

*Subsampling as an economic consequence of
using whole genome sequence data in
landscape genomics:
how to maximize environmental information
from a reduced number of locations?*

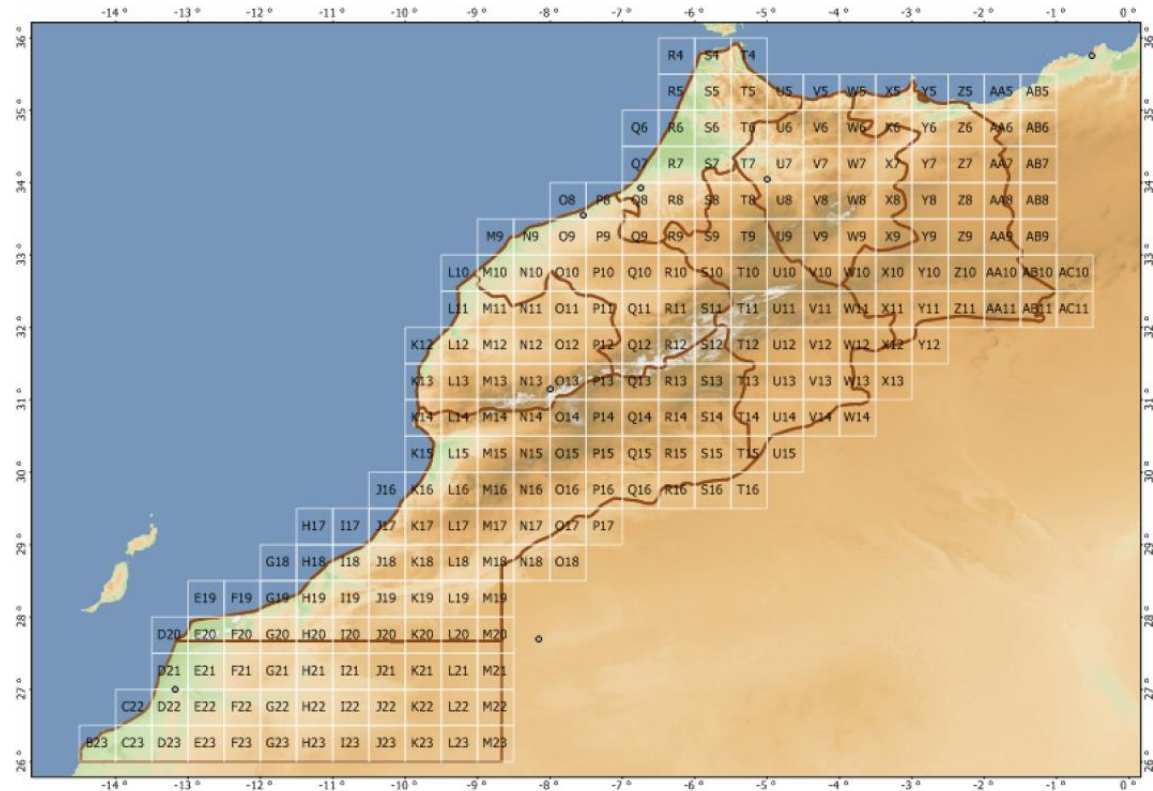
Kevin Leempoel, Sylvie Stucki, Stéphane Joost

Laboratory of Geographic Information Systems (LASIG), Ecole
Polytechnique Fédérale de Lausanne (EPFL), Switzerland

NEXTGEN project in Morocco

Objectives

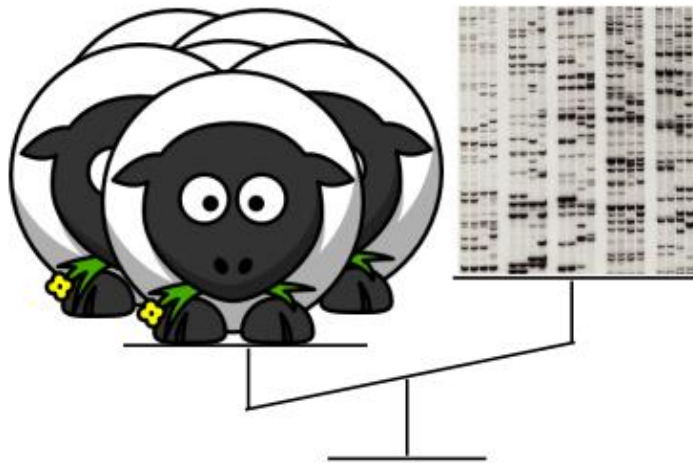
- Whole Genome Sequence of sheep and goat
- Identify SNPs under selection
- Assess which environmental variables are significant
- Determine if SNPs under selection occur in the same genes in both species



Stucki 2014

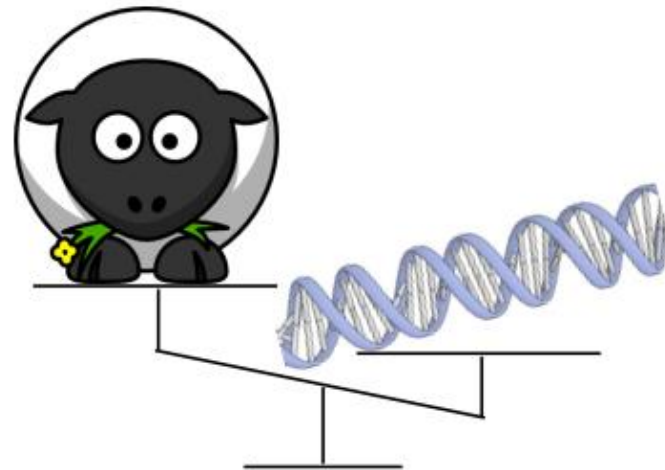
WGS : a constraint for landscape genomics studies

Genetic markers
(ex. AFLP, microsatellites)



1000 individuals / 10 - 1000 markers

Whole Genome Sequencing (WGS)



100 individuals / 10 000 markers

Adapted from Stucki 2014

➤ Requires a careful choice of individuals to sequence

Identifying relevant variables and sampling strategy

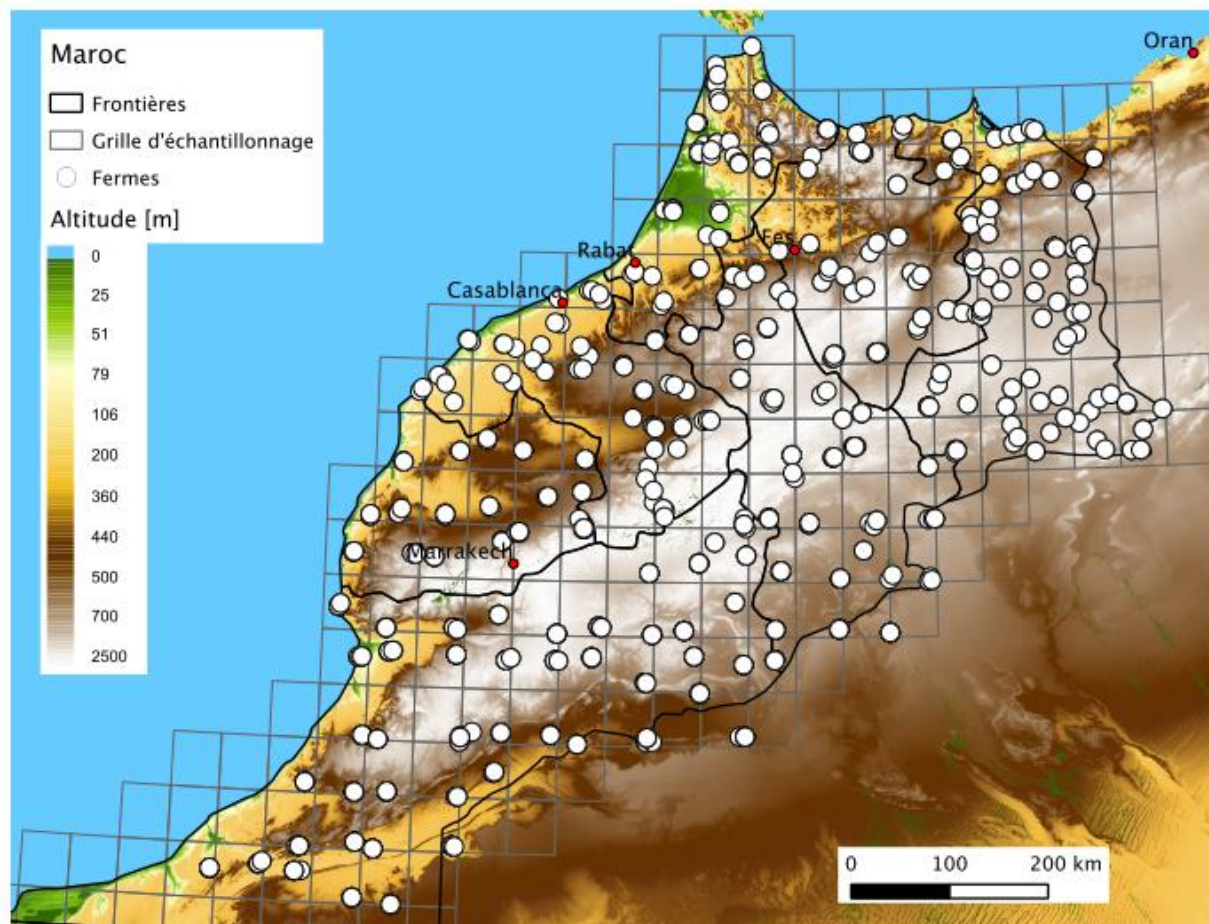
Sampling on a regular grid

3 farms in each cell, 3 individuals per farm

1283 sheeps and 1412 goats reduce to 164 per species

Climatic variables obtained from CRU (resolution : 10')

- Temperature
- Pluviometry
- Solar radiation



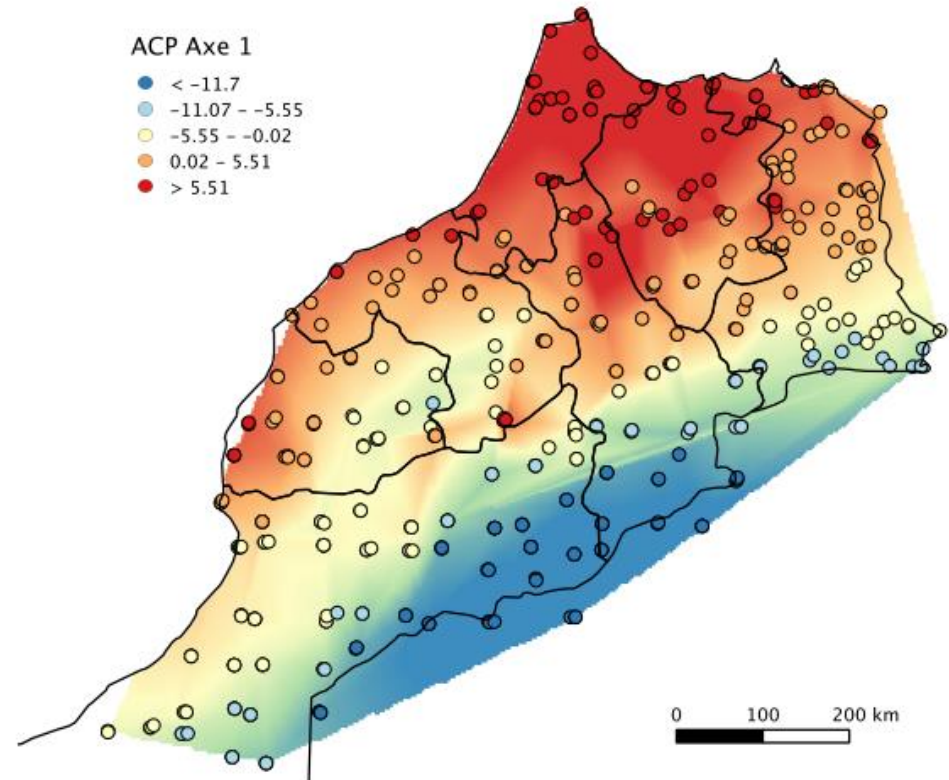
Adapted from Stucki 2014

A stratified sampling methodology to maximise environmental information

Principal component analysis of the entire set of variables (117)

allows to choose samples as different as possible

96% of the variance explained by the first 7 axes of the PCA



Adapted from Stucki 2014

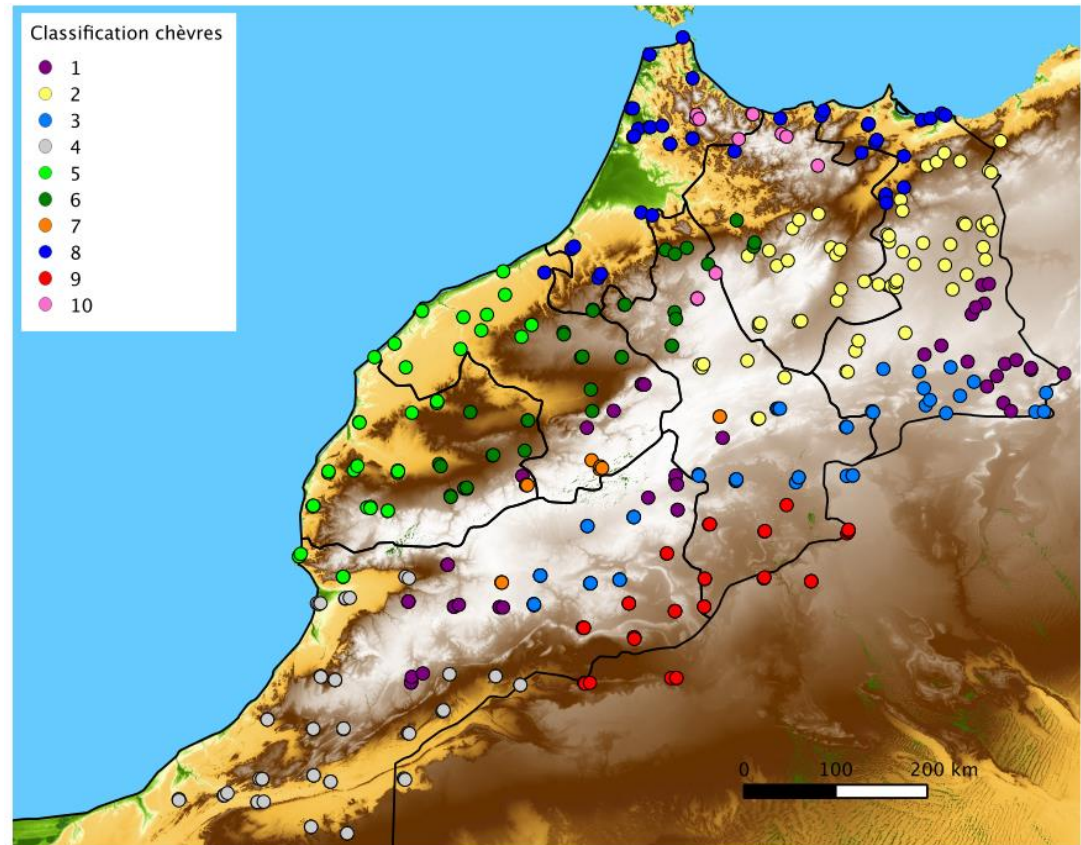
A stratified sampling methodology to maximise environmental information

Hierarchical agglomerative clustering

Classification in function of the climatic conditions

Index of spatial distribution to have samples as evenly distributed as possible

Resulted in regroupement based on spatial and altitudinal proximity



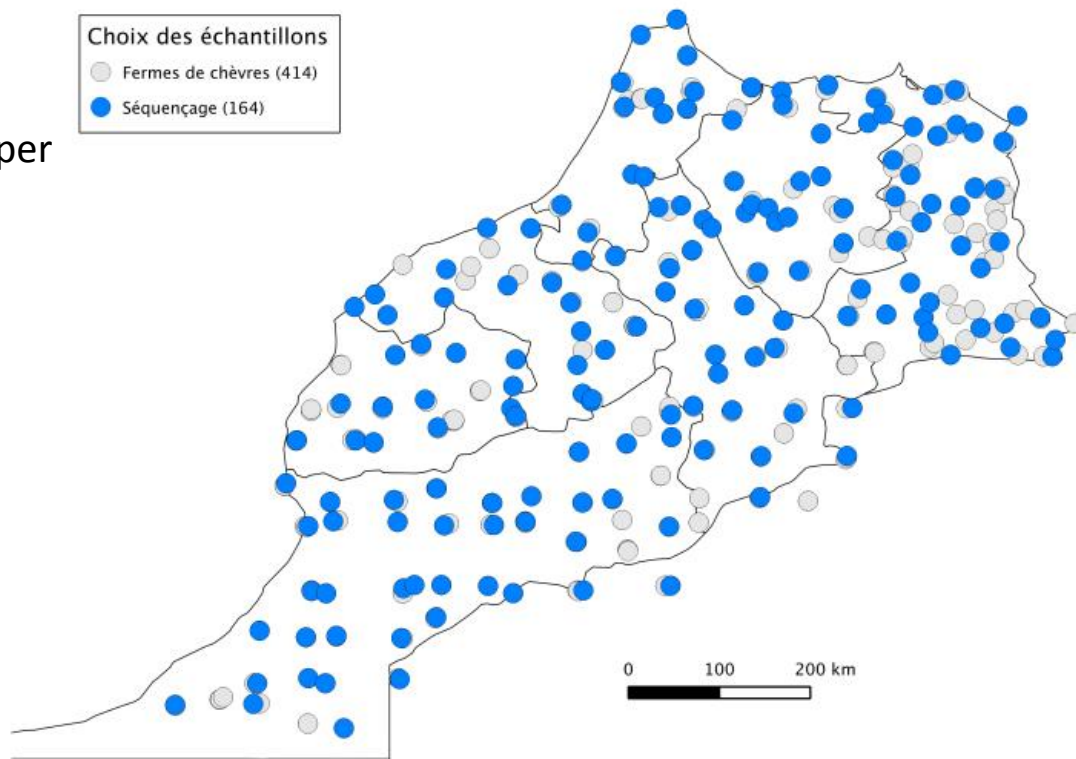
Adapted from Stucki 2014

Results and evaluation

Resulted in a sample of 164 individuals per species.

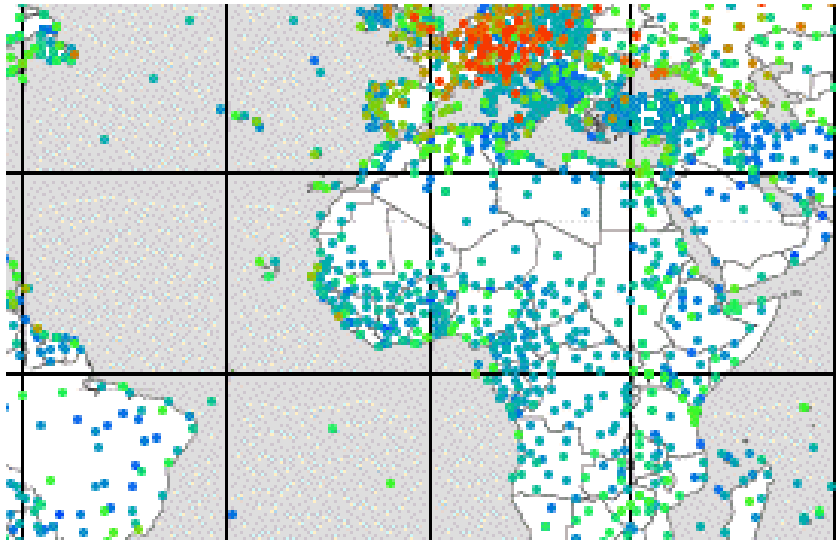
Representativeness of races was also considered

Chosen sheeps



Adapted from Stucki 2014

Reliability of environmental data in function of scale

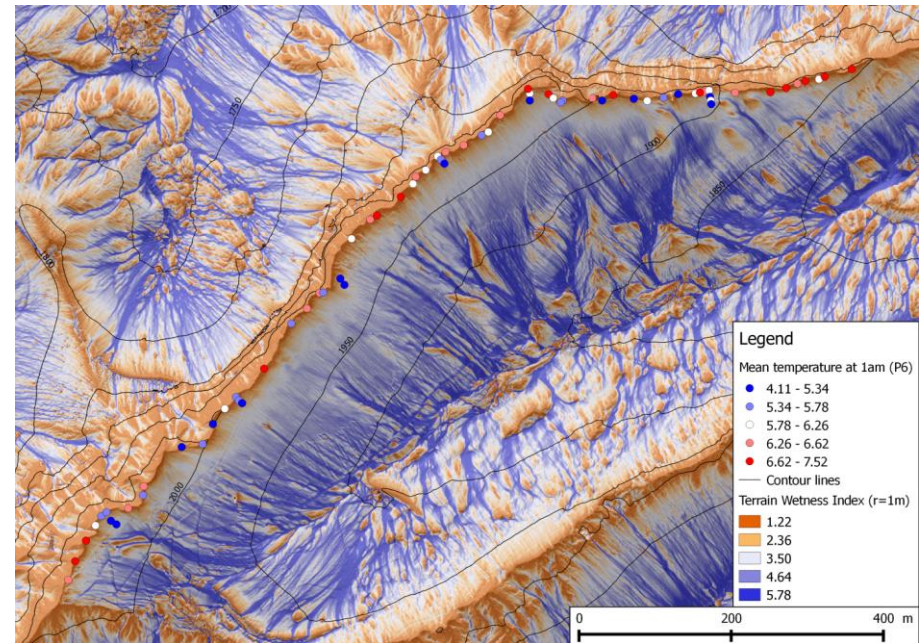


Weather stations used in CRU datasets

- Distance to weather stations
- Precision and resolution of DEMs
- Artefacts in remote sensing data

Relevance of using multi-scale VHR DEM-derived variables in landscape and evolutionary ecology as a surrogate for important climatic variables such as humidity and temperature

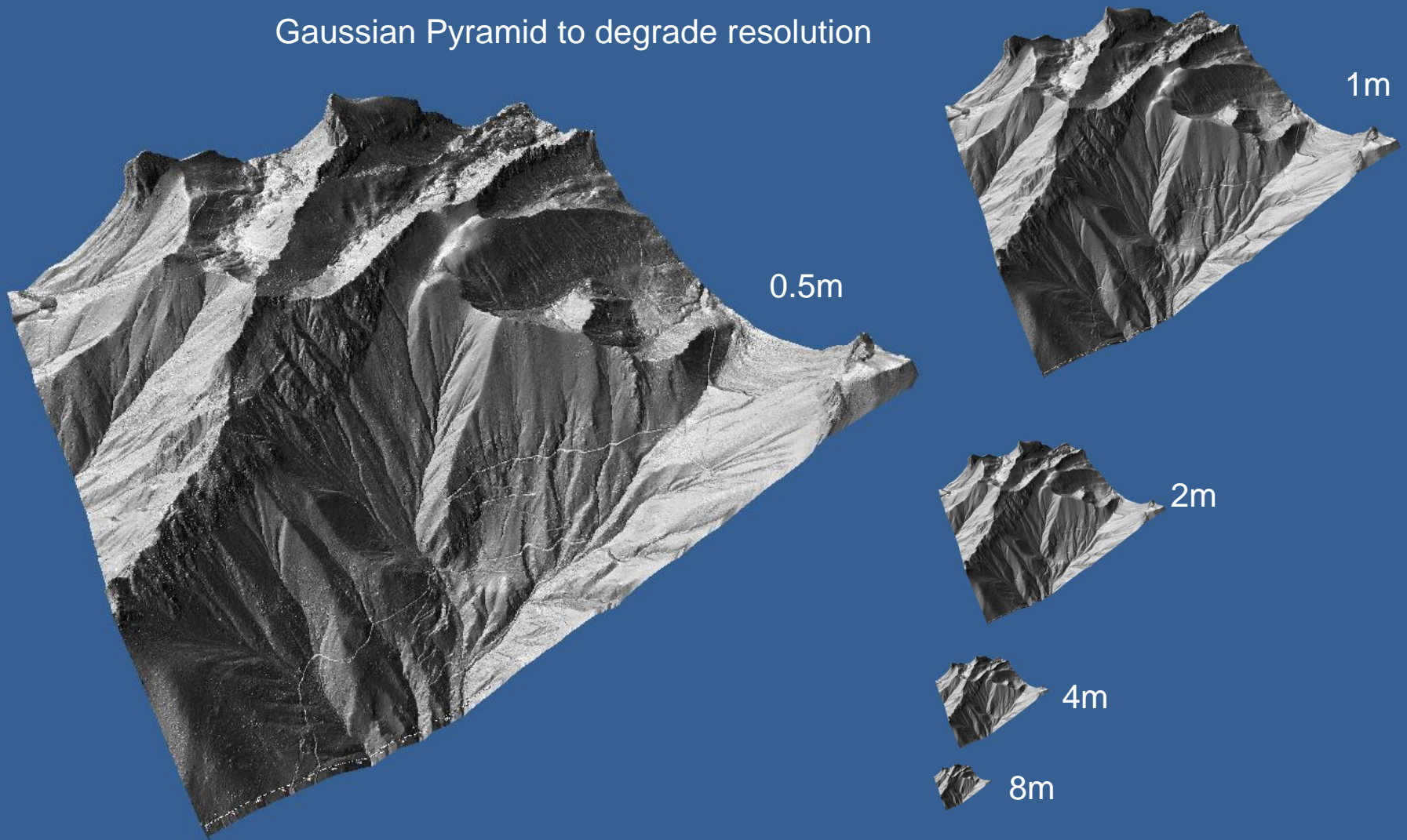
(Leempoel et al., hopefully in 2014)



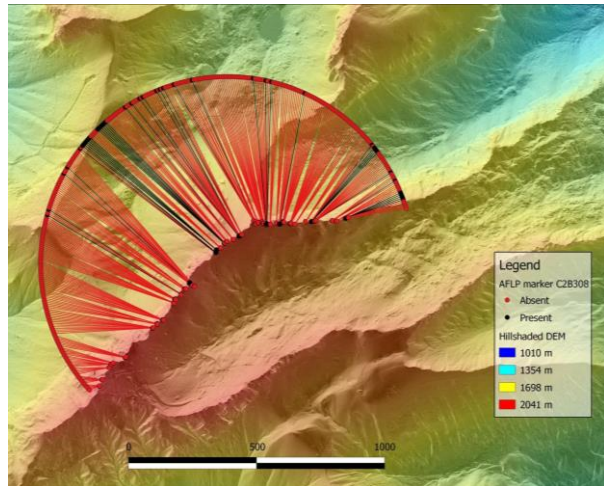
Les Rochers-de-Naye (CH)

Influence of resolution on variables and association models

Gaussian Pyramid to degrade resolution

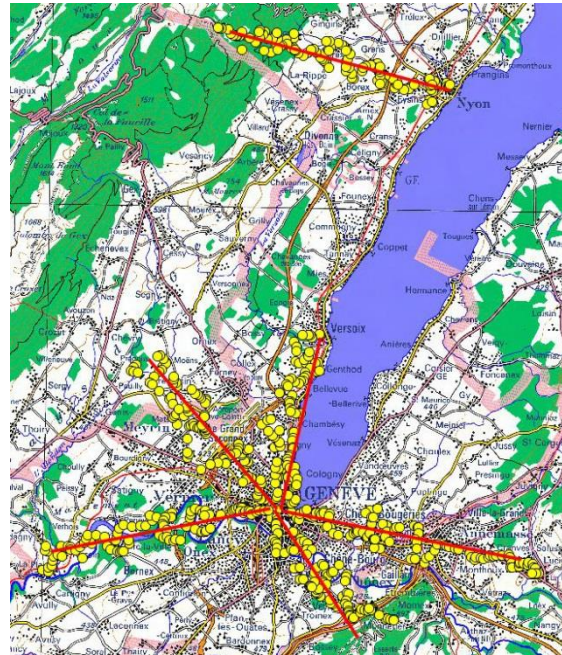


At which scale do they matter ?



Les Rochers-de-Naye

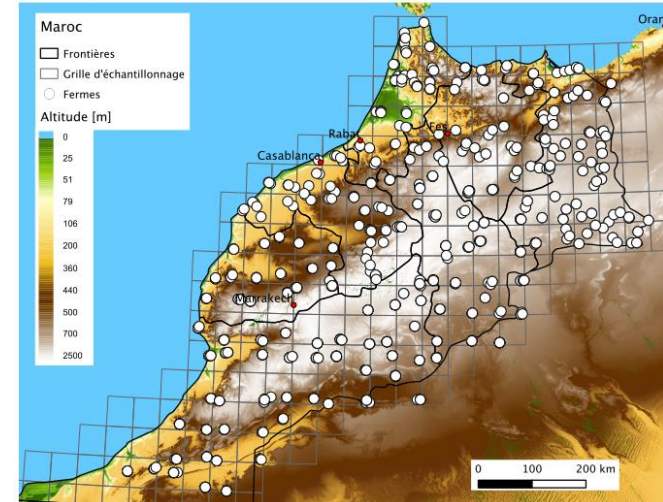
- Local scale (1.5km)
- Alpine plant
- Multi-resolution



Adapted from Lassueur et al. (2007)

UrbanGene

- Local scale (10x10km)
- Urban plant
- Resolution and window size

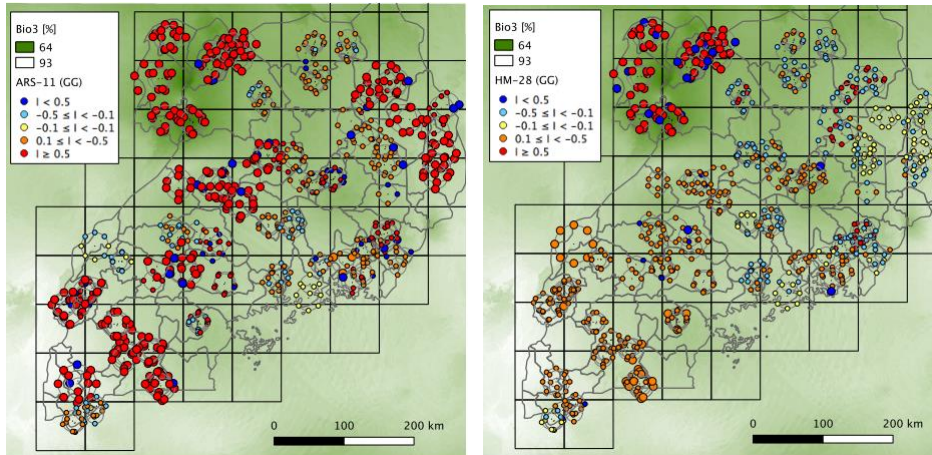


NEXTGEN Morocco

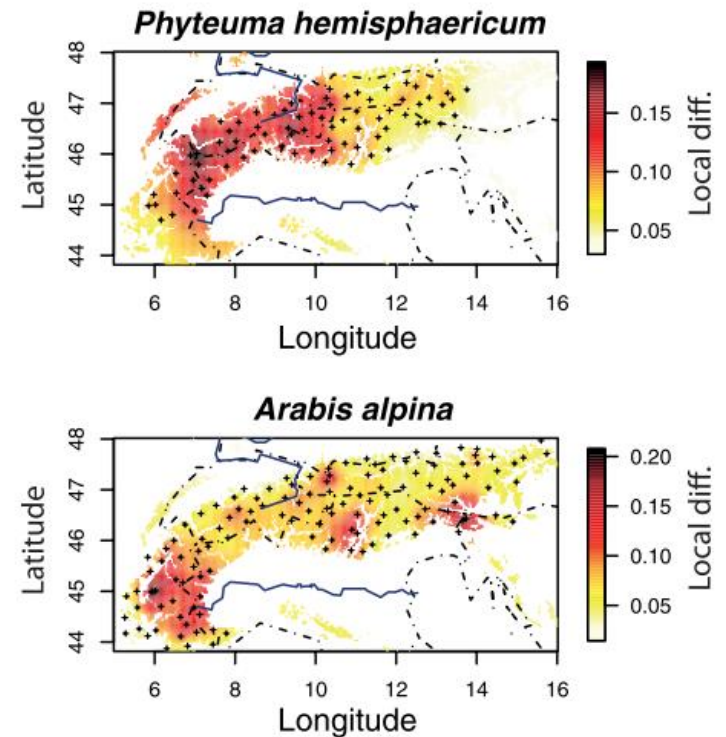
- Regional scale (800x600km)
- Sheeps and Goats
- Resolution and window size

Exploiting spatial tools to analyse the scale and spatial component

- Local index of spatial association (LISA)
- Geographically weighted regression (GWR)
- LocalDiff



Univariate LISA



Heatmap of **local genetic differentiation** computed across the Alps

Thank you for your attention



EUROPEAN
SCIENCE
FOUNDATION
GENOMIC-RESOURCES

EPFL

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

LABORATORY OF GEOGRAPHIC INFORMATION SYSTEMS

LASIG 12

LIVESTOCK GENOMIC RESOURCES – CARDIFF 17 JUNE 2014