Innovative Management of Animal Genetic Resources (IMAGE): investigating local adaptation in European sheep through landscape genomics

*Licia Colli[1], Oliver Selmoni[2], Mario Barbato[1], Marcello Del Corvo[1], Elia Vajana[2], Elisa Eufemi[1], Badr Benjelloun[3], Stéphane Joost[2], **Paolo Ajmone-Marsan[1]** & the IMAGE consortium[4]*

1 Department of Animal Science, Food and Nutrition (DIANA), Biodiversity and Ancient DNA Research Centre (BioDNA), Università Cattolica del Sacro Cuore, Piacenza, Italy
2 Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
3 Institut National de la Recherche Agronomique Maroc (INRA-Moroc), Centre Régional de Beni Mellal, Beni Mellal, Morocco
4 INRA, Animal Genetics and Integrative Biology Unit, Jouy-en-Josas, France

EAAP Ghent, August 28, 2019

IMAGE project

IMAGE
European Project n° 677353

Innovative Management of Animal Genetic rEsources 'IMAGE' project (imageh2020.eu)

selected under the H2020 call :

SFS-7-2014/2015: Genetic resources and agricultural diversity for food security, productivity and resilience

topic B. [2015]
Management and sustainable use of genetic resources

Michèle
Tixier Boichard

IMAGE project

Objectives of IMAGE

➢ **to enhance the use of genetic collections**

➢ **to upgrade animal gene bank management**

➢ **to demonstrate the benefits brought by gene banks:**

- ▪ **To minimise genetic accidents**

- ▪ **To allow the livestock sector to respond to new environmental constraints and market needs**

through genomics, bio-informatics, reproductive biotechnologies

28 partners, March 2016 to February 2020 ; 7 M€
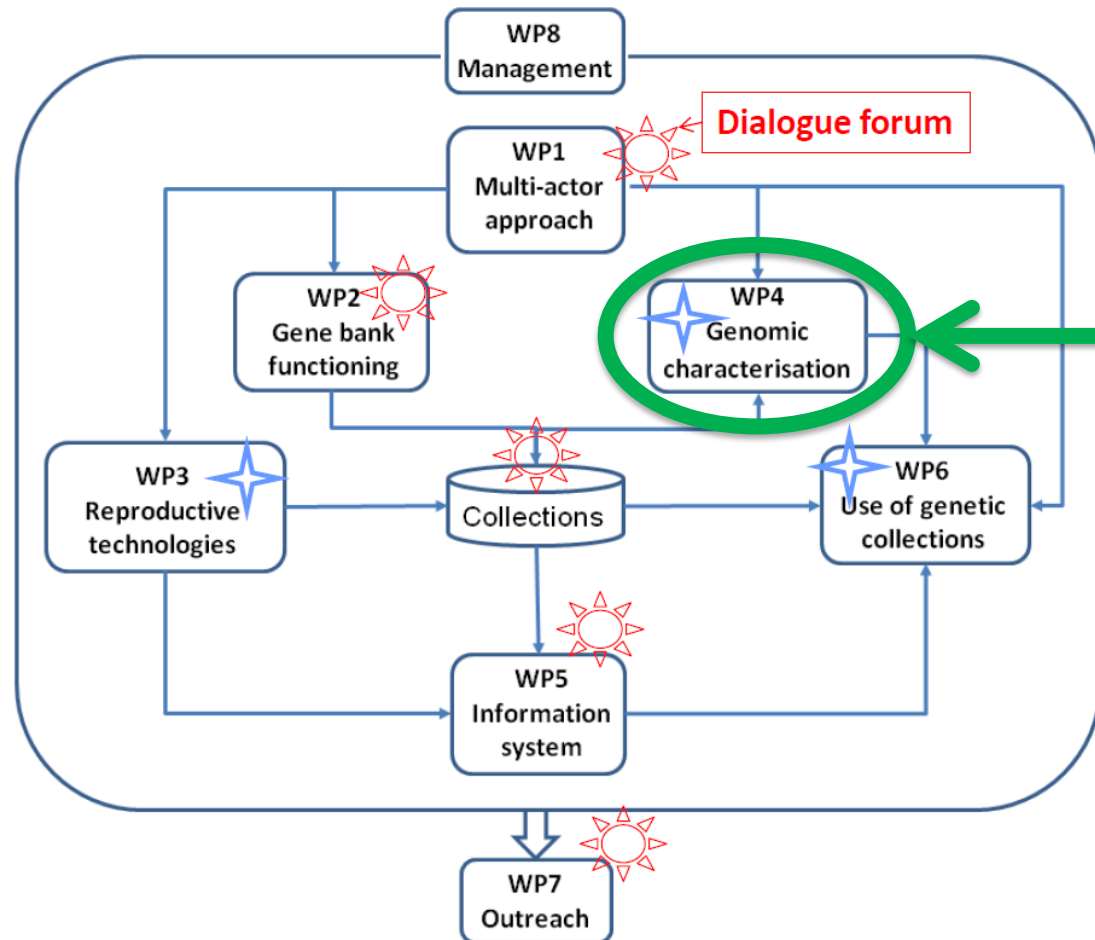
IMAGE project

UNIVERSITÀ CATTOLICA del Sacro Cuore

IMAGE — Innovative Management of Animal Genetic Resources

HORIZON 2020 — THE FRAMEWORK PROGRAMME FOR RESEARCH AND INNOVATION

WP structure

All species

Selected species for case studies

WP8 Management

Dialogue forum

WP1 Multi-actor approach

WP2 Gene bank functioning

WP4 Genomic characterisation

WP3 Reproductive technologies

Collections

WP6 Use of genetic collections

WP5 Information system

WP7 Outreach

UCSC EPFL Adaptation in sheep

The origin and dispersal of domestic livestock species in the Fertile Crescent

Melinda A. Zeder PNAS 2008;105:33:11597-11604

Adaptation
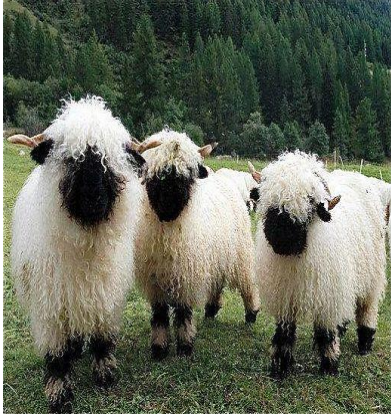
Following domestication sheep breeds have adapted to very different and harsh environments, spanning mountain to desert areas and extreme conditions in terms of humidity, temperature and elevation (FAO, 2015)
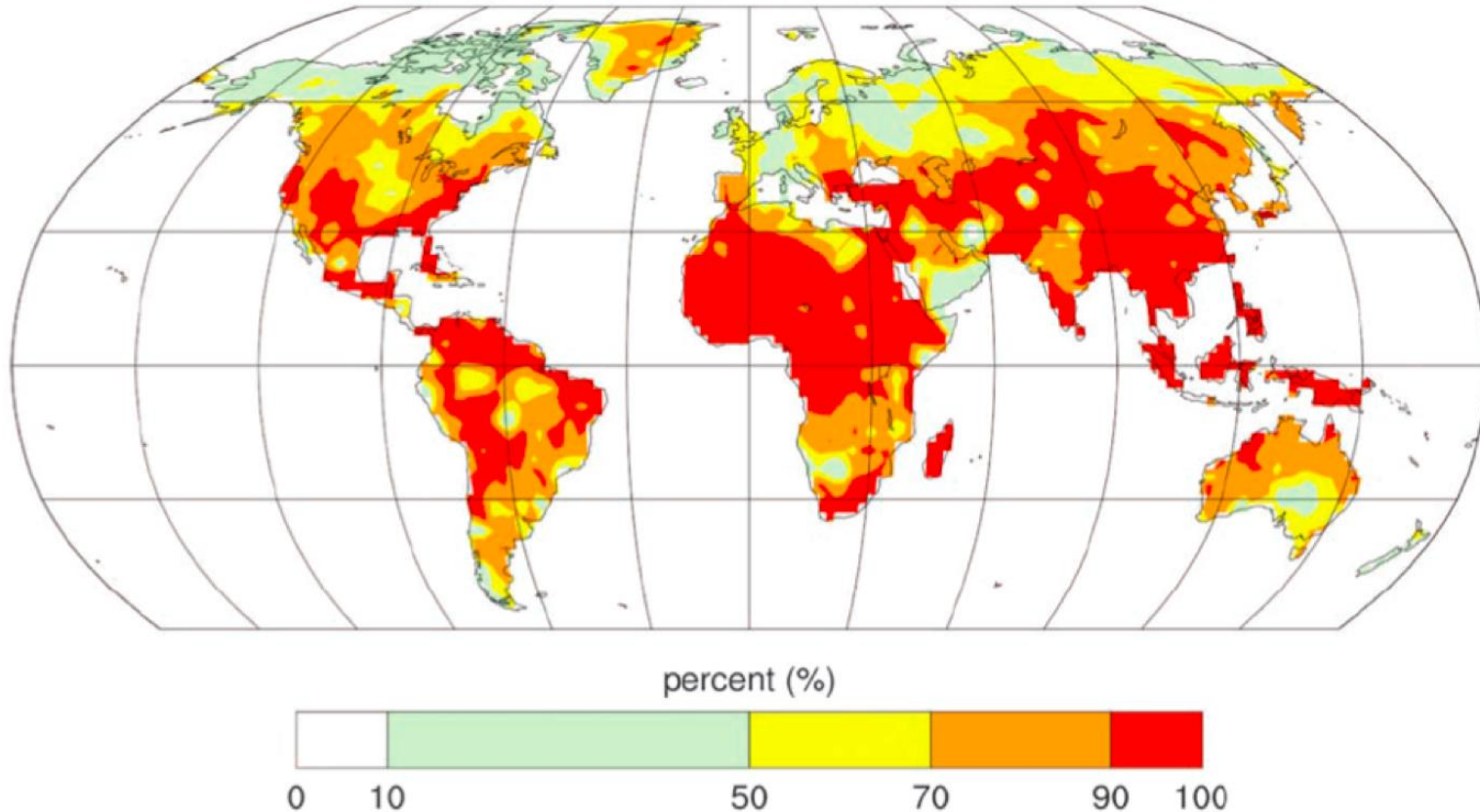
Summers in 2080-2100 Warmer than Warmest on Record

Likelihood (in percent) that future average summer temperatures will exceed the highest summer temperature observed on record for 2090. From Science, 2009, 323: 240.

Landscape genomics

Goal: understanding **molecular mechanisms underlying livestock adaptation** to **environmental challenges**

Integrated coanalysis
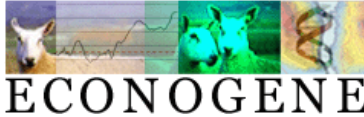
Environmental data

Landscape genomics

Genomic data
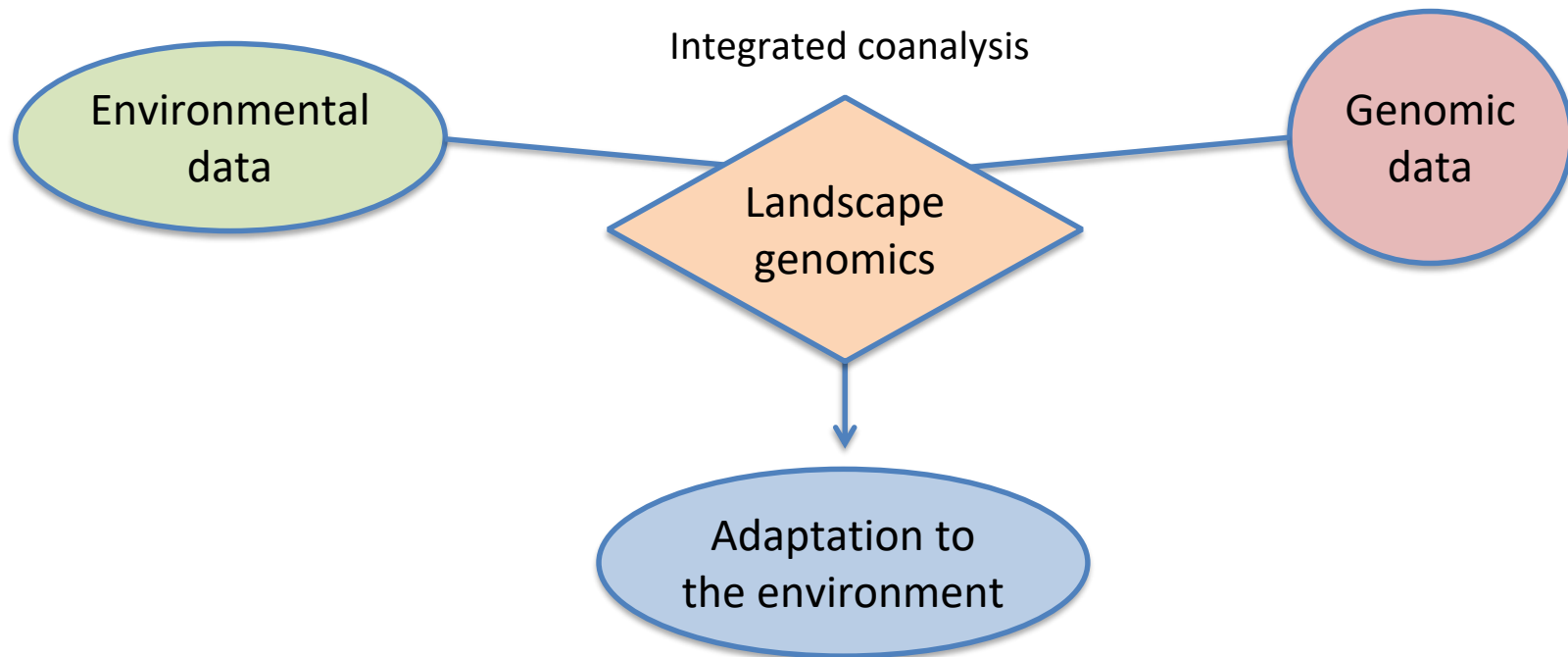
Adaptation to the environment
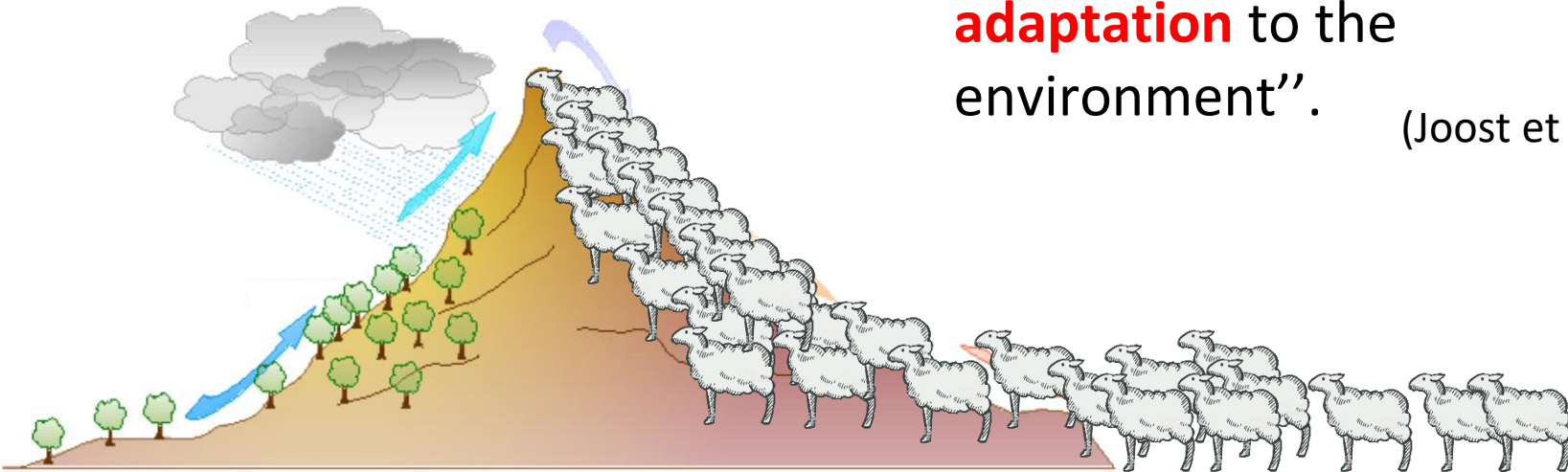
The Landscape genomics approach

Stephane Joost

Mike Bruford

Landscape genomics
''attempts to explain the pattern of spatial genetic variation in relation to **local adaptation** to the environment''.
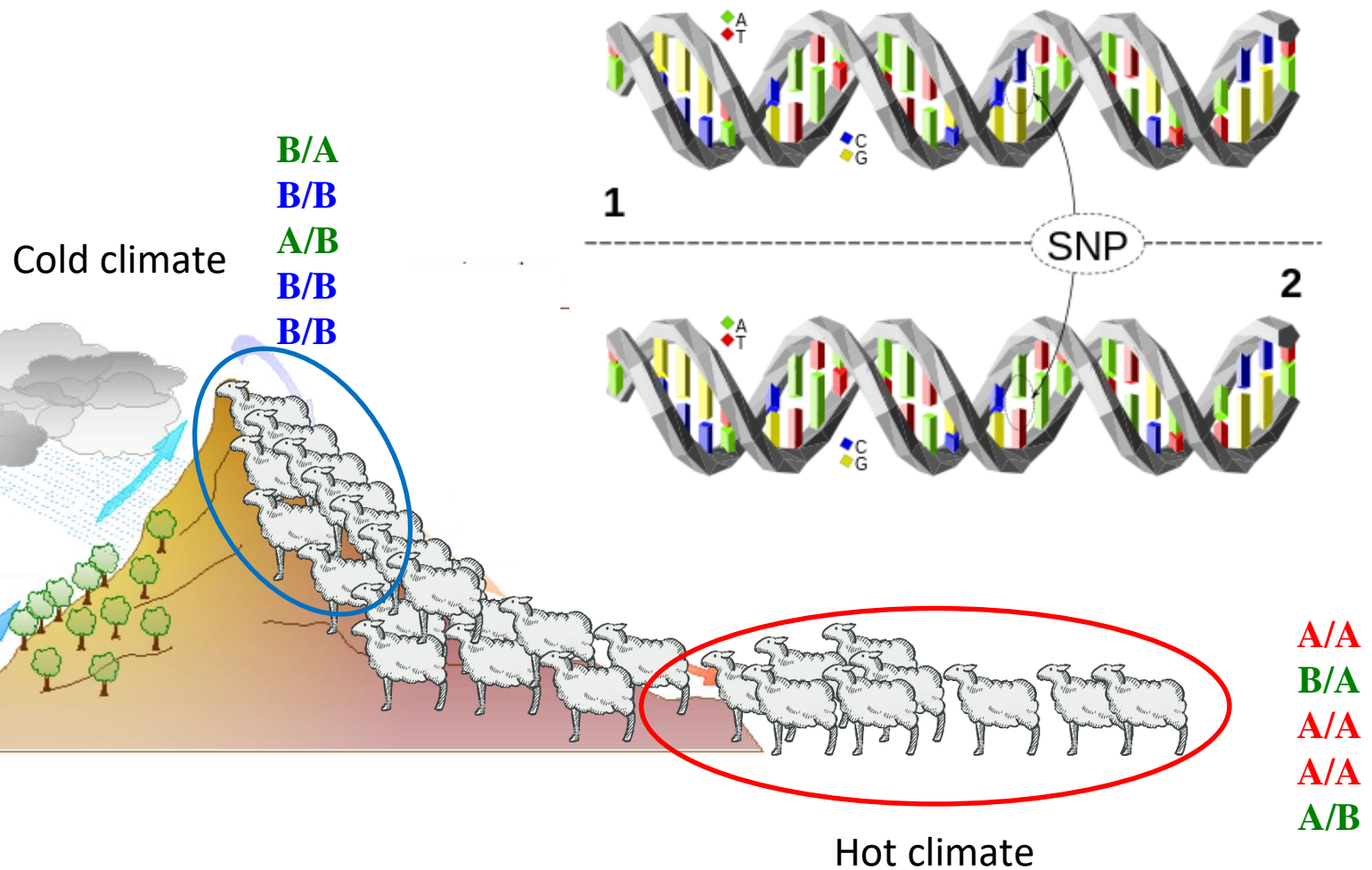
(Joost et al., 2013)

The Landscape genomics approach

IMAGE sheep landscape genomics study

```
┌─────────────────┐         ┌──────────────────────┐         ┌──────────────────┐
│  50x50 Km grid  │ - - - → │ 19 bioclimatic       │ - - - → │ Cluster analysis │
│                 │         │ variables            │         │                  │
└─────────────────┘         │ + Elevation          │         └──────────────────┘
                            └──────────────────────┘
```

8 Main clusters
32 subclusters
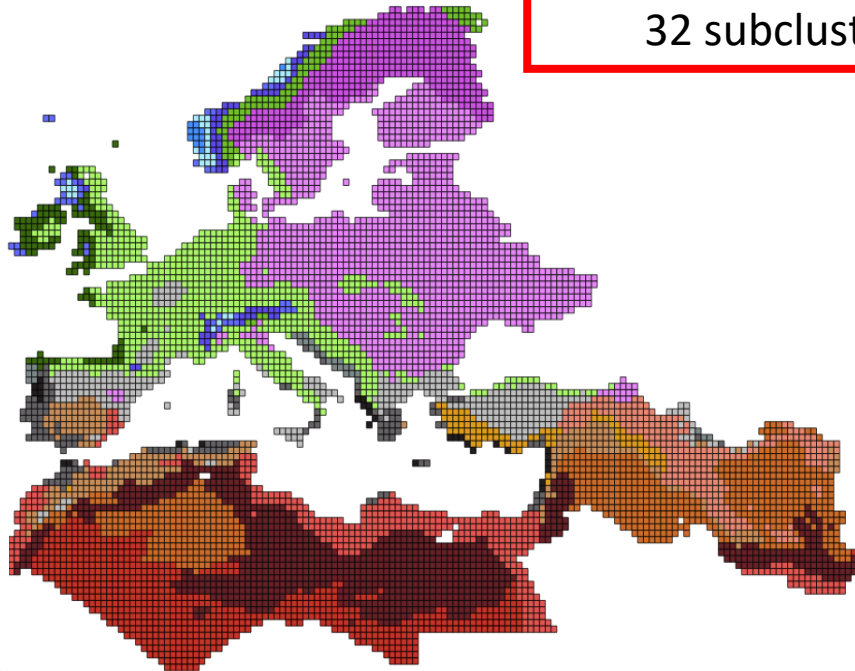
BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (BIO2/BIO7) (* 100)
BIO4 = Temperature Seasonality (standard deviation *100)
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of Coldest Month
BIO7 = Temperature Annual Range (BIO5-BIO6)
BIO8 = Mean Temperature of Wettest Quarter
BIO9 = Mean Temperature of Driest Quarter
BIO10 = Mean Temperature of Warmest Quarter
BIO11 = Mean Temperature of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month
BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter

UNIVERSITÀ CATTOLICA del Sacro Cuore

Oliver Selmoni (LASIG)

Selection of sheep samples for HD genotyping and landscape genomics analysis (Task 4.4) and whole-genome sequencing (Task 4.2)
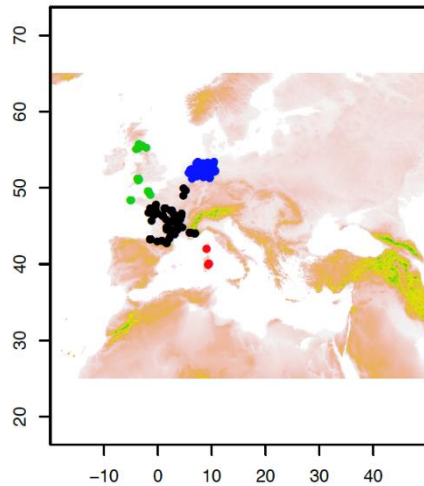
- 706 georeferenced sheep having HD SNP data were retrieved from previous projects of IMAGE Partners and assigned to the 8 environmental clusters and 32 subclusters;

- Additional 965 geo-referenced biological samples/DNA were made available by IMAGE partners;

- among them, 672 were selected for HD genotyping to ensure both the representativeness of the environmental sub-clusters and the within-breed variation.

- 57 animals (one per breed for all breeds having sufficient DNA quality/amount) have been fully sequenced 12X (other 43 will follow, chosen on the basis of landscape genomics results)

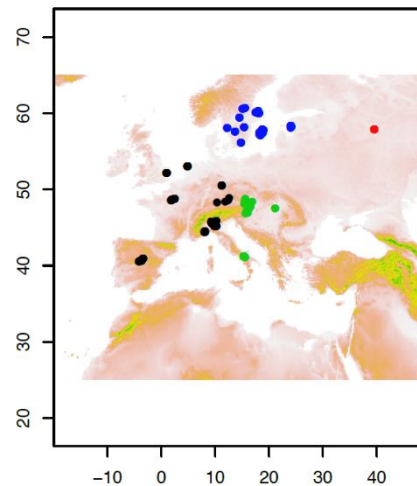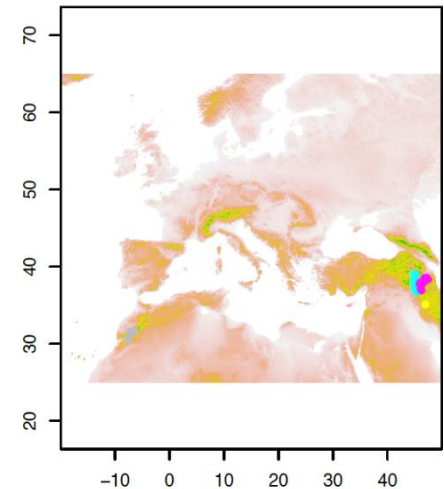Samples distribution among environmental clusters/subclusters

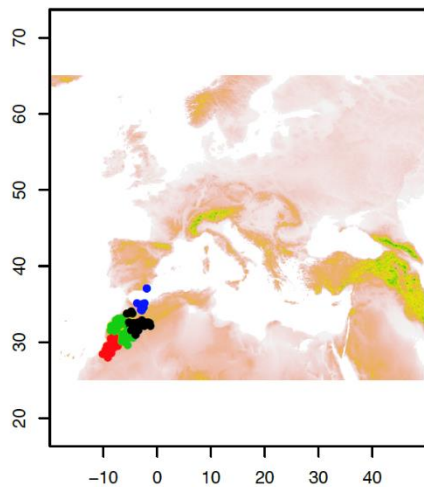1 Continental - humid
2: Mediterranean - dry
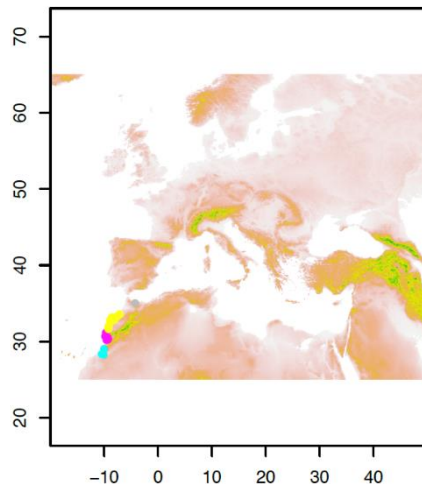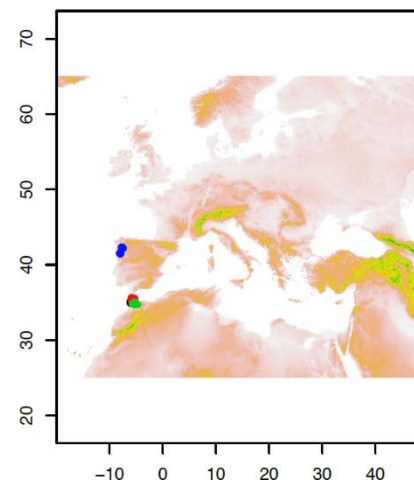3: Continental - cold
4: Continental - arid
5: Desert
6: Atlantic - dry
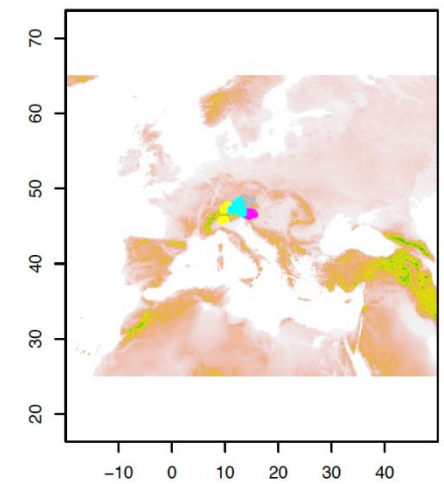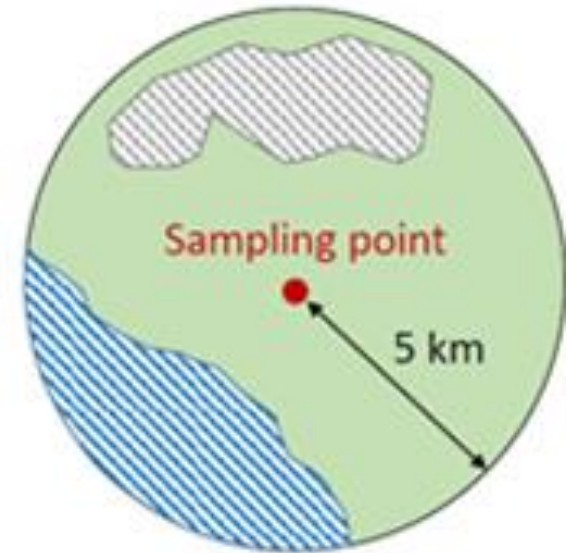7: Atlantic - wet
8: Alpine
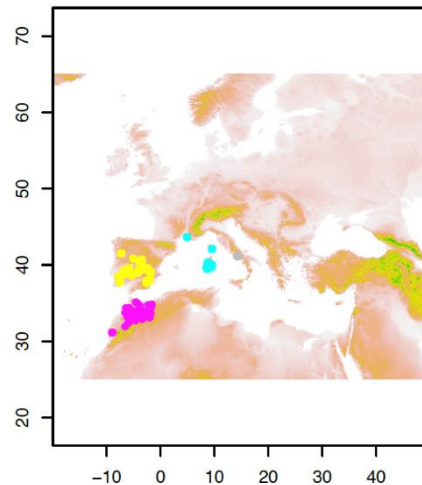
- Considered **a buffer area** with a **radius of 5 km** centered on the sampling point.

- Land cover discrimination was applied to **remove artificial surfaces** (urban land cover) **and water bodies** from the buffer area.
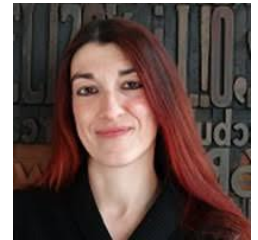

Sampling point

5 km

▨ Artificial/Urban Area

▨ Water Area

▢ Area from which we retrieved the environmental variables and computed the statistics

HD SNP genotypes
- 71 domestic breeds from 12 Countries (Austria, Germany, Estonia, France, Iran, Italy, Morocco, Netherlands, Portugal, Spain, Sweden, United Kingdom)

Licia Colli

WGS
- Newly produced (N=57 → 1 animal per domestic breed with enough good quality DNA available)
- Additional N=43 to be produced on animals chosen following the first round of landscape genomics analyses

Dataset and QC

Raw dataset

1431 animals (74 breeds)

600,000 SNPs

Quality Check

Step 1: SNPs with >2% missing data removed

Step 2: Animals with >2% missing data removed

Step 3: SNPs with MAF<1% removed

PCA: pruned for LD>0.2 and N

166,004 SNPs, 1,216 individuals
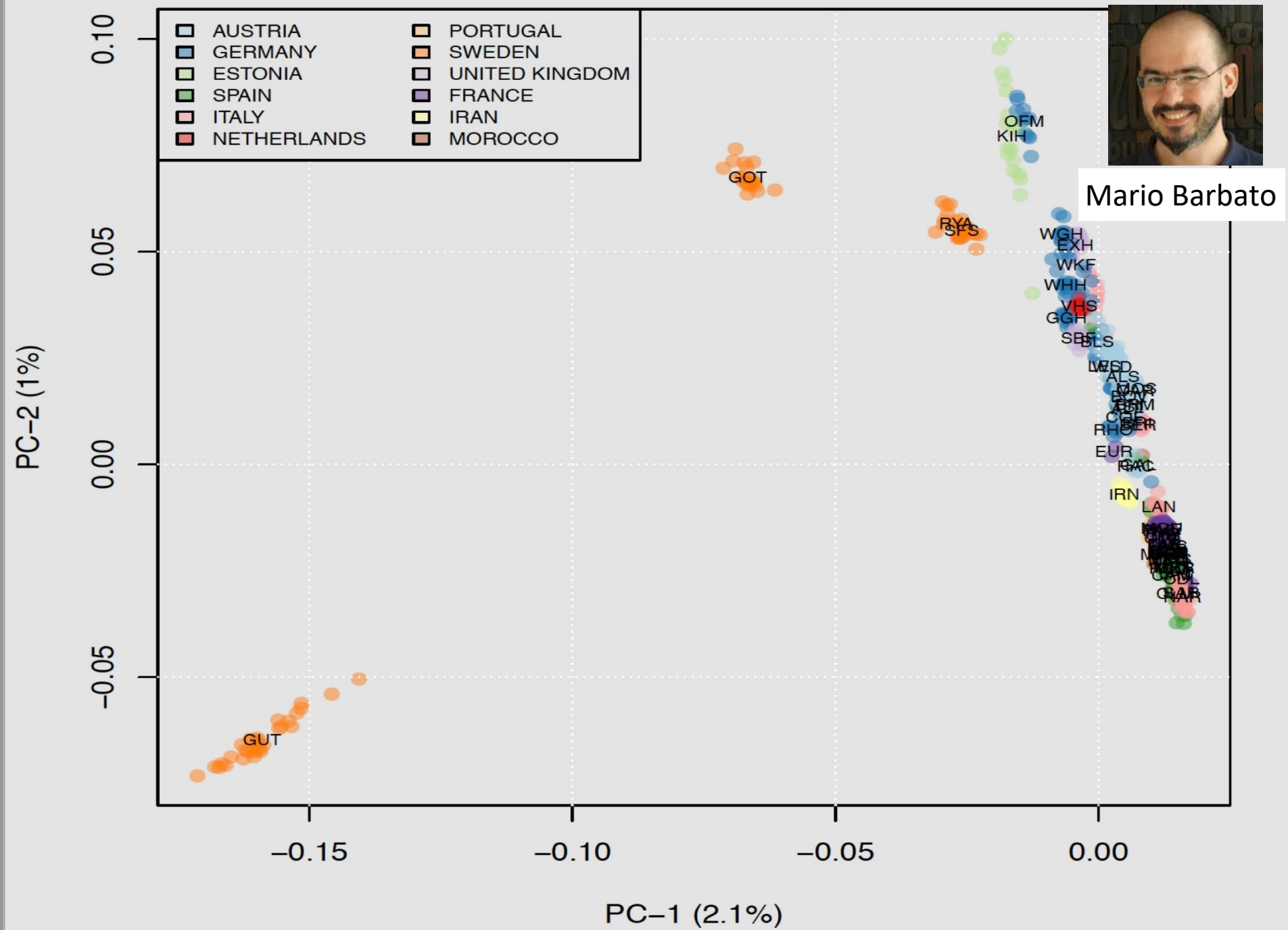
SamBada: pruned for IBD>0.18
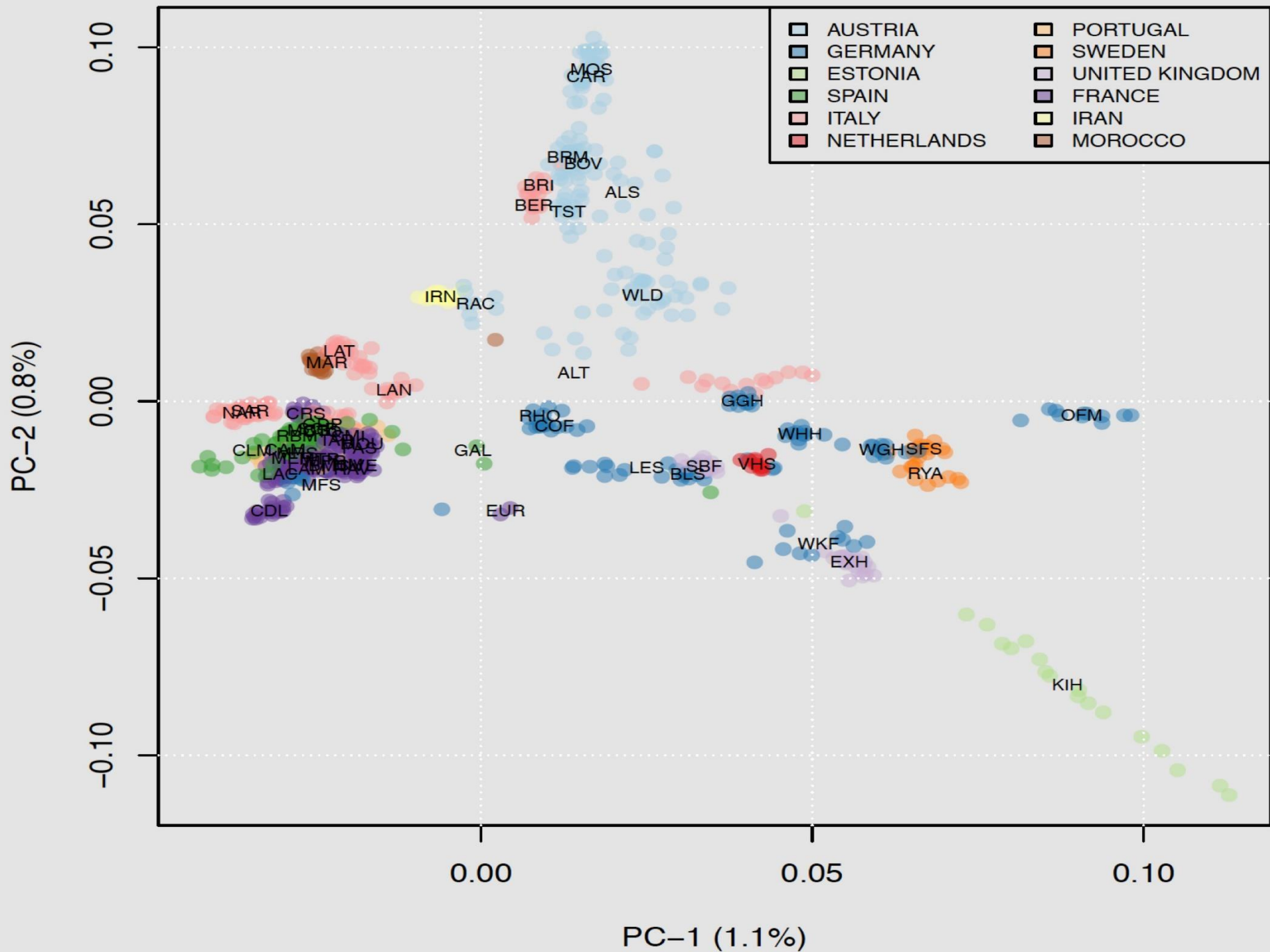
428,463 SNPs, 821 individuals

Fst: pruned for LD>0.5

322,732 SNPs, 1,216 individuals

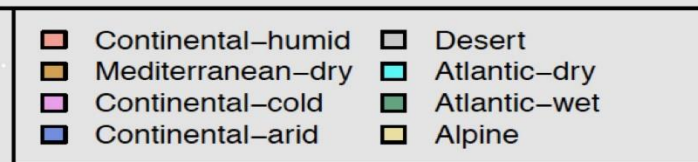**PCAdapt: pruned for LD>0.2**

511,623 SNPs, 1,399 individuals

Cleaned dataset
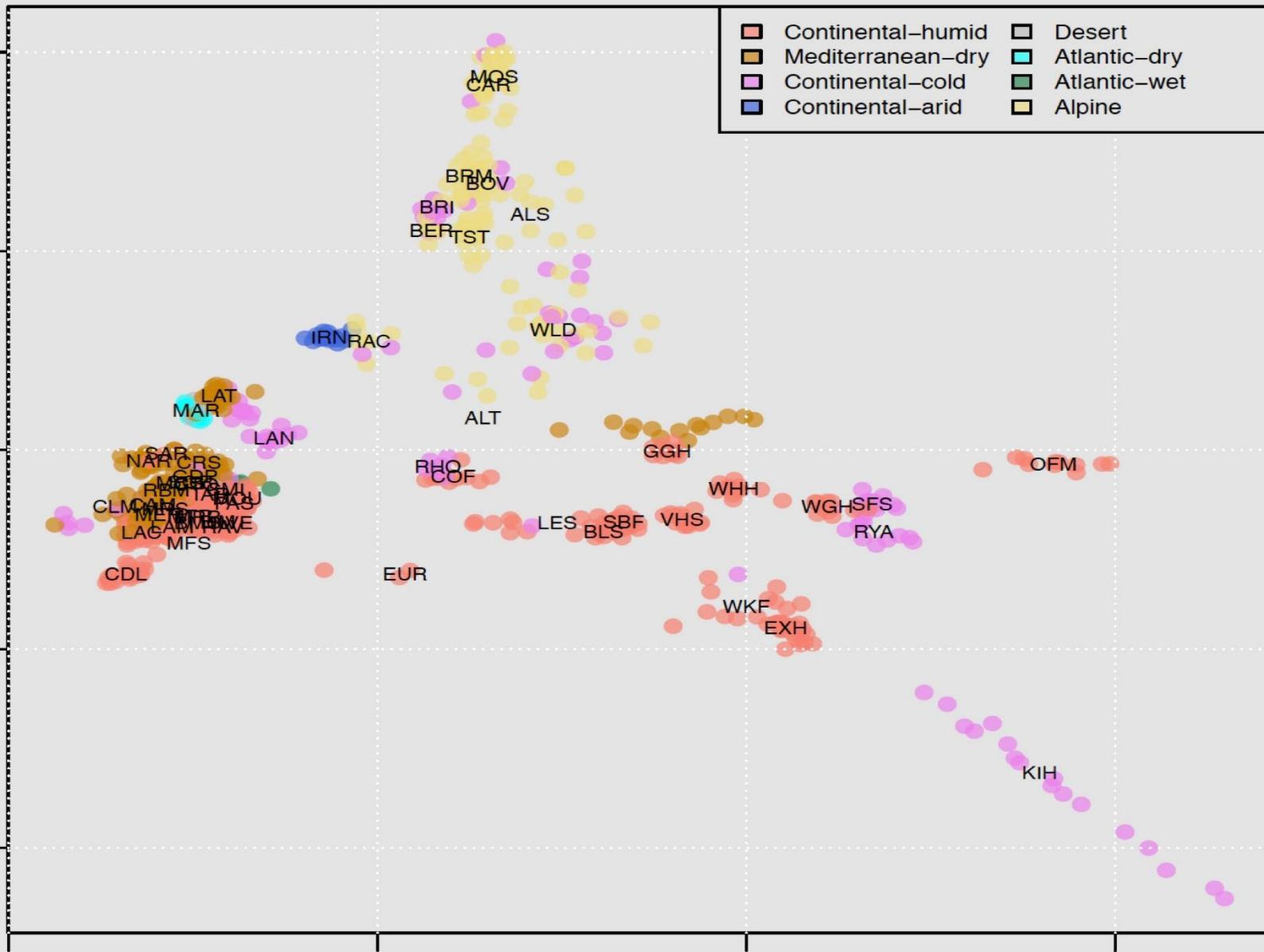
519,169 SNPs

1,399 individuals

Call rate 0,9992

Mario Barbato

Given a specific genotype, NM was specified as:

$$ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{s=1}^{S} \beta_s x_{si}$$

Likelihood of having the genotype at sampling site considering 6 variables accounting for population structure

where β_0 represents the model intercept, β_s the regression coefficient related to the *s-th* population structure predictor, and x_{si} the *i-th* observation of the *s-th* population structure predictor.

AM was specified as:

$$ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_Z z_i + \sum_{s=1}^{S} \beta_s x_{si}$$

Likelihood of having the genotype at sampling site considering 6 variables accounting for population structure + environmental variable

where β_Z is the regression coefficient for the environmental variable Z, and z_i is the *i-th* observation of Z.

This way, NM is nested within AM (i.e. NM=AM when β_Z=0). Then, a LR test was performed for each genotype to test if the inclusion of Z (i.e. in turn ψ_R or γ) led to a significantly improved explanation of the genotype spatial distribution. Particularly, as SAMβADA returns log-likelihood (LogL) values, the LR test was specified as:
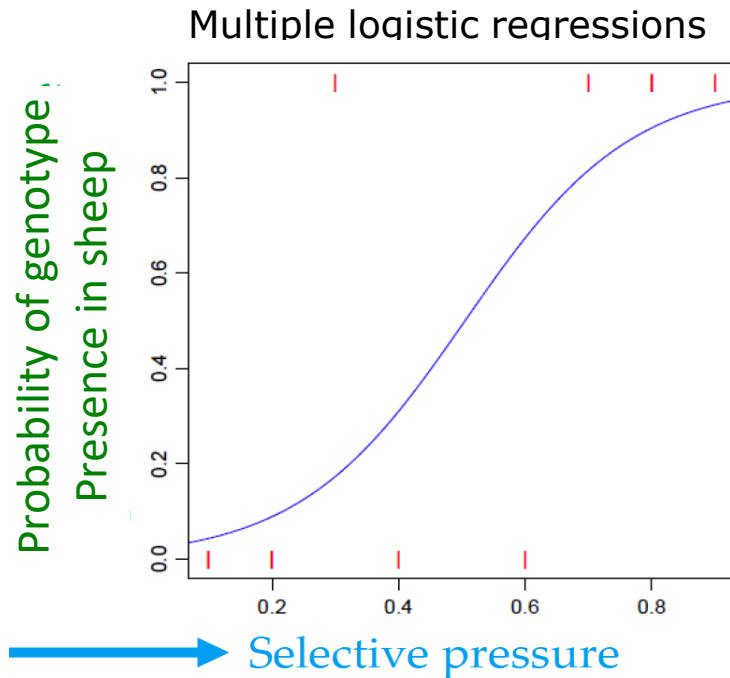
$$D = -2(\text{LogL NM} - \text{LogL AM})$$

Likelihood ratio test

- Software Samβada (Stucki et al., 2016);

Multiple logistic regressions

Probability of genotype Presence in sheep

Selective pressure

High performance computation of landscape genomic models including local indicators of spatial association

S. STUCKI,* P. OROZCO-TERWENGEL,† B. R. FORESTER,‡ S. DURUZ,* L. COLLI,§ C. MASEMBE,¶ R. NEGRINI,§,** E. LANDGUTH,†† M. R. JONES,†† THE NEXTGEN CONSORTIUM,‡‡ M. W. BRUFORD,† P. TABERLET§§,¶¶ and S. JOOST*

Rapid identification and interpretation of gene–environment associations using the new R.SamBada landscape genomics pipeline

Solange Duruz[1] | Natalia Sevane[2] | Oliver Selmoni[1] | Elia Vajana[1] | Kevin Leempoel[3] | Sylvie Stucki[1] | Pablo Orozco-terWengel[4] | Estelle Rochat[1] | Susana Dunner[2] | The NEXTGEN Consortium | The CLIMGEN Consortium | Michael W. Bruford[4] | Stéphane Joost[1]

Almost 130 million models to be considered

Landscape genomics

Stephane Joost

Oliver Selmoni

Elia Vajana

Parallelization…… LD pruning……. Analysis by cromosome…….

In the meantime………

(outlier) F$_{st}$
(across populations)

Luikart et al, 2003 Net Rev Genet
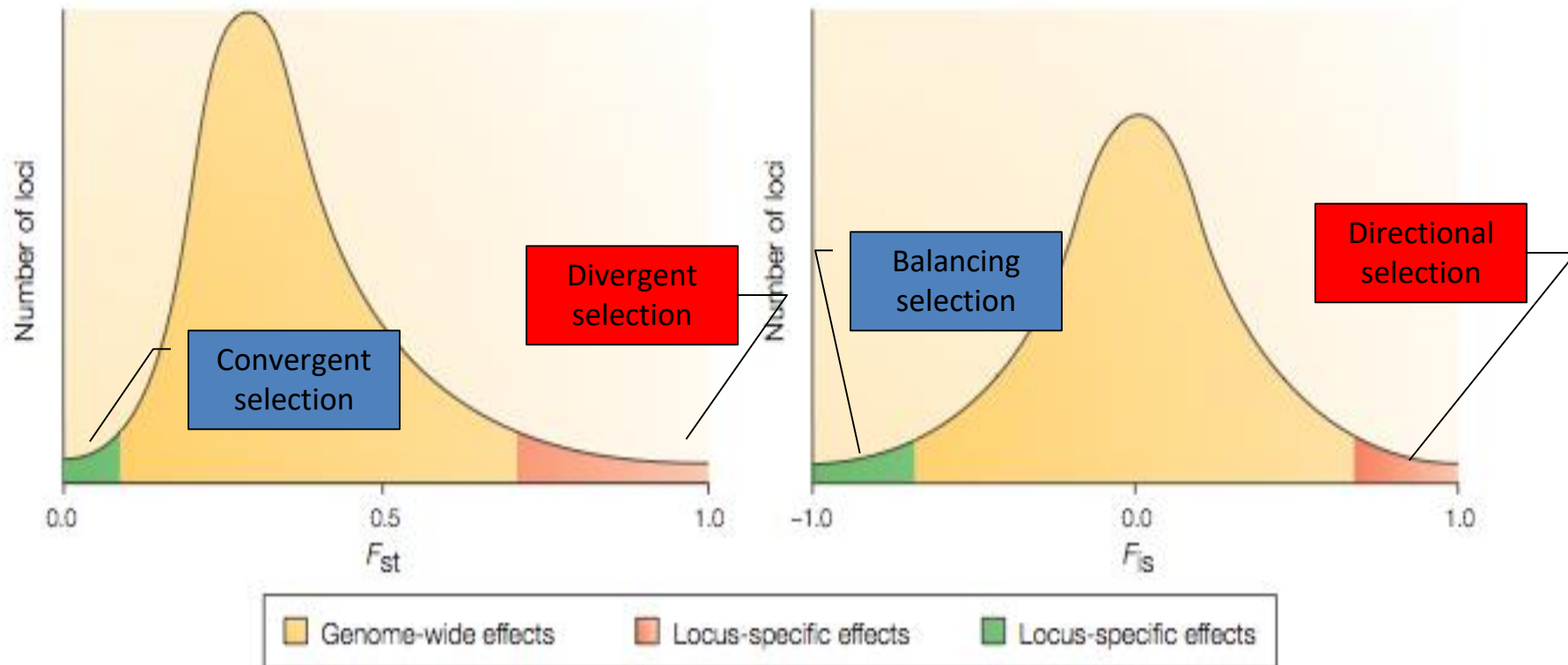
- Convergent selection
- Divergent selection
- Balancing selection
- Directional selection

Number of loci — F_{st}

Number of loci — F_{is}

☐ Genome-wide effects ☐ Locus-specific effects ☐ Locus-specific effects

PCAdapt

SPECIAL ISSUE: POPULATION GENOMICS WITH R

pcadapt: an R package to perform genome scans for selection based on principal component analysis

KEURCIEN LUU,* ERIC BAZIN† and MICHAEL G. B. BLUM*

*Laboratoire TIMC-IMAG, UMR 5525, CNRS, Université Grenoble Alpes, Grenoble, France, †Laboratoire d'Ecologie Alpine UMR 5553, CNRS, Université Grenoble Alpes, Grenoble, France

Markers outliers with respect to how they are related to population structure ascertained with principal component analysis, are candidate to be under selection

- Computed a vector containing K z-scores measures to what extent a SNP is related to the first K principal components.

- Mahalanobis distance computed for each SNP to detect outliers for which the vector of z-scores does not follow the distribution of the main bulk of points.

- Takes into account population structure

- No need to group animals into populations

- Handles admixed individuals

The number of K is user selected on the basis of eigenvalue screeplot

Eigenvalues that correspond to random variation lie on a straight line whereas the ones corresponding to population structure depart from the line.

Recommended the use of Cattell's rule (Cattell 1966): keep eigenvalues to the left of the straight line

Scree Plot – K = 50
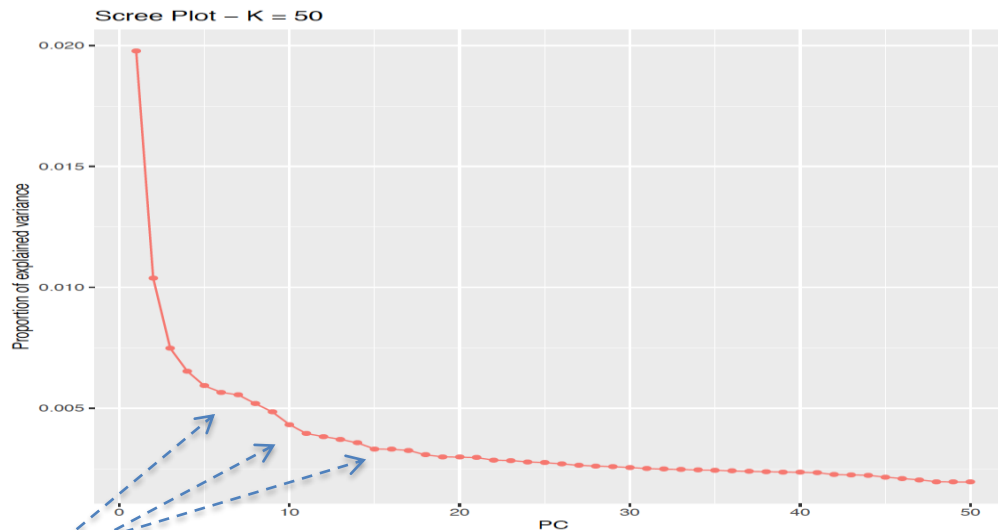
?? K=5, 10, 17?

Correction for population structure considering K=5, 10 and 17 and retained SNPs significant with all three corrections…….

Locus under selection:
- within the top 0.1% Fst values (as in Kijas 2012);
- part of a positive sweep (single outliers not considered).

Mario Barbato

FST - thinning: 0.2
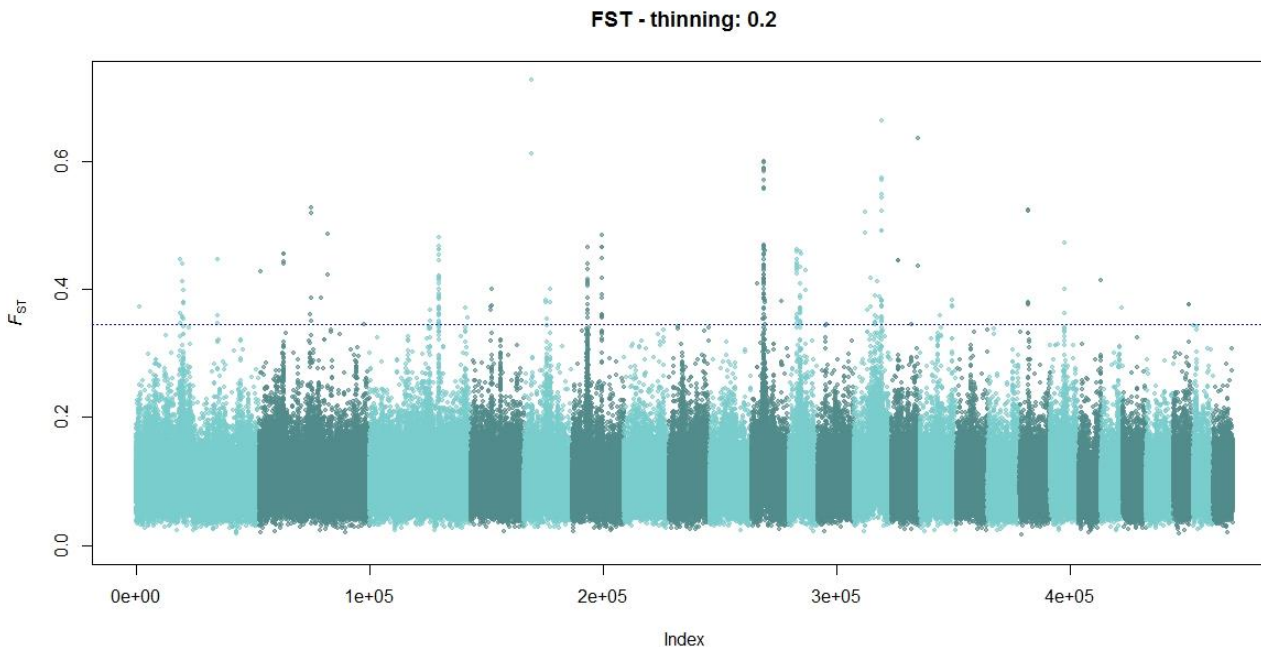
52 genes in ± 2.5Kb* flanking the 155 SNPs candidate under selection

* half of the average distance between adjacent SNPs

PCAdapt

Elia Vajana

- Bonferroni correction for multiple testing with alpha=0.001;

- 690 significant SNPs common to the three K corrections;

- 149 genes in the ±2.5Kb* interval flanking significant SNPs

*mean LD decay in sheep

Marcello Del Corvo
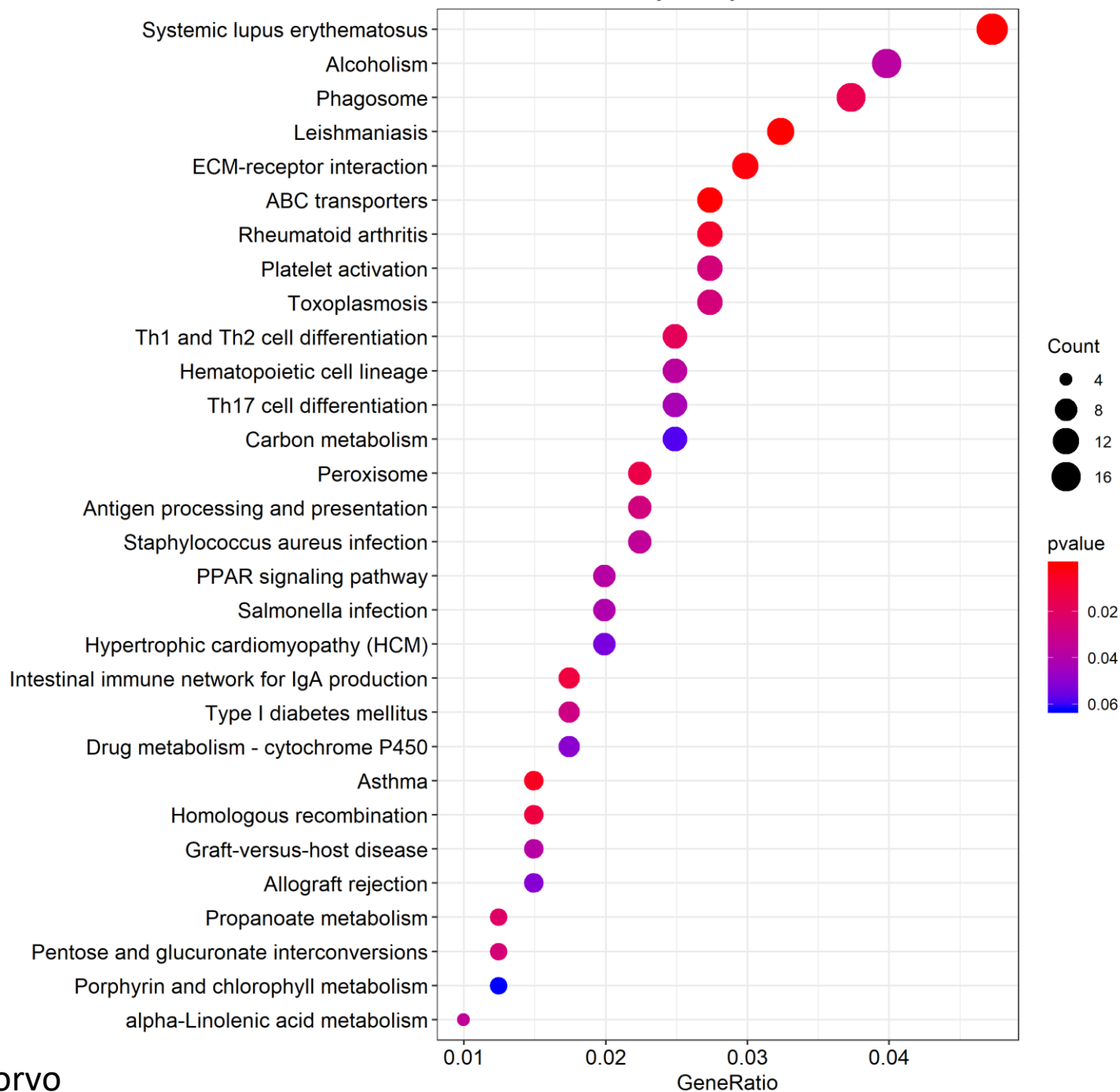
Genes common to Fst and PCAdapt

(10 + 2 different members of a same gene family)

OAR	ensembl_gene_id	Wiki-name	start	end	Trait
3	ENSOARG00000014852	**BBS10**	112151866	112154616	Senses
3	ENSOARG00000019622	**NELL2**	141536582	141928837	Brain development
3	ENSOARG00000002947	**MSRB3**	154219234	154397986	Reproduction
5	ENSOARG00000005541	**ARHGEF37**	58459873	58495649	Metabolism
6	ENSOARG00000002944	**IBSP**	36837847	36848482	Morphology
7	ENSOARG00000020323	**INO80**	33519925	33611601	Chromatin remodeling
10	ENSOARG00000011926	**KATNAL1**	30731527	30780628	Brain disease
13	ENSOARG00000003825	**GPR158**	25018952	25320219	Reproduction
13	ENSOARG00000008957	**RALY**	62852214	62901057	Pigmentation
19	ENSOARG00000004312	**TNNC1**	48403558	48406316	Hearth disease

Cubilin, intrinsic factor, cobalamine receptor

BMC Veterinary Research

BMC Nephrology

RESEARCH ARTICLE   Open Access

Inherited selective cobalamin malabsorption in Komondor dogs associated with a *CUBN* splice site variant

John C. Fyfe[1,2], Shelby L. Hemker[1,6†], Alycia Frampton[3], Karthik Raj[3], Peter L. Nagy[4,7†], Kristi J. Gibbon[5,8†] and Urs Giger[3*]

CASE REPORT   Open Access

Isolated proteinuria due to *CUBN* homozygous mutation – challenging the investigative paradigm

Kushani Jayasinghe[1,2,3,4], Susan M. White[4,5,6], Peter G. Kerr[1,2], Duncan MacGregor[7], Zornitza Stark[3,4,5,6], Ella Wilkins[3,4,6], Cas Simons[3,4], Andrew Mallett[3,4,8,9†] and Catherine Quinlan[3,4,10*†]

Theriogenology 93 (2017) 55–61

Contents lists available at ScienceDirect

Theriogenology

journal homepage: www.theriojournal.com

Cobalamin supplementation during *in vitro* maturation improves developmental competence of sheep oocytes

Federica Zacchini [a], Paola Toschi [b], Grazyna Ewa Ptak [a, c, *]

[a] *Institute of Genetics and Animal Breeding, Polish Academy of Sciences, 05-552 Jastrzebiec, 36a Postepu str., Poland*
[b] *Faculty of Veterinary Medicine, University of Teramo, via Renato Balzarini, 64100 Teramo, Italy*
[c] *National Research Institute of Animal Production, ul. Krakowska 1, 32-083 Balice n. Krakow, Poland*

Methionine-S-sulphoxide reductase 3

- Selection signature in sheep

- calving easy in cattle

- changes in gestational age at birth in humans

GSE Genetics Selection Evolution

RESEARCH ARTICLE **Open Access**

CrossMark

High-resolution analysis of selection sweeps identified between fine-wool Merino and coarse-wool Churra sheep breeds

Beatriz Gutiérrez-Gil[1*] ⓘ, Cristina Esteban-Blanco[1,2], Pamela Wiener[3], Praveen Krishna Chitneedi[1], Aroa Suarez-Vega[1] and Juan-Jose

DNA methylation shows genome-wide association of *NFIX*, *RAPGEF2* and *MSRB3* with gestational age at birth

Hwajin Lee,[1,2†] Andrew E Jaffe,[1,3,4†] Jason I Feinberg,[1] Rakel Tryggvadottir,[1] Shannon Brown,[3] Carolina Montano,[1,2] Martin J Aryee,[1,5] Rafael A Irizarry,[1,4] Julie Herbstman,[3,6] Frank R Witter,[7] Lynn R Goldman,[8,9] Andrew P Feinberg[1,2,4*†] and M Daniele Fallin[1,2,3,4†]

BMC Genomics

RESEARCH ARTICLE **Open Access**

Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds

Mahdi Saatchi[1], Robert D Schnabel[2], Jeremy F Taylor[2] and Dorian J Garrick[1,3*]

Integrin- binding sialoprotein

- major structural component of the bone matrix

- involved in bone diseases

- affects skeletal development in mice

- Associated to sheep height and weight

GSE Genetics Selection Evolution

RESEARCH ARTICLE **Open Access**

Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight

Hawlader A. Al-Mamun[1,2], Paul Kwan[2], Samuel A. Clark[1], Mohammad H. Ferdosi[1], Ross Tellam[3] and Cedric Gondro[1*]

PLOS | ONE

OPEN ACCESS Freely available online

Skeletal Development of Mice Lacking Bone Sialoprotein (BSP) - Impairment of Long Bone Growth and Progressive Establishment of High Trabecular Bone Mass

Wafa Bouleftour[1], Maya Boudiffa[1], Ndeye Marième Wade-Gueye[1], Guénaëlle Bouët[1], Marco Cardelli[2], Norbert Laroche[1], Arnaud Vanden-Bossche[1], Mireille Thomas[1], Edith Bonnelye[3], Jane E. Aubin[2], Laurence Vico[1], Marie Hélène Lafage-Proust[1], Luc Malaval[1*]

- Selection signatures can be searched in managed and natural populations withouth the need for phenotype recording;

- many methods available. Methods correcting for population structure are to be preferred;

- all produce a number of false positives but comfirmation across methods, experiments and species helps in identifying reliable signals.

- The identification of the feature (gene, promoter, enhancer, miRNA…..) and the variant under selection is very challenging;

- even more challenging is the detection of the selective pressure. It is most often done *ex-post* on the basis of (largely incomplete) knowledge of gene function or on the basis of GWAS results. Also, functional annotation of genes is presently biased towards traits of interest for the human species;

Conclusions 3

- Landscape genomics needs no phenotype but precise georeferentiation of samples and partially bypasses the problem of selection pressure identification;

- both selection signatures and landscape genomics are still in the discovery phase. The use of results in practical breeding is still in its infancy;

Thank you for your attention.