results - sheep

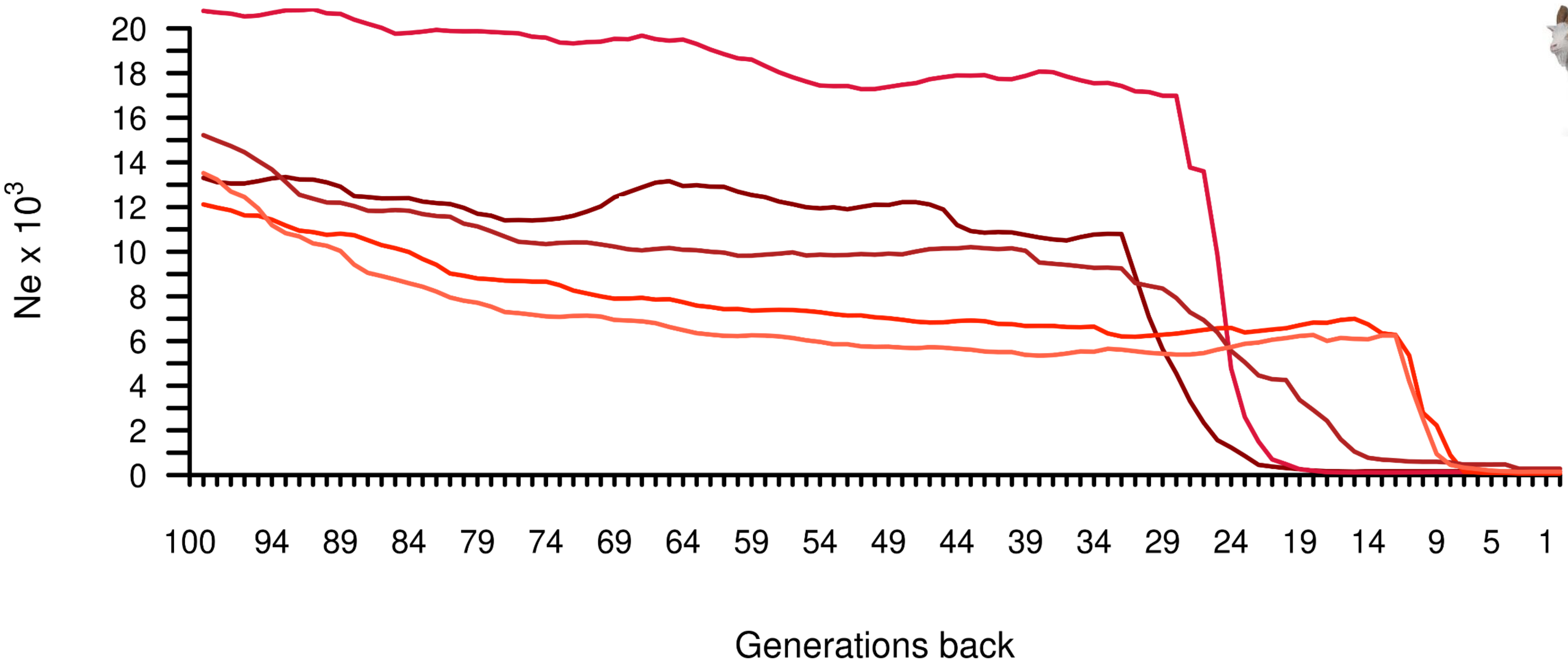




Erasmus+ project 2021-1-SK01-KA220-HED-000032068

ISAGREED
Co-funded by the European Union

# Results - goat

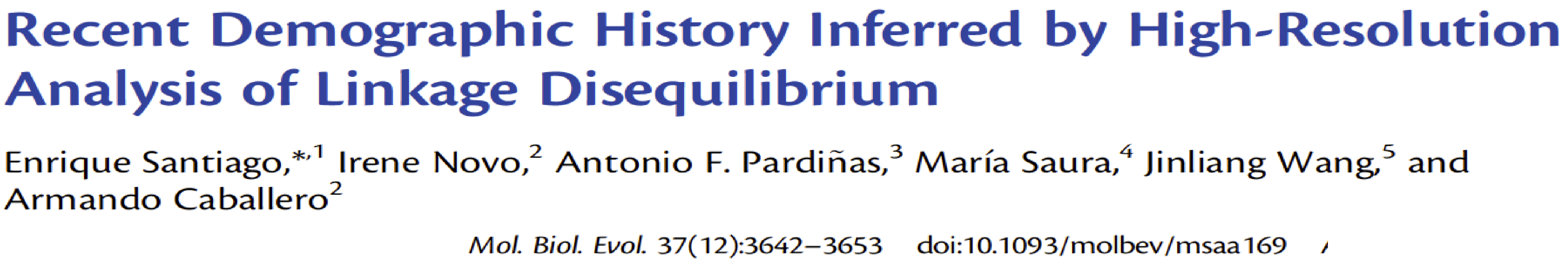


ISAGREED

Co-funded by the European Union

# Conclusion – Case study

**Software GONE**



Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium

Enrique Santiago,*[1] Irene Novo,[2] Antonio F. Pardiñas,[3] María Saura,[4] Jinliang Wang,[5] and Armando Caballero[2]

*Mol. Biol. Evol.* 37(12):3642–3653 doi:10.1093/molbev/msaa169

# Reference

https://www.nhkladruby.cz/fotogalerie

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

ISAGREED
Co-funded by the European Union

ISAGREED

## Partners:

Slovak University of Agriculture Nitra

Mendel University in Brno

Siedlce University of Natural Sciences and Humanities

Czech University of Life Sciences Prague

# 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.*

Luboš Vostrý
vostry@af.czu.cz

ISAGREED

Co-funded by the European Union