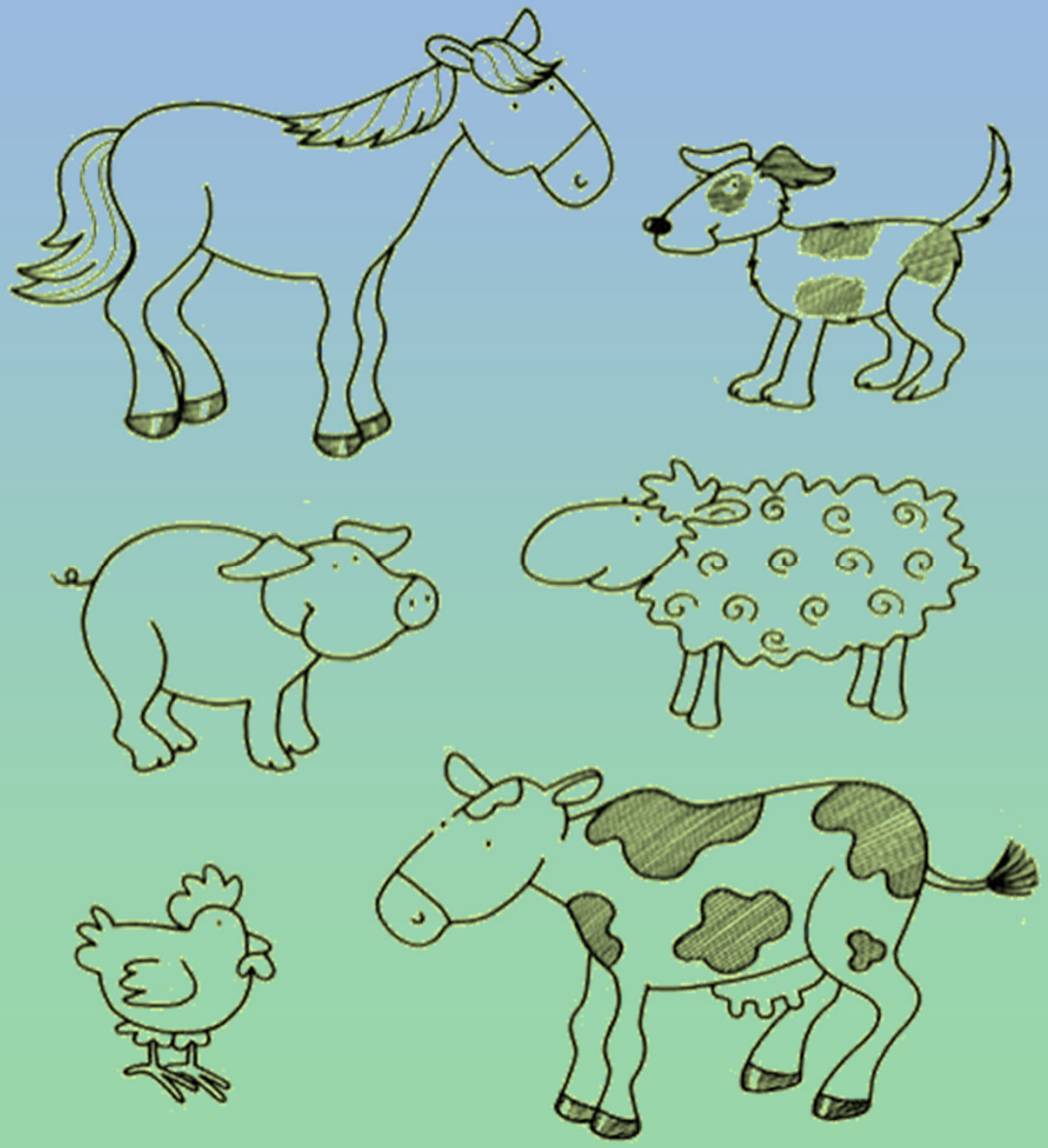
Genomic characterization of animal genetic resources

Johannes A. Lenstra, Utrecht University

FAO Webinar December 14, 2022



Genomic characterization of animal genetic resources

Johannes A. Lenstra, Utrecht University

FAO Webinar December 14, 2022

New FAO guidelines: most comprehensive description of

Sample → DNA isolation → dataset (SNPs, WGS) → analysis

The FAO Guidelines will be recommended in the Authors' Guidelines of the ISAG journal Animal Genetics

Genomic characterization of animal genetic resources

Johannes A. Lenstra, Utrecht University

FAO Webinar, December 14, 2022

- Datasets more and more powerful
- Analysis more and more sophisticated
- Basic analysis still relies on the golden trio

1. PCA → survey, clines
2. Model Based Clustering (STRUCTURE, ADMIXTURE) → Genomic components
3. Genetic distances → Breed relationships, phylogeny

Genomic characterization of animal genetic resources

Johannes A. Lenstra, Utrecht University

FAO Webinar, December 14, 2022

But:

1. **What's in a breed?** Check prior to downstream analysis!

2. The basic analyses by **PCA and Model Based Clustering (STRUCTURE, ADMIXTURE)** are often distorted by uneven sampling and inbreeding and are often misinterpreted

3. NeighborNet of **F_{ST} genetic distances** + **f_3 / f_4 / BSAA** are the most informative methods for inferring genetic history

1. What's in a breed?

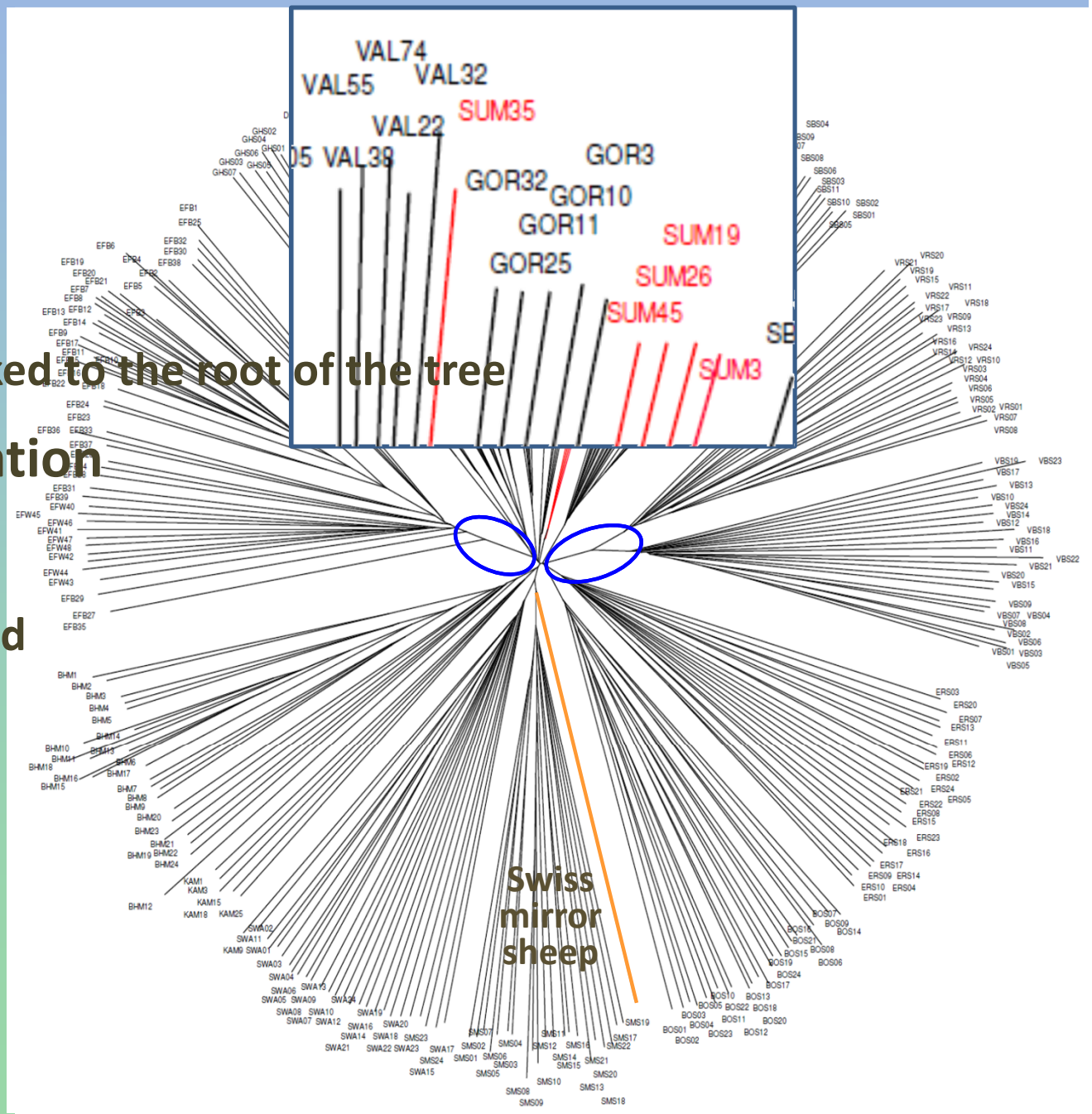
← NJ trees of allele sharing distances

Errors

- identical samples
- mislabelling
- crossbred outliers linked to the root of the tree

Pattern of differentiation

- genetic isolation →
- PCA, ADMIXTURE spoiled by inbreeding bias



1. What's in a breed?

← NJ trees of allele sharing distances

⚠ **Errors**

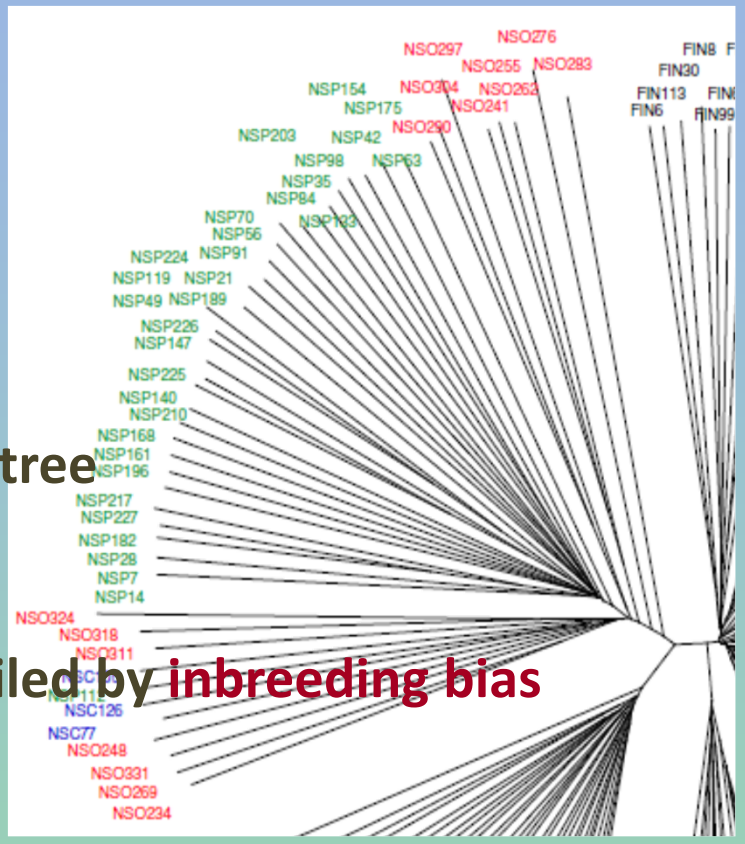
- 😟 **identical** samples
- 😟 **mislabelling**
- 😟 crossbred **outliers** linked to the root of the tree

🔍 **Pattern of differentiation**

- 🙂 genetic **isolation** → 😟 PCA, ADMIXTURE spoiled by **inbreeding bias**
- 😯 degree of breed differentiation

1. What's in a breed?

← NJ trees of allele sharing distances



⚠ **Errors**

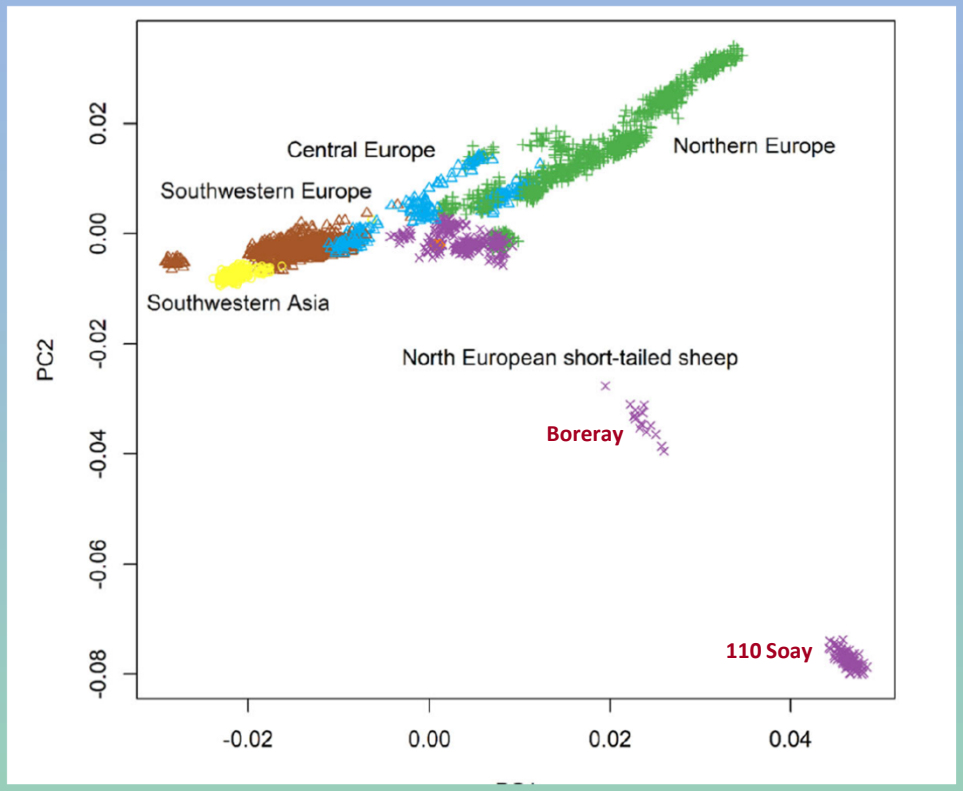
- 😟 **identical** samples
- 😟 **mislabelling**
- 😟 crossbred **outliers** linked to the root of the tree

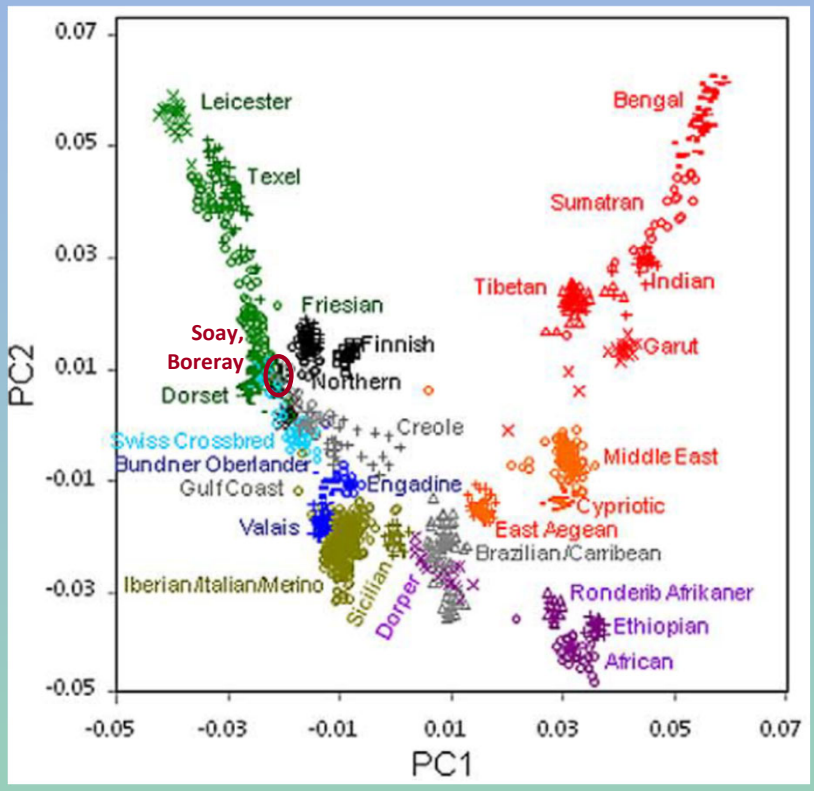
🔍 **Pattern of differentiation**

- 🙂 genetic **isolation** → 😟 PCA, ADMIXTURE spoiled by **inbreeding bias**
- 😯 degree of breed differentiation
- 😯 **subdivision** within breeds → keep separate for downstream analysis!

2. PCA often distorted by uneven sampling and inbreeding

→ Gradual clines by ancient migrations not displayed





European and Asian sheep, ~20/breed,
110 Soay, highly inbred Scottish island sheep

☹ PC2 reflects Soay ↔ others

☹ Boreray from nearby island intermediate

Overlapping samples, **20 Soay**

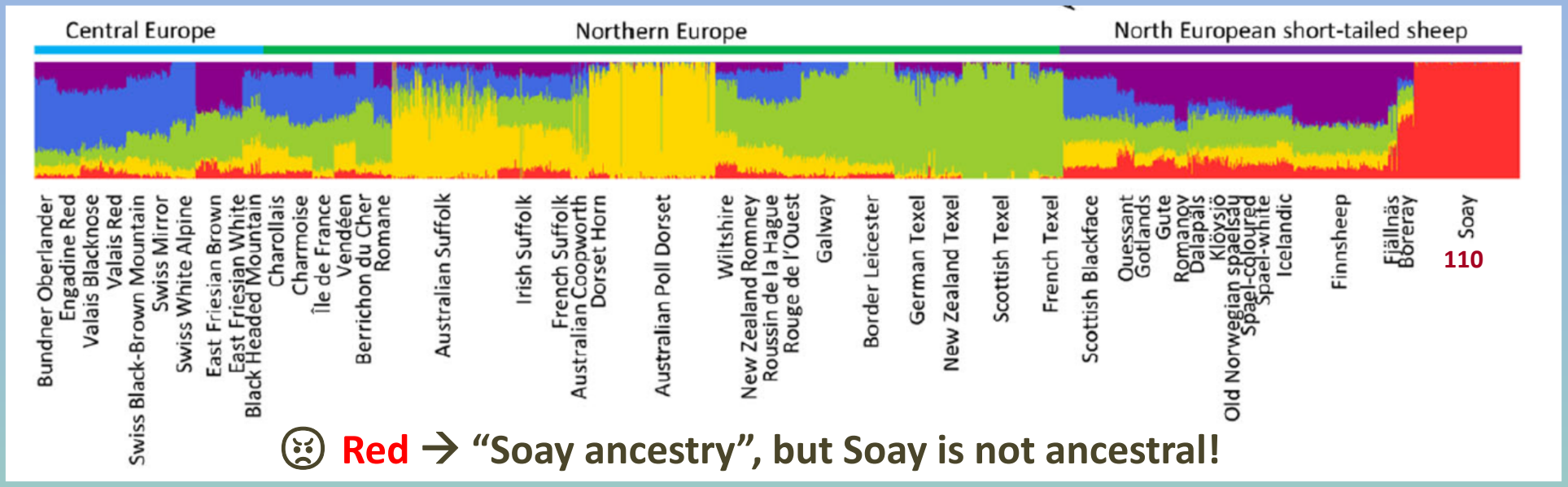
☺ Soay and Boreray plotted according to their ancestry

😊 PC1 and PC2 show several clines

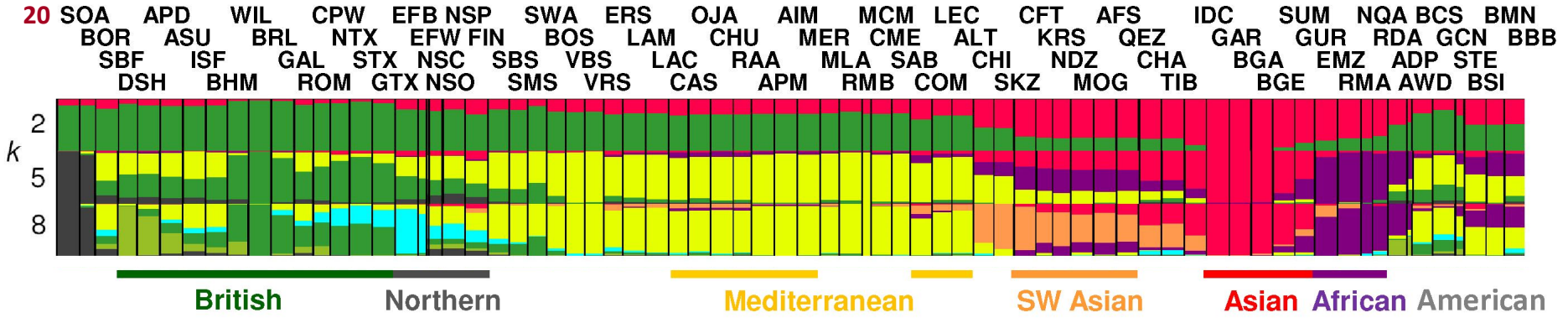
😄 Powerful alternative: **Supervised PCA**:

- PCs are set by a subset of samples
- Other samples are interpolated within this subset

2. ADMIXTURE also distorted by uneven sampling and inbreeding



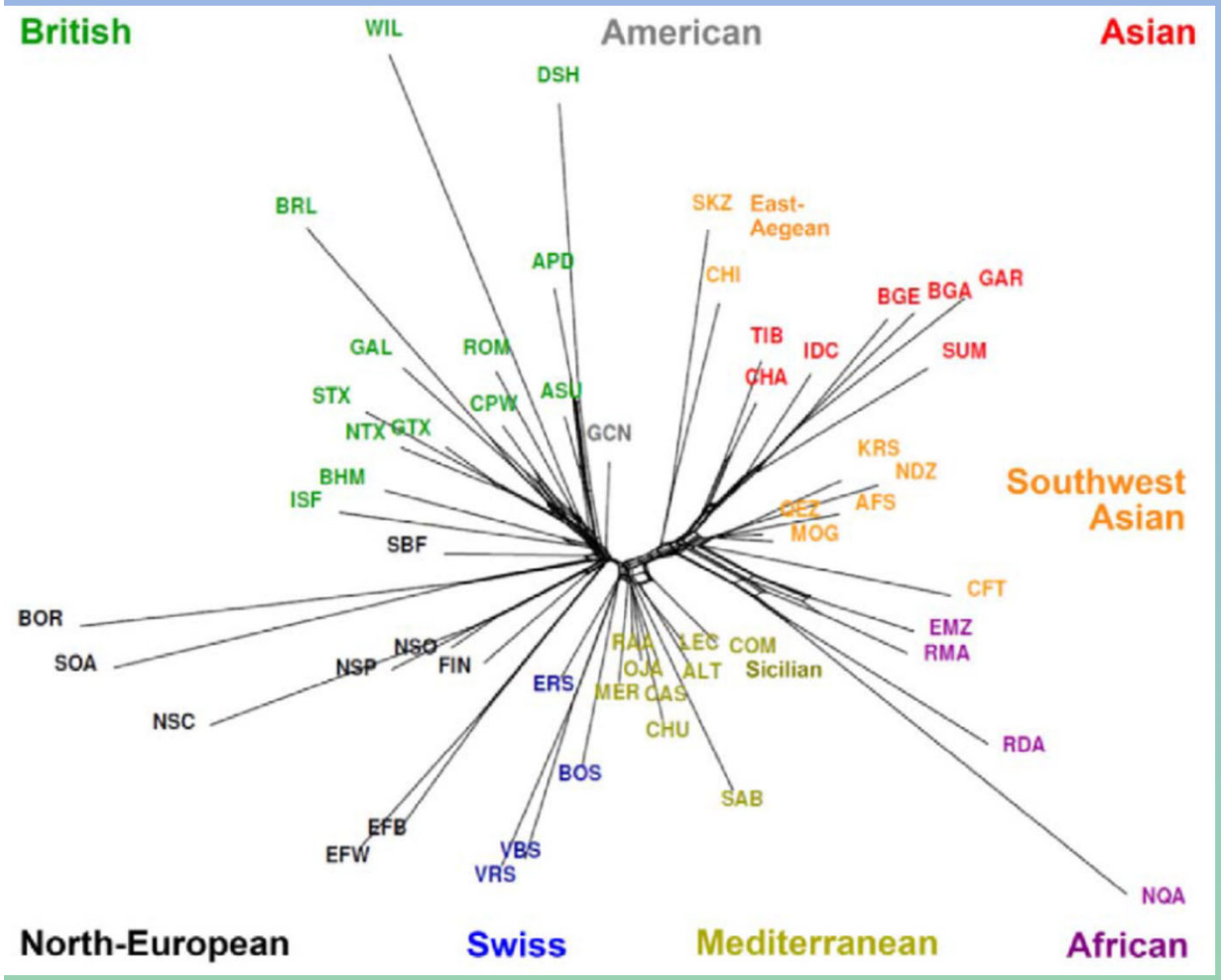
😠 **Red → "Soay ancestry", but Soay is not ancestral!**



😟 **Dark grey shows a reduced Soay+Boreray ancestry ~ Nordic ancestry**

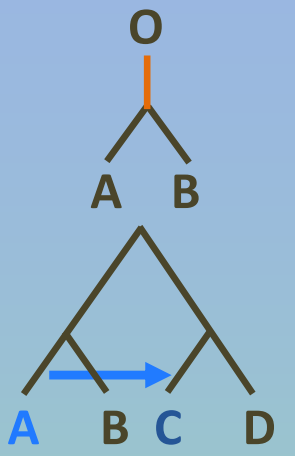
Inferred clusters are a group or related animals, but are ~~not always~~ almost never an ancestral component !!

3. NeighborNet of F_{ST} genetic distances



- ☺ Not influenced by sampling or inbreeding bias
- ☺ Shows regional clusters
- 😄 Lines reconstruct gene flows → link with history
- 😮 Reticulations by crossbreeding only partially shown

3. f_3 / f_4 / BSAA are the most informative methods for inferring crossbreeding



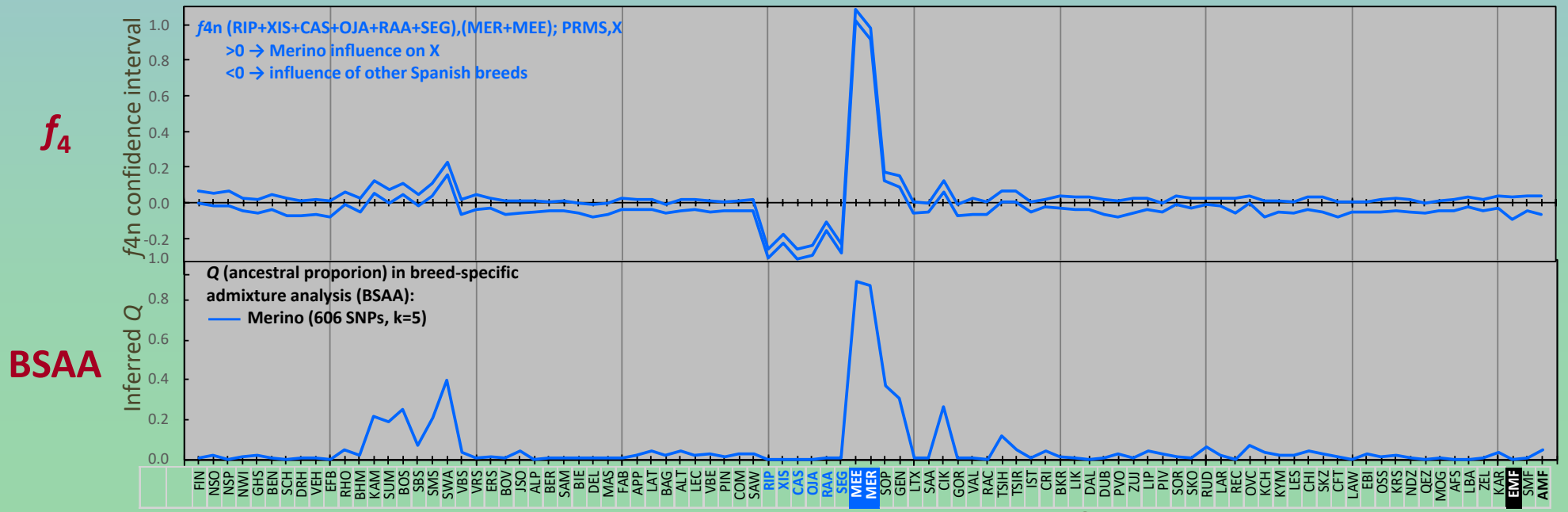
f_3, coancestry coefficient, gene flow shared by A and B
after diverging from outgroup; increases by gene flow A → B

f_4, admixture coefficient, correlation of alleles of breeds
in different part of the tree; >0 by gene flow from A → B

BSAA (breed-specific admixture analysis): model based clustering (STRUCTURE, ADMIXTURE) with Ancestry Informative Markers (AIMs): SNPs that differentiate the donor and nearby other breeds

f_4

BSAA



Swiss

Merino Hungary: forgotten Merino incrossing in 18th

Genomic characterization of animal genetic resources

Johannes A. Lenstra, Utrecht University

FAO Webinar, December 14, 2022

Keep in mind

- Using sophisticating programs does not imply that you are dumb and the computer is bright.
- Those software developers may lose sight of the biological reality
- “All models are wrong, but some are useful”. Many models are useless!

Do's and Don’ts

- Don’t cut-and-paste from computer output in your paper
- **Understand** the algorithm
- Do **validate** a program as in the lab: positive and negative controls, varying parameters, omitting strategic breeds, splitting the dataset, etc.
- Read about **human molecular genetics** on advanced methodology