Bio**geo**informatics of livestock genomic resources

Stéphane JOOST

Lab of Geographic Information Systems (LASIG)

Ecole Polytechnique Fédérale de Lausanne
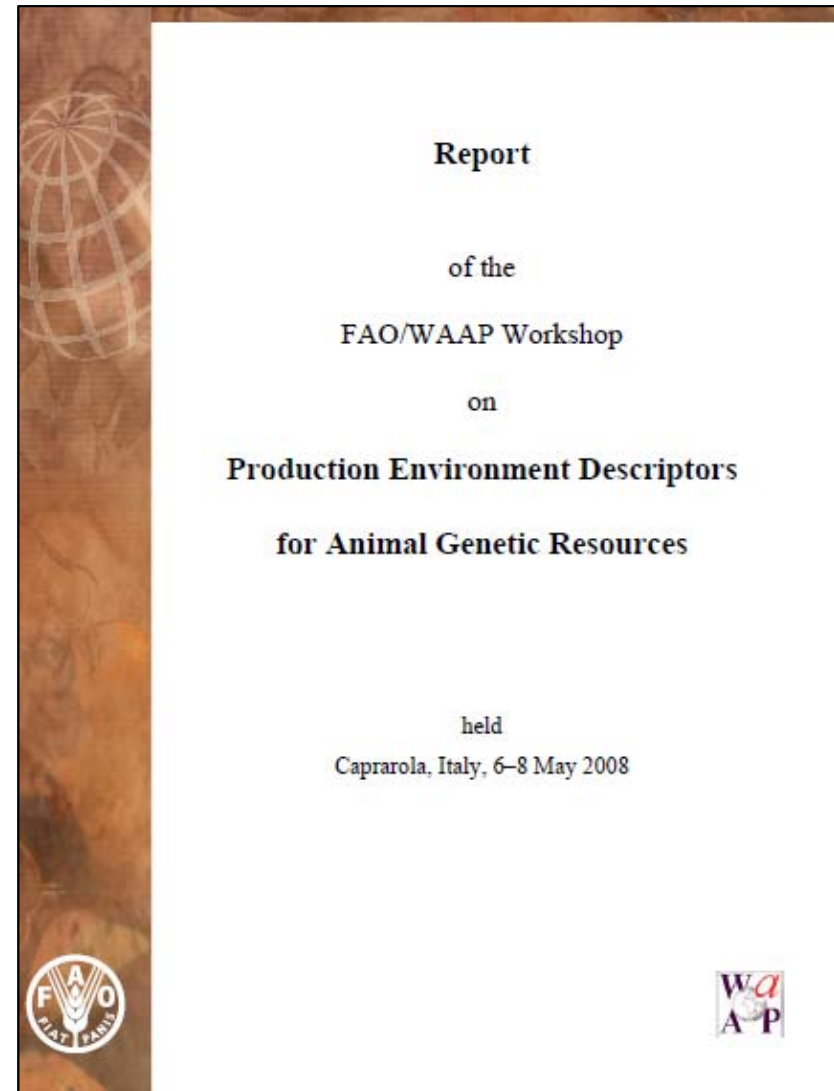
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

GEOEXPERIMENT

GEOgraphy matters!

Federal Emergency Management Agency (FEMA), Flooded Oklahoma 2007 (Creative Commons)

Report on PEDs for FAnGR in 2008

- Edited by FAO & WAAP

Report

of the

FAO/WAAP Workshop

on

Production Environment Descriptors

for Animal Genetic Resources

held
Caprarola, Italy, 6–8 May 2008

Recommendation

Conclusions

The analysis presented above leads to four main conclusions for the further development of PEDs:

1. The selection of criteria, indicators and variables to be included in the descriptive scheme should be guided by their relation to adaptive traits and the degree to which they affect animal performance.

2. Natural and management environment should be distinguished to account for the difference between external variables (not under the control of the livestock keeper) and internal factors (controlled by livestock keepers) affecting the adaptedness of species/breeds and animal performance. This distinction also facilitates the operationalization of data collection.

3. The georeferencing of breed locations in order to enable linkages to information available in other georeferenced databases should be initiated as soon as possible. This will decrease the required data collection and data entry by National Coordinators for the Management of Animal Genetic Resources and enable a continuous refinement of the description of production environments, as it is expected that an increasing number of datasets, with improving resolutions, will become available in the near future.

GeoSituation in 2014

Non-exhaustive review of national and international projects including sampling of livestock individuals
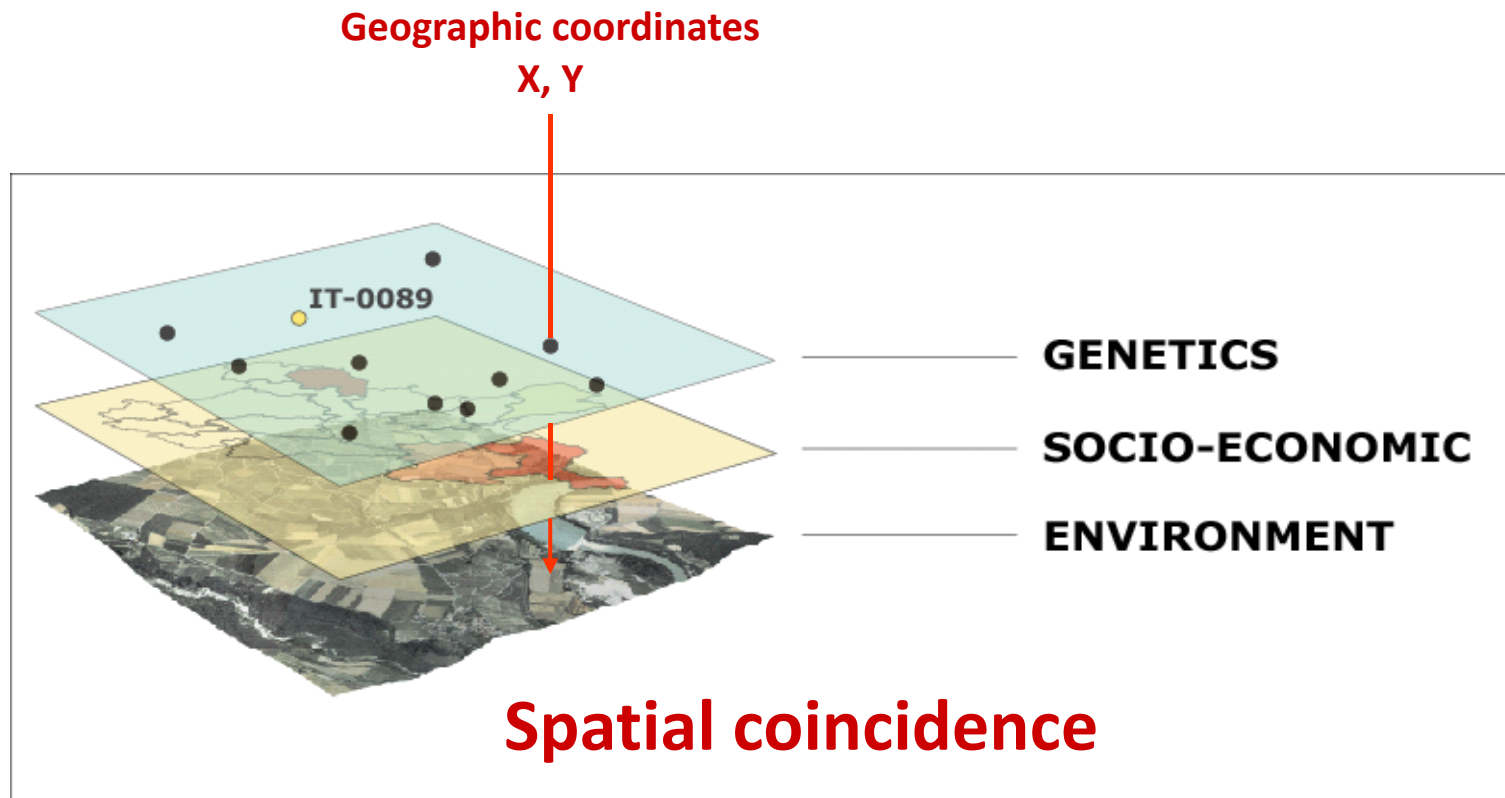
- 2000-2014: 2/15 = 13%

ISI Web of science

- «Livestock» & «GIS» 2000-2007 = 127 = 9.8/y
- «Livestock» & «GIS» 2008-2014 = 170 = 24.2/y **+**
- «Livestock» & «GPS» 2000-2007 = 34 = 4.2/y
- «Livestock» & «GPS» 2008-2014 = 116 = 16.5/y **+**

A key function

Geographic coordinates
X, Y

IT-0089

GENETICS

SOCIO-ECONOMIC

ENVIRONMENT

Spatial coincidence

FAnGR conservation

- Management and conservation of livestock genetic resources imply breed prioritization, and therefore decision making

- Decision making rests on the simultaneous analysis of several criteria…

- … in order to identify and to favour sustainable breeding conditions

Data categories

1. Population and evolutionary genetics
2. Animal husbandry practices
3. Socio-economic and socio-demographic data in the regions where animals are bred
4. Environmental information: climatic and geophysical characteristics of the places where animals are bred
5. Political and administrative boundaries: geographical units where policies have to be applied
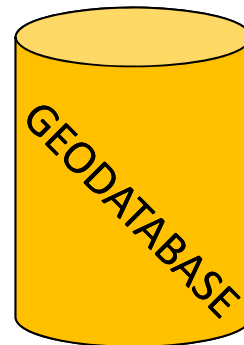
Data integration

Administrative boundaries
Socio-Economic data
Socio-demographic data

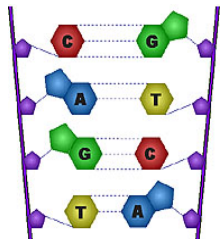Environmental data:
topography, climate,
soil, etc.
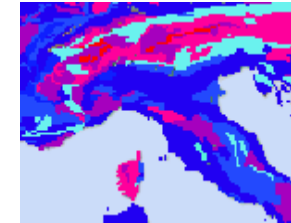
**Geographic coordinates
X, Y**

GEODATABASE

Sampling
Genetic data

GIS

Questionnaires
Husbandry pratices

FAnGR monitoring

- FAO Global plan (2007) requires countries to monitor their FAnGR

- Countries like Austria, Germany, Great Britain have developed a monitoring system

- GenMon prototype under development in Switzerland (Duruz 2014)

- The system uses geographic information to assess endangerment

- And to communicate information by means of thematic maps

GenMon-CH

Welcome

Welcome on GenMon-CH, an open Web-GIS application for the monitoring of Farm Animal Genetic Resources (FAnGR).

This application is designed to rank breeds according to four criteria: Genetic diversity (estimated from pedigree data), Introgression, Geographical concentration and Agriculture sustainability (from Socio

The ranking of the breed is shown in the following table, while more information for each breeds (table, graphs, maps) are available if you click on more info.

Please refer to the tutorial for more information and for test data to try the application.

Summary table

| Breed name	Last year of data	Number individals last GI	Average inbreeding last GI	Effective pop size range	Pedig Index (0-1)	Introg Index	Geog index (km)	BAS Index (0-1)	Global Index (0-1)	More Info
SN	2012	33306	0.103 ■	50-100 ■	0.008 ■	0 ■	13.02 ■	0.78 ■	0.393	more
FM	2013	26877	0.0571 ■	50-100 ■	0.383 ■	0.114 ■	57.66 ■	0.75 ■	0.454	more
BFS	2012	43341	0.0467 ■	50-100 ■	0.474 ■	0 ■	51.58 ■	0.707 ■	0.703	more
SBS	2012	37712	0.0411 ■	50-100 ■	0.523 ■	0 ■	59.53 ■	0.719 ■	0.746	more
BVO	2014	30469	0.033 ■	■	0.594 ■	0.013 ■	58.52 ■	0.74 ■	0.812	more
WAS			■		■	■		■		more
Test			■		■	■	■	■		more

Base Layer
- ○ Mean Inbreeding
- ◉ Max Inbreeding
- ○ Mean introgression
- ○ Number of animals
- ○ Socio-Economical Index

Overlays
- ☑ Selection
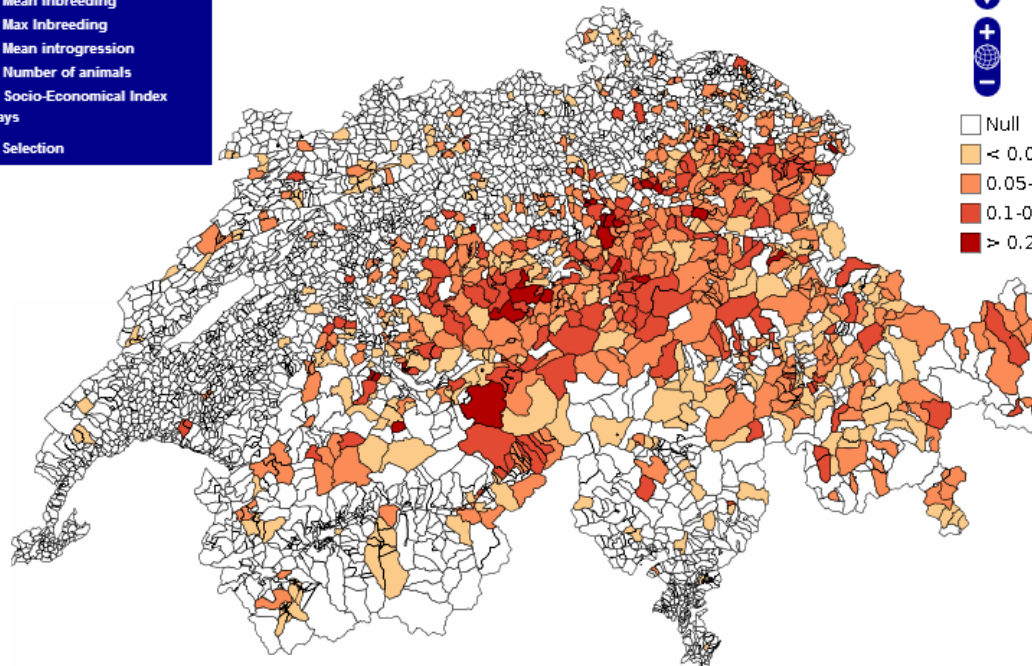
Legend:
- □ Null
- < 0.05
- 0.05-0.1
- 0.1-0.2
- > 0.2

Add data for this breed (BVO): [Choose File] No file chosen [Add]

See the format in the tutorial.

Or go to the assisted upload section [Assisted upload]

See spatial distribution (BVO): [Go to map]

PDF: PopRep Population Report BVO

PDF: PopRep Inbreeding Report

Effective population size

Table: Effective population size according to different computations
[Method] [Ne]

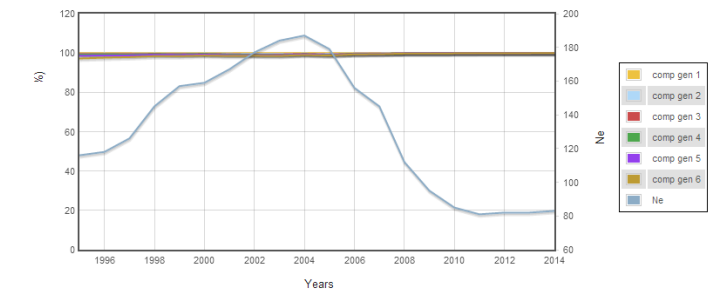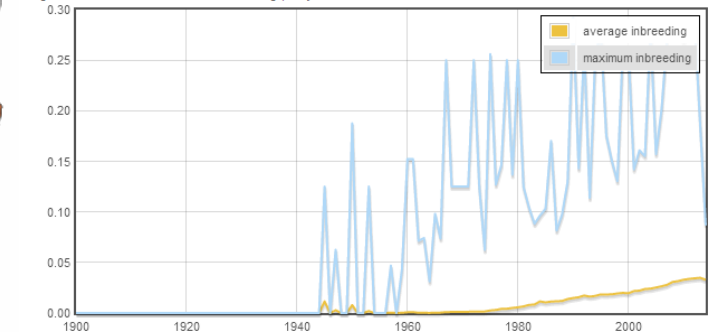Figure: Effective population size (Ne$_{\Delta F_P}$) and pedigree completeness

Legend:
- comp gen 1
- comp gen 2
- comp gen 3
- comp gen 4
- comp gen 5
- comp gen 6
- Ne

Figure: Maximum and mean inbreeding per year
- average inbreeding
- maximum inbreeding

Geointelligence

- Breed, demography, **bio**logy, population genetics

- Geodata (sampling design, geocoordinates)
- Communication skills, thematic mapping
- Database, spatial database, web protocols, web design
- GIS software, programming languages, data processing

- Computer science

ECOLOGY

Time to Tap Africa's Livestock Genomes

Olivier Hanotte,[1] Tadelle Dessie,[2] Steve Kemp[3]

Fortunately, the fields of genetics and genomics (3–5) offer a new start for the sustainable improvement of African livestock productivity. Landscape genomics links genome-wide information to geo-environmental resource analysis to identify potentially valuable genetic material. Typically, researchers will perform a genome-wide scan on a number of animals from populations living in different habitats or across an ecological cline (from dry to wet areas, for instance).

Landscape genomics
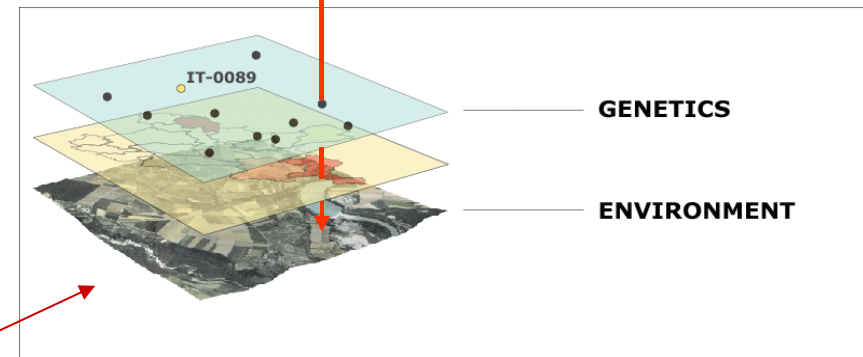
- Correlative approaches and spatial statistics

Genome-wide information

- Paradigm shift and transition phase

High performance computation of landscape genomic models integrating local indices of spatial association

Sylvie Stucki [1,*], Pablo Orozco-terWengel [2], Michael W. Bruford [2], Licia Colli [3], Charles Masembe [4], Riccardo Negrini [3,5], Pierre Taberlet [6,7], Stéphane Joost [1,*] and the NEXTGEN Consortium [8]
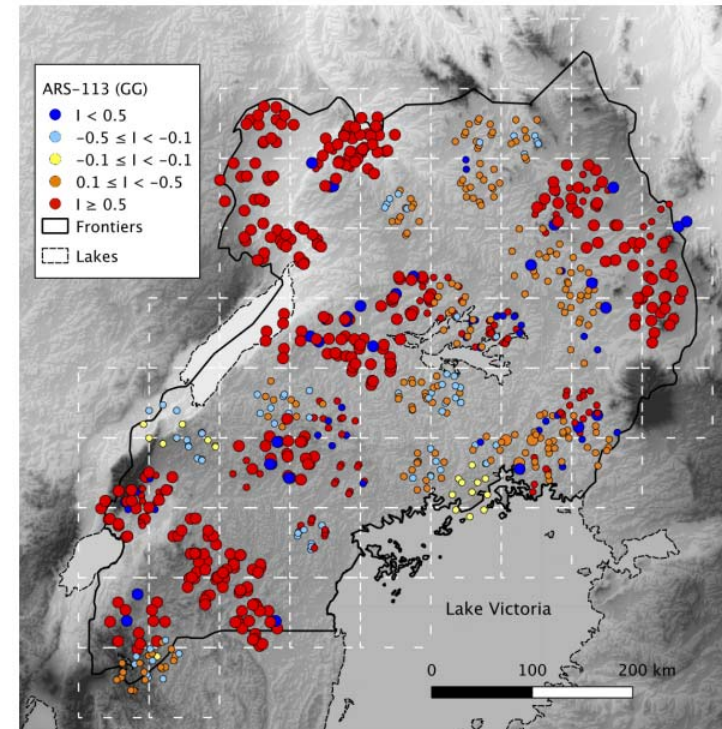
http://lasig.epfl.ch/**sambada**

Computation time (in hours)

Spatial statistics

	41,215 SNPs 804 samples	634,849 SNPs 102 samples
Samβada	1.2	2.9
Samβada biv.	8.7	18.4
BayEnv	41.3	62.,2
LFMM	3.2	16.0
LFMM (mono)	6.1	58.1

23 environmental variables

ARS-113 (GG)
- $I < 0.5$
- $-0.5 \leq I < -0.1$
- $-0.1 \leq I < -0.1$
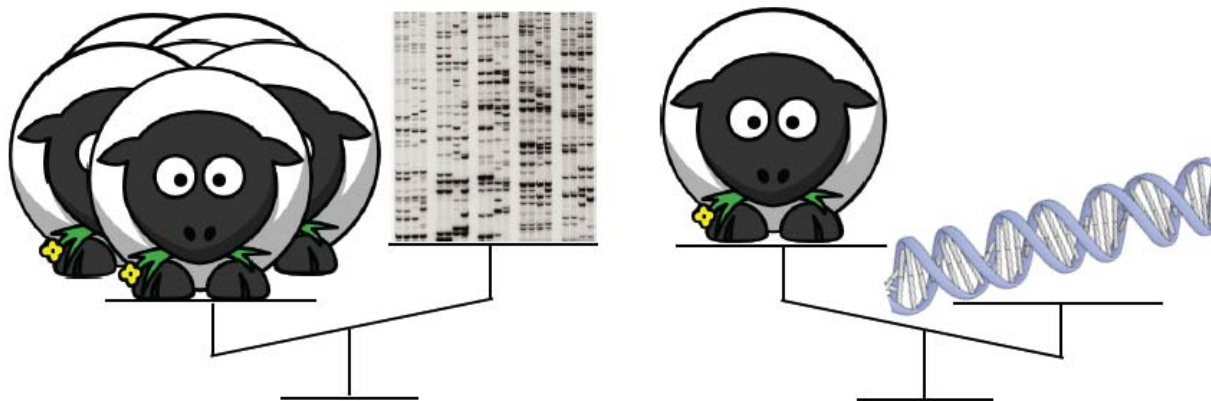- $0.1 \leq I < 0.5$
- $I \geq 0.5$
- Frontiers
- Lakes

Lake Victoria

0   100   200 km

(a) ARS-113 (GG)

Whole genome sequence (WGS) data

- WGS data available but still expensive (transition phase)

- Standard sampling strategy but...

- High cost of WGS limits the number of individuals to be sequenced

Picture: Sylvie Stucki

Sub-sampling

- How to maximize environmental information from a reduced number of locations in landscape genomics ?
- GIS, PCA and hierarchical clustering (Stucki 2014)
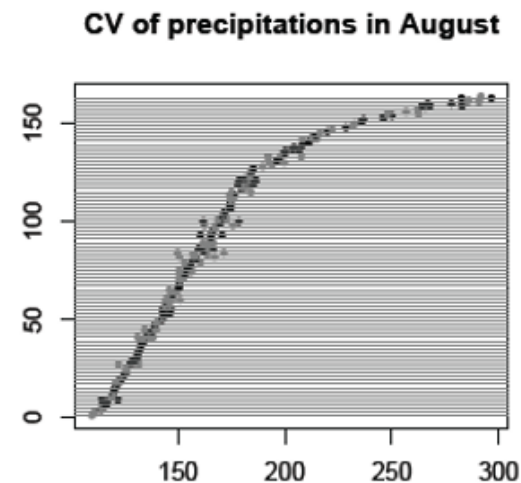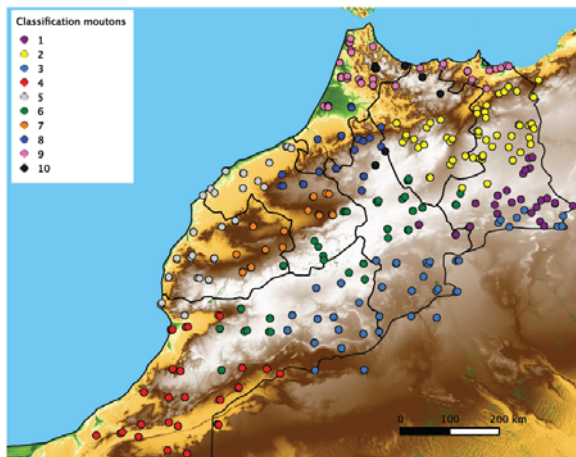
Geointelligence

- Breed, **bio**logy, population genetics, genomics

- Geodata (sampling design, geocoordinates)
- Communication skills, thematic mapping
- GIS software, programming languages
- Spatial statistics, High Performance Computing (HPC) for data processing

- Computer science

Biogeoinformatics

- In both cases, an original combination of skills including molecular biology, computer science and geographic information science is necessary
- New knowledge will be extracted from the present data tsunami (big molecular data, big environmental data, big socio-economic data)…
- … only if we train a new generation of students/scientists able to develop innovative transdisciplinary and efficient geocomputing tools

One among 20

- …most important problems we need to overcome for effective livestock genomic resources conservation

- to enforce the recording of geographical coordinates of any sampled animal as a **standard rule**

- so that we can fully benefit from the power of biogeoinformatics !

Thank you for your attention !