Biogeoinformatics

Stéphane JOOST
Lab of Geographic Information Systems (LASIG)
Ecole Polytechnique Fédérale de Lausanne

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

GIS Lab EPFL

- GIS application design
- Spatial Decision Support Systems
- Geospatial data infrastructures
- GIS for urban studies and design

- Spatial statistics and analysis

Computational Landscape Genetics Group
- Joost
- Widmer
- Leempoel
- Duruz
- Rochat
- Cesconeto

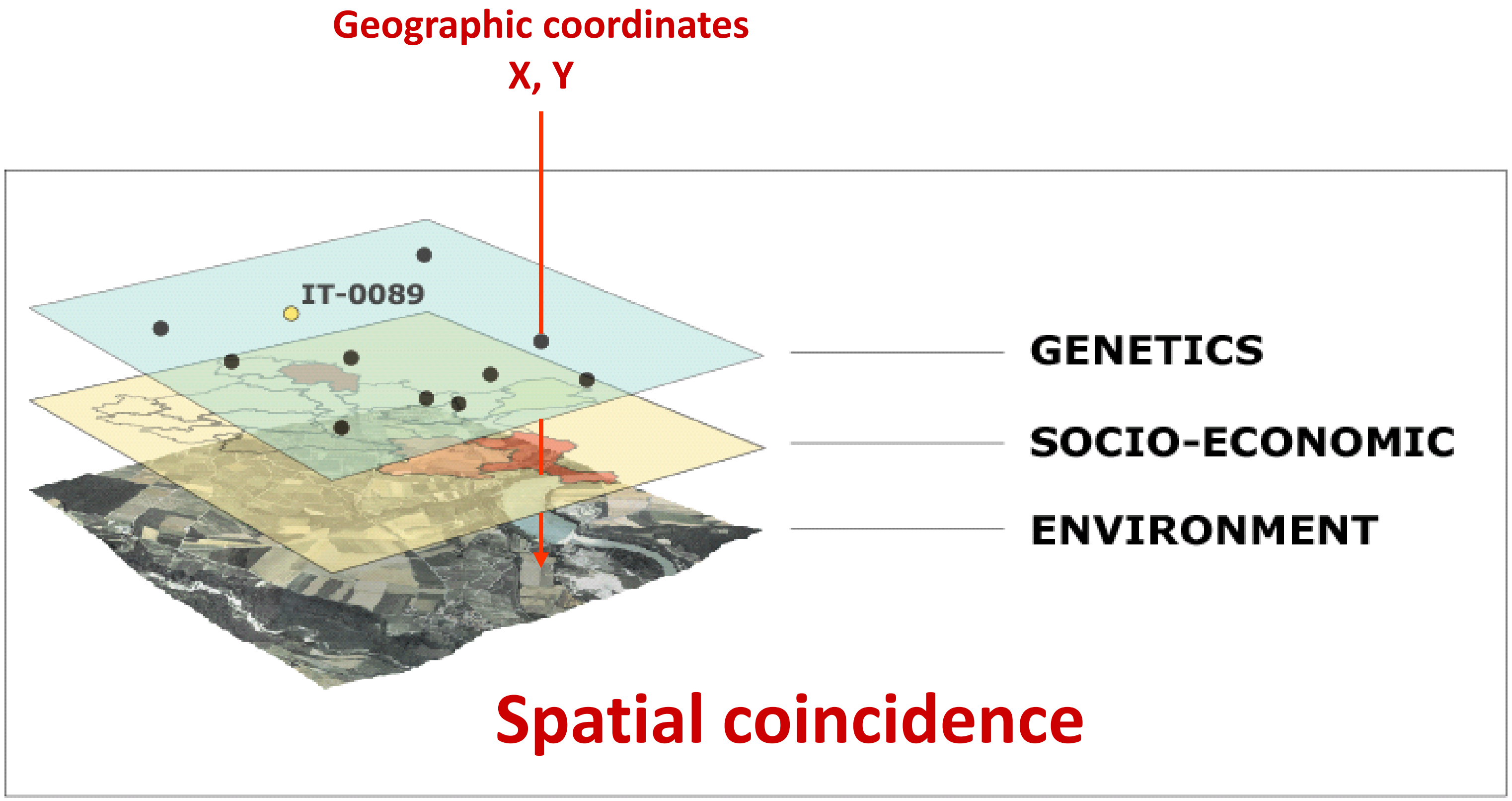
Research goals

- Make use of GIScience methods and geodata to contribute to :
  - The elaboration of efficient decision-making support approaches to favor the conservation of plant/animal genetic diversity
  - The advancement of our understanding of mechanisms controlling the evolution of species (adaptation to local environment in particular)

GIScience as insider

- Since 2001, immersion in the evolutionary biology community
- Close collaboration with biologists, geneticists, vets, etc.
- **Transdisciplinarity** vs interdisciplinarity
- A real appropriation of the research field
- Members of the team are biologists
- Projects submitted to life science divisions

A key function



Conservation and monitoring of Farm animal Genetic Resources

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

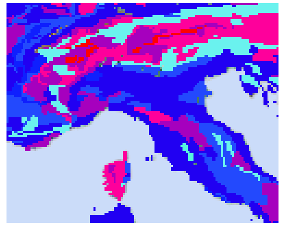
Data categories

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

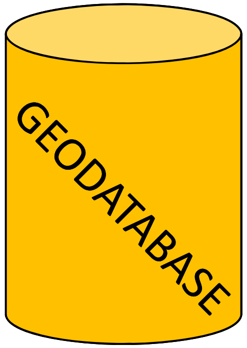
Data integration

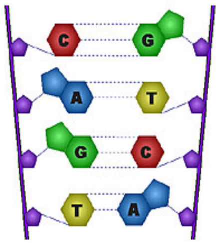
Administrative boundaries
Socio-Economic data
Socio-demographic data

Environmental data:
topography, climate,
soil, etc.



Geographic coordinates
X, Y





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).

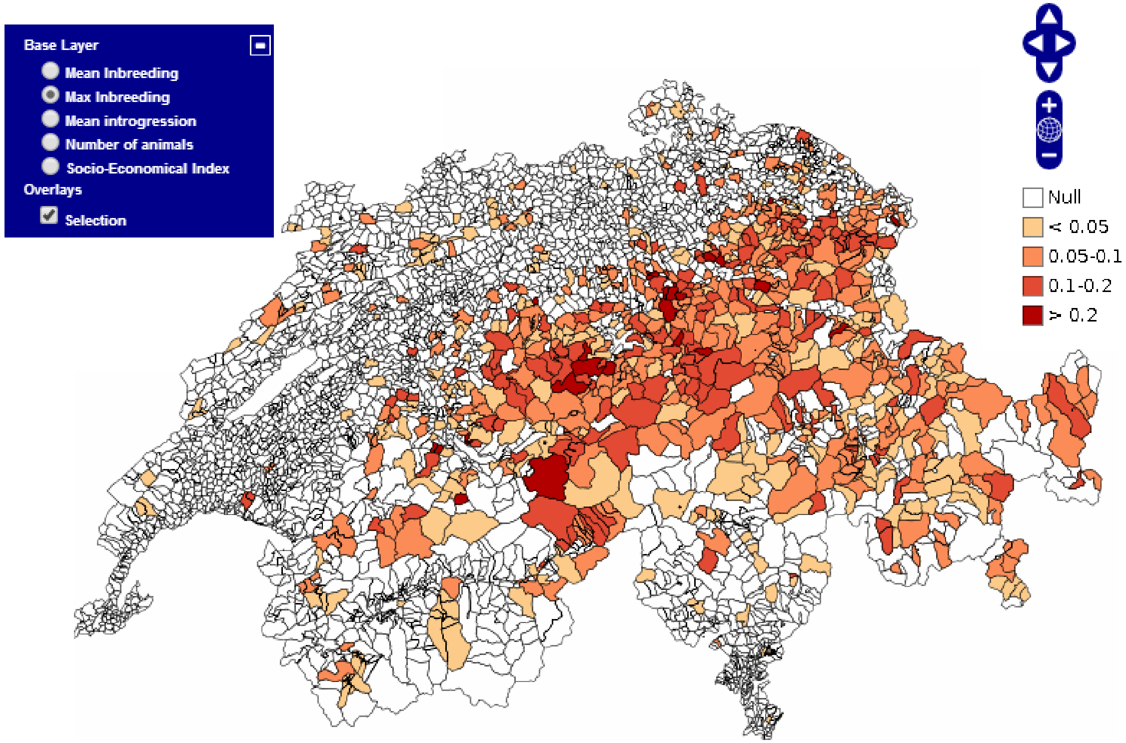
This application is designed to rank breeds according to four criteria: Genetic diversity (estimated from pedigree data), Introgression, Geographical concentration and Agriculture sustainability (from Socio

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 individuals 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



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 ($Ne_{\Delta Fp}$) and pedigree completeness

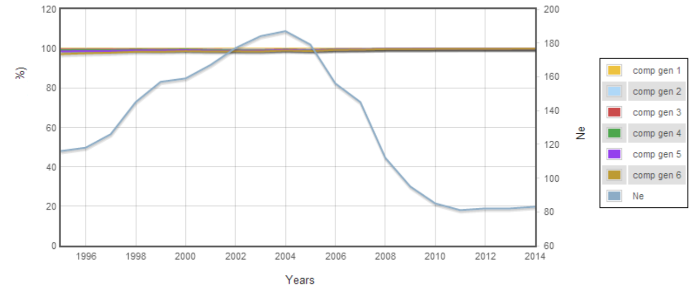
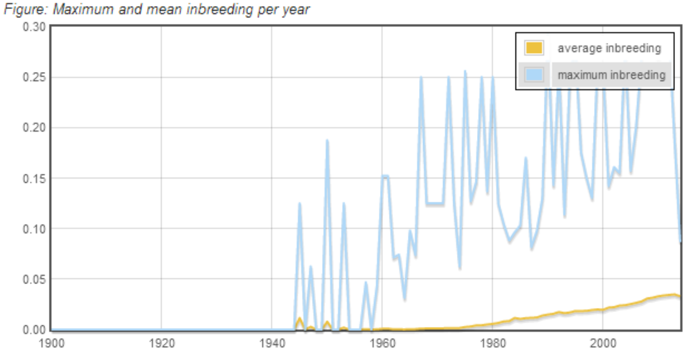


Figure: Maximum and mean inbreeding per year



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

Understanding of mechanisms controlling the evolution of species

PERSPECTIVES

Science, 2010

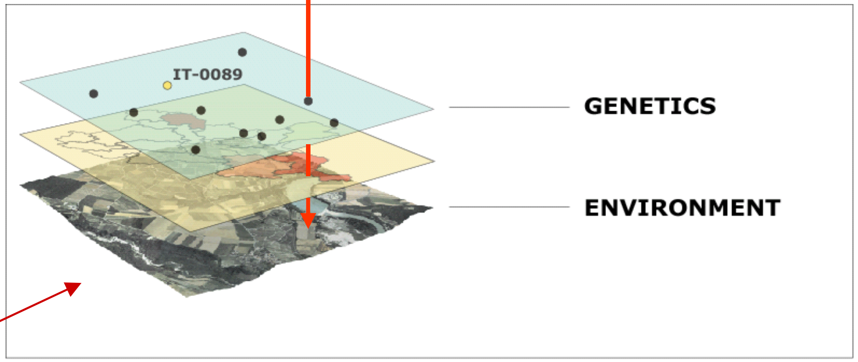
ECOLOGY

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

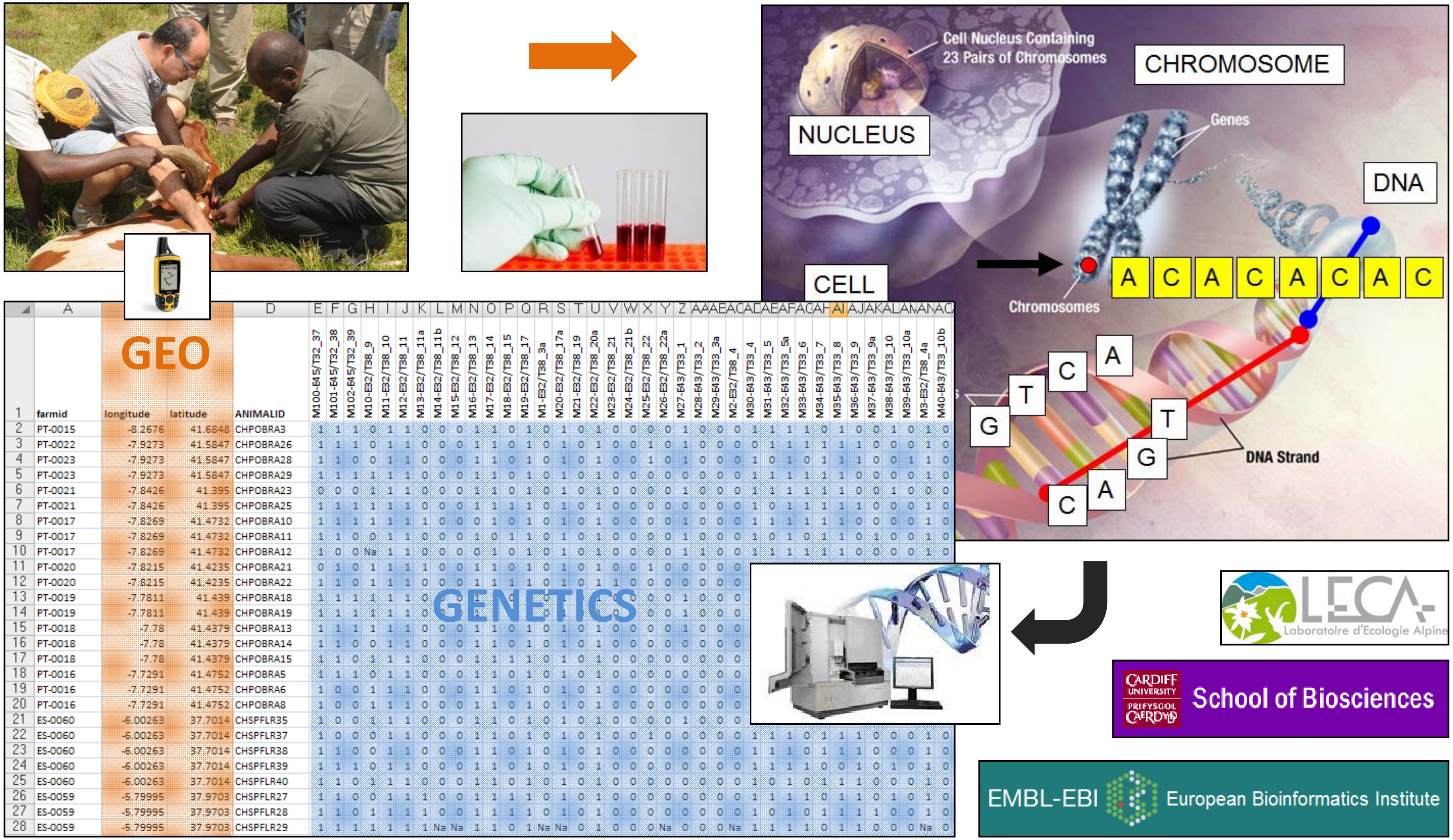


- Correlative approaches and spatial statistics

Genome-wide information

- Paradigm shift and transition phase

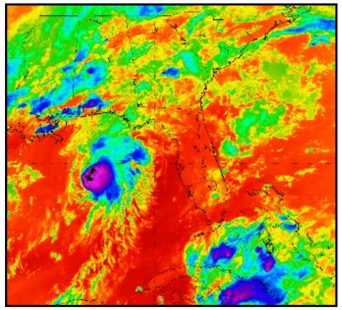
Genetic data

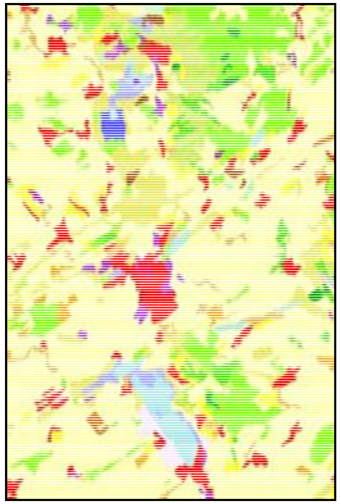


Environmental data characterizing sampling locations

GEO | **GENETICS** | **ENVIRONMENT**

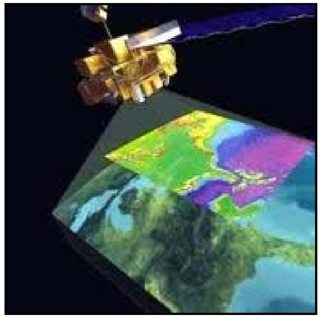
1	farmid	animalid	OARJMP29_allele2_137	OARJMP29_allele2_139	OARJMP29_allele2_141	OARJMP29_allele2_143	OARJM e2_145	OARJM e2_147	OARJM e2_149	OARJM e2_151	OARJM e2_153	OARJM e2_155	OARJM e2_157	OARJMP29_allele2_159	OARJMP29_allele2_161	OARJMP29_allele2_163	OARJMP29_allele2_165	OARJMP29_allele2_167	wndjan	altitude	wndfeb	wndmar	wndapr
1044	PL-4005	OAPLPOM25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		5.1	22	4.6	5	4.4
1045	PL-4005	OAPLPOM26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		5.1	22	4.6	5	4.4
1046	PL-4006	OAPLPOM01	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0		5.3	153	4.8	4.9	4.3
1047	PL-4006	OAPLPOM15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		5.3	153	4.8	4.9	4.3
1048	PL-4006	OAPLPOM24	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		5.3	153	4.8	4.9	4.3
1049	PL-4007	OAPLPOM05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		5.3	250	4.8	5	4.5
1050	PL-4007	OAPLPOM16	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		5.3	250	4.8	5	4.5
1051	PL-4008	OAPLPOM09	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0		5.2	166	4.8	5	4.4
1052	PL-4008	OAPLPOM19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		5.2	166	4.8	5	4.4
1053	PL-4008	OAPLPOM20	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0		5.2	166	4.8	5	4.4
1054	PL-4009	OAPLPOM10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		5.5	87	5	5.2	4.6
1055	PL-4009	OAPLPOM21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		5.5	87	5	5.2	4.6
1056	PL-4010	OAPLPOM08	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0		5.4	208	4.9	5.1	4.5





esa

European Environment Agency



CRU

NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE

Logistic regression

Individuals | **Genetic markers** | **Environmental variables**

1	farmid	animalid	OARJMP29_allele2_13	OARJMP29_allele2_13	OARJMP29_allele2_14	OARJMP29_allele2_14	OARJMP29_allele2_14	OARJMP29_allele2_14	OARJMP29_allele2_14	OARJMP29_allele2_15	OARJMP29_allele2_15	OARJMP29_allele2_15	OARJMP29_allele2_15	OARJMP29_allele2_159	OARJMP29_allele2_161	OARJMP29_allele2_163	OARJMP29_allele2_165	OARJMP29_allele2_167	wndjan	altitude	wndfeb	wndmar	wndapr
1044	PL-4005	OAPLPOM25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.1	22	4.6	5	4.4
1045	PL-4005	OAPLPOM26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.1	22	4.6	5	4.4
1046	PL-4006	OAPLPOM01	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	5.3	153	4.8	4.9	4.3
1047	PL-4006	OAPLPOM15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.3	153	4.8	4.9	4.3
1048	PL-4006	OAPLPOM24	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	5.3	153	4.8	4.9	4.3
1049	PL-4007	OAPLPOM05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.3	250	4.8	5	4.5
1050	PL-4007	OAPLPOM16	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	5.3	250	4.8	5	4.5
1051	PL-4008	OAPLPOM09	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	5.2	166	4.8	5	4.4
1052	PL-4008	OAPLPOM19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.2	166	4.8	5	4.4
1053	PL-4008	OAPLPOM20	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	5.2	166	4.8	5	4.4
1054	PL-4009	OAPLPOM10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.5	87	5	5.2	4.6
1055	PL-4009	OAPLPOM21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.5	87	5	5.2	4.6
1056	PL-4010	OAPLPOM08	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	5.4	208	4.9	5.1	4.5

Multiple parallel logistic regressions

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]

Cornell University Library
arXiv.org > q-bio > arXiv:1405.7658
Quantitative Biology > Populations and Evolution

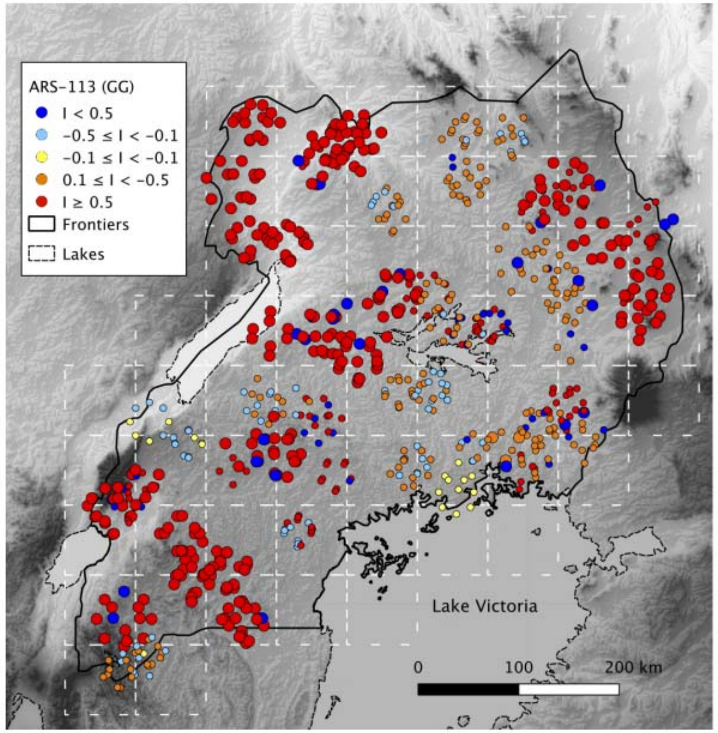
http://lasig.epfl.ch/**sambada**

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



(a) ARS-113 (GG)

MBE Advance Access published October 13, 2014

Adaptations to Climate-Mediated Selective Pressures in Sheep

Feng-Hua Lv,[1] Saif Agha,[2,3] Juha Kantanen,[4,5] Licia Colli,[6,7] Sylvie Stucki,[2] James W. Kijas,[8,] Stéphane Joost,[2] Meng-Hua Li,*[,1] and Paolo Ajmone Marsan[6,7]

[1]CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences (CAS), Beijing, China

[2]Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[3]Department of Animal Science, Faculty of Agriculture, Ain Shams University, Cairo, Egypt,

[4]Biotechnology and Food Research, MTT Agrifood Research Finland, Jokioinen, Finland

[5]Department of Biology, University of Eastern Finland, Kuopio, Finland

[6]Istituto di Zootecnica, Facoltà di Agraria, Università Cattolica del Sacro Cuore, Piacenza, Italy

[7]Biodiversity and Ancient DNA Research Center—BioDNA, Università Cattolica del Sacro Cuore, Piacenza, Italy

[8]CSIRO Livestock Industries, St Lucia, Brisbane, Qld, Australia

***Corresponding author:** E-mail: menghua.li@ioz.ac.cn.

Associate editor: Yuseob Kim

anisms likely underlying environmental adaptation in domestic animals.

Impact of Climate on Energy Metabolism and Endocrine and Autoimmune Regulation

Our results suggest that the process of autochthonous sheep breeds adaptation to extreme climates is principally mediated by complex, integrated energy metabolic responses, as observed in rodents (Rezende et al. 2004). Climate is known to have an important impact on animal physiology and fit-

McManus et al. 2011). Consequently, long-term thermal stress can result in energy metabolic adaptation, as well as heat and cold tolerance, in particular breeds. Meanwhile, variation in animal morphology, including body size (large vs. small) and shape (fat-tailed vs. thin-/short-tailed), also follows basic thermoregulatory principles to dissipate or conserve body energy in different climates. In addition, several breeds (e.g., Norwegian White Sheep) grow particularly large in cold regions that are dry/warm in early spring and warm/wet in late summer, where conditions favor continued grass growth (Nielsen et al. 2013); however, several breeds grow small in the

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

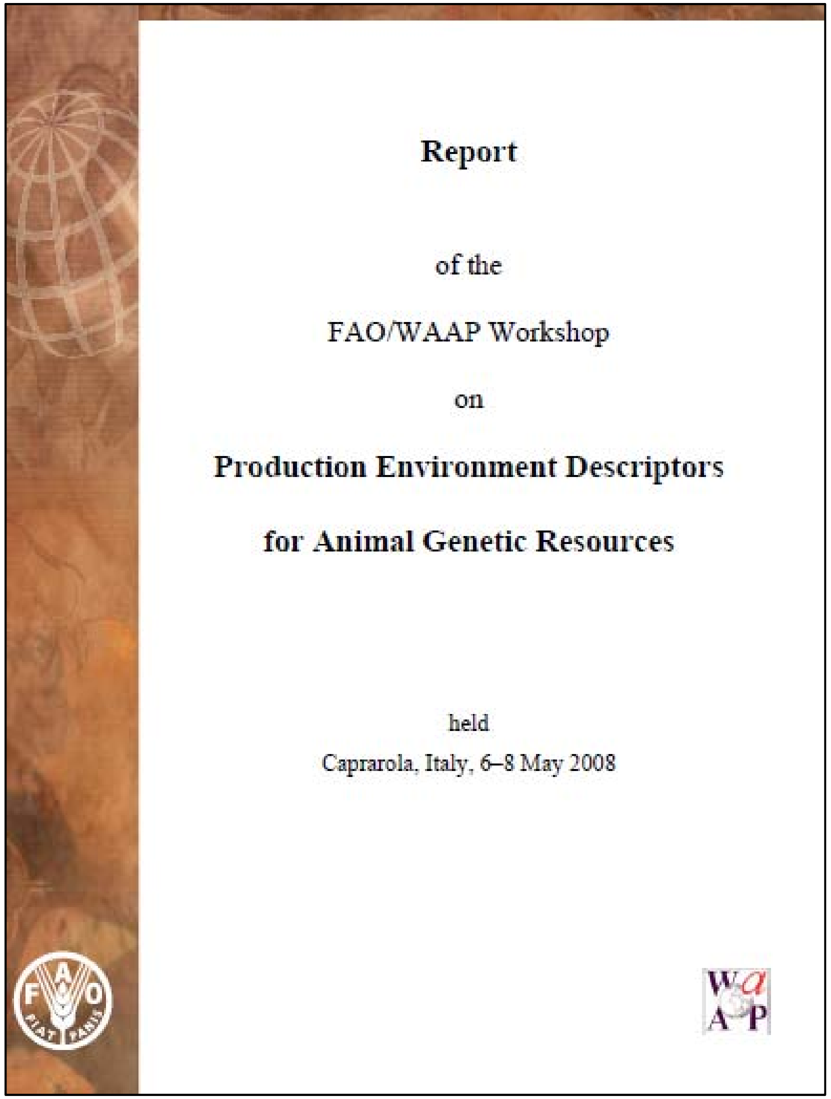
- Computer science

Limit

- An 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

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

X,Y

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.

Conclusion

- 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 (large molecular, environmental, and socio-economic datasets)
- … only if we train a new generation of students/scientists able to develop innovative **transdisciplinary** and efficient geocomputing tools

Thank you for your attention !