

Adaptation of Mediterranean local cattle breeds to climate changes: a preliminary study on genome-wide diversity

S. Casu¹, R. Ciampolini², S. Mastrangelo³, L. Flori^{4,5}, K. Moazami-Goudarzi⁴, T. Sechi², F. Casabianca⁶, A. Lauvie⁶, M. Gautier⁷, A. Carta², B. Portolano³, D. Laloë⁴

¹AGRIS-Sardegna S.S. 291 Sassari-Fertilia – Km. 18,600, 07100 Sassari

²Dipartimento di Scienze Veterinarie, Università di Pisa, 56100 Pisa Italy

³Dipartimento Scienze Agrarie e Forestali, Università degli Studi di Palermo, 90128 Palermo, Italy

⁴Inra/AgroParisTech, GABI, 78352 Jouy-en-Josas, France

⁵Cirad, INTERTRYP, 34000 Montpellier, France

⁶Inra, LRDE, 20250 Corte, France

⁷Inra/IRD/Cirad/Montpellier SupAgro, CBGP, 34988 Montferrier-sur-Lez, France

Corresponding author: scasu@agrisricerca.it

Summary

In recent decades, changes in climate have caused impacts on natural and human systems on all continents. Mediterranean countries will be particularly affected by this phenomenon, with growing temperatures and reduced rainfall during summer months. Understanding how species and ecosystems respond to climate change has become a crucial focus in ecology, biodiversity conservation and management. The recent availability of genome-wide SNP panels allows providing background information concerning genome structure in domestic animals, opening new perspectives to livestock genetics. Several approaches have been performed in landscape genomics to detect adaptation to different climate pressure by correlations between genomic data and climate variables. The International Project GALIMED "Genetic Adaptation of Bovine Livestock and production systems in MEDiterranean region", funded by INRA (metaprogram ACCAF), develops an integrated approach that combines the analysis of SNP markers, bioclimatic variables and farming system information to identify genotypes and breeding practices able to respond to climate change. The general aim of the research is to identify genomic regions related to adaptation to climate change in these local cattle. In this work in particular, the genetic variability within breed and the genetic relationships between Italian and Corse local cattle breeds, that are part of this project, will be presented. Individuals of Cinisara (71), Modicana (72), Maremmana (25), Piemontese (21), Romagnola (21), Sarda (30), Sardo-Modicana (28) and Corse (31) cattle breeds, were genotyped using Illumina Bovine SNP 50k BeadChip. Piemontese e Romagnola genotyping data were public available, whereas the other breeds are part of the GALIMED project. Farming systems data were collected by interviewing breeders. Geographic coordinates and 19 bioclimatic variables were also available. A Principal Components Analysis (PCA) was performed either on SNPs data and climatic variables. Afterwards a Co-inertia (CIA) analysis was realized to detect a possible common structure between such different information. We present here first results of genetic variability and climatic data. After edits, 43,625 SNPs were retained for the analysis. Breeds are clearly differentiated according to geography and climate. The first axis of PCA on individual genotypes differentiates Modicana from other breeds, and shows that Sardo-Modicana is close to Modicana, but

is clearly admixed to another breed. The second axis differentiates Cinisara from a cluster "Romagnola/Maremmiana" and shows a large variability of the Modicana breed. The CIA coefficient between molecular data and both geographic and climatic information is equal to 0.38, significantly greater than 0 (p -value=0.001), suggesting evidence of genetic adaptation to different climatic pressures of the studied breeds. Further analyses are on going to identify the genetic regions with a potential adaptive role. The results will provide to farmers, technicians and industrial partners, a solid scientific foundation to reconsider their objectives and selection criteria and to improve their farming practices to prepare their livestock to the new environmental conditions.

Introduction

In recent decades, changes in climate have caused impacts on natural and human systems on all continents. Mediterranean countries will be particularly affected by this phenomenon, with growing temperatures and reduced rainfall during summer months, and recurrent heat waves and droughts. Understanding how species and ecosystems respond to climate change has become a crucial focus in ecology, biodiversity conservation and management. Despite the growing diffusion of some cosmopolite specialized cattle breeds, several autochthonous breeds, well adapted to their— often marginal — environment, are still bred in the Mediterranean area. In fact, cattle production represents an important resource for the economy of hilly and mountain areas, in which other economic activities are limited. In a time of rapid and unpredictable change in climate, understanding the genetic bases of adaptive ability of livestock to different environments may contribute to identify conservation/breeding objectives and to develop more efficient methodologies for genetic improvement to face the consequences of climate change.

The recent advent of high-throughput and cost-effective genotyping techniques makes it possible to provide a detailed genome wide assessment of the genetic structure and relationships among cattle populations. SNP panels open up new perspectives, in particular for the investigation of genome diversity within and among individuals and populations, fine scale marker-trait association in the search of causative genes and mutations, and for the identification of signatures left by selection. This last application provides an attractive prospect for the identification of genomic regions sustaining traits linked to adaptability (adaptation to extreme climates and poor fodder and resistance to disease) often very difficult, expensive and even impossible to record.

In order to investigate the genetic basis of cattle adaptation abilities, an international Project GALIMED "Genetic Adaptation of Bovine Livestock and production systems in MEDiterranean region", has been funded by INRA (metaprogram ACCAF). The GALIMED project proposes a multidisciplinary approach combining population genetic, environmental, climatic, livestock management and production systems data to characterize Mediterranean cattle breeds, quantify biodiversity and describe the covariation of these different factors in order to detect genetic variants and selection signals associated with adaptive traits (to extreme climates and poor fodder and disease resistance) with particular emphasis on climatic adaptation. At present 21 local breeds from the north and the south of the Mediterranean basin have been included in the project. These breeds have been locally settled for centuries and reared under traditional extensive production systems. Thus, they are supposed to have undergone adaptive selection for a long time frame. In this paper, first genetic characterization of 5 Italian cattle breeds (Cinisara and Modicana from Sicily; Sarda and Sardo-Modicana from Sardinia; Maremmiana, from Tuscany) and Corse cattle, included in the GALIMED project will be presented. Piemontese and Romagnola (from north Italy) genotyping data, publicly available, were also included in the analysis. The aims of the

work were to study the genetic relationships between these local breeds of the Central Mediterranean Area, reconstruct their genetic structure and identify genomic regions related to adaptation to climate change.

Materials and methods

Animals and Genotypings

Individuals of Cinisara (CIN, n=71), Modicana (MOD, n=72), Maremmana (MAM, n=25), Piemontese (PMT, n=21), Romagnola (RMG, n=21), Sarda (SAR, n=30), Sardo-Modicana (SAM, n=28) and Corse (COR, n=31) cattle breeds, were genotyped using Bovine SNP50k v2 BeadChip (Illumina). Piemontese and Romagnola genotyping data were publicly available, whereas the other breeds are part of the GALIMED project. Genotypings were performed by Porto Conte Ricerche s.r.l. (Alghero, Italy), Labogena (INRA, Jouy-en-Josas, France) and Dipartimento Scienze Agrarie e Forestali (University of Palermo, Italy).

Only the SNPs located in the autosomes were considered. Markers were filtered according to quality criteria that included: 1) missing SNP genotype rate <0.05; 2) call frequency ≥ 0.98 ; 3) minor allele frequency (MAF; ≥ 0.01). Nine individuals with more 1600 missing genotypes were also discarded. After editing 41030 SNPs and 290 individuals were retained. File editing was carried out using PLINK (Purcell et al. 2007).

Genetic diversity within and between breeds

Within-breed genetic diversity was estimated with the following parameters: number of monomorphic loci (npSNP), average MAF, observed (H_o) and expected (H_e) heterozygosity, inbreeding coefficient (F) and effective population size (N_e) estimated from the extent of linkage disequilibrium (LD) (Tenesa et al, 2007).

In particular, the historical effective population size (N_e) for each breed was estimated, without considering mutation, as

$$N_e = \frac{n(1 - r^2) + 1}{4c(nr^2 - 1)}$$

where n is the breed chromosome sample size (2*number of individuals), r^2 measures the average level of LD between pairs of autosomal SNPs from 0.01 to 1 Mb apart, and c is the distance between the SNPs in Morgans (we assumed 1 Mb = 0.01 Morgans). N_e was estimated for $t = 1/2c$ generations ago, thus from 5000 to 50. Basic genetic diversity indices were obtained using PLINK (Purcell et al, 2007) and author developed EXCEL worksheets.

In order to investigate genetic relationship between breeds, a principal component analysis (PCA) was performed on individual genotypes using the glPca function of the Adegenet R package (Jombart et Ahmed, 2011).

Geographical and bioclimatic variables

In order to study the influence of climate on the investigated breeds, bio-climatic variables characterizing the traditional rearing area of each were downloaded from the worldclim database (<http://www.worldclim.org>). Geographic coordinates for localizing the area were provided by contributors. To characterize the rearing environment and reduce the number of variables and then the dimensionality of the subsequent analyses a PCA was also performed on these 19 variables, while latitude, longitude and altitude were used as supplementary variables. The corresponding screenplot and correlation circle were then plotted.

Joint analysis between genetic data and climatic parameters

In order to detect a possible common structure between genetic data and bioclimatic variable, a co-inertia analysis (CIA) was performed. Co-inertia analysis (CIA, Doledec and Chessel, 1994; Dray et al, 2003) is a multivariate method for coupling two tables looking for a possible co-structure of two type of information. Results of the CIA consist in two sets of scores (genotypes and geography/climate) of maximum covariance. CIA maximizes the product of the correlation by the projected variances, ensuring that co-inertia axes will have both a good correlation with the initial variables and real meaning for each of the two data sets. Additionally we measure the overall similarity using a multivariate extension of the Pearson correlation coefficient called the RV-coefficient. The RV-coefficient is calculated as the total co-inertia (sum of eigenvalues of a co-inertia analysis) divided by the square root of the product of the squared total inertias (sum of the eigenvalues) from the individual COAs. It ranges between 0 and 1. A RV-coefficient close to 1 indicates a very high degree of co-structure.

Results and discussion

The basic genetic diversity indices are reported in Table 1. MAM and RMG showed the highest number of monomorphic SNPs and consequently the lowest MAF values and gene diversity (H_e). MOD, MAM and RMG presented the lowest N_e and a highest level of inbreeding, suggesting a narrow genetic base and mating between close relatives. CIN and SAR breeds showed the lowest number of monomorphic SNPs and the highest H_e and N_e . Similar values of gene diversity were observed for the COR and PMT breeds, with lowest values of F and N_e . The SAM has intermediate values of all indices.

Table 2: Estimates of genetic diversity indices between cattle breeds

Breed	npSNP	$H_o \pm s.d$	$H_e \pm s.d$	MAF \pm s.d	$F \pm s.d$	N_e
Cinisara	589	0.323 ± 0.155	0.332 ± 0.152	0.250 ± 0.146	0.05 ± 0.07	934
Modicana	1957	0.310 ± 0.176	0.307 ± 0.167	0.230 ± 0.153	0.09 ± 0.06	454
Piemontese	2305	0.337 ± 0.180	0.327 ± 0.157	0.247 ± 0.149	0.01 ± 0.01	657
Corsa	2006	0.328 ± 0.170	0.328 ± 0.156	0.248 ± 0.149	0.03 ± 0.04	779
Maremmiana	4623	0.315 ± 0.195	0.300 ± 0.173	0.225 ± 0.156	0.07 ± 0.03	287
Romagnola	4339	0.306 ± 0.192	0.300 ± 0.173	0.225 ± 0.156	0.10 ± 0.03	295
Sarda	1479	0.325 ± 0.165	0.332 ± 0.154	0.251 ± 0.148	0.04 ± 0.06	873
Sardomodicana	2413	0.325 ± 0.179	0.320 ± 0.162	0.240 ± 0.151	0.04 ± 0.03	536

Number of monomorphic SNPs (npSNP), Average minor allele frequency (MAF), observed heterozygosity (H_o), expected heterozygosity (H_e), inbreeding coefficient (F) and standard deviation (s.d.), and effective population size (N_e) 50 generations ago.

Results of PCA on individual genotypes are displayed in Figure 1 where arrows are drawn from the breed barycenter (position of the breed) to each animal and colors are according to breeds. Inertia ellipses were drawn for each breed. The first axis clearly separated MOD from other breeds, and showed that SAM was closed to MOD, but is clearly admixed to another breed. The second axis separated CIN from a cluster RMG/MAM and showed a large variability for the MOD breed.

Figure 1 PCA based on genotypic data for individuals in Italian and Corse local cattle breeds plotted for the first and second components

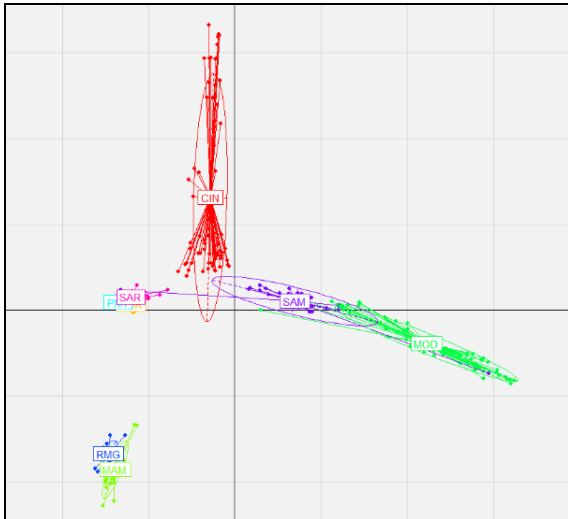
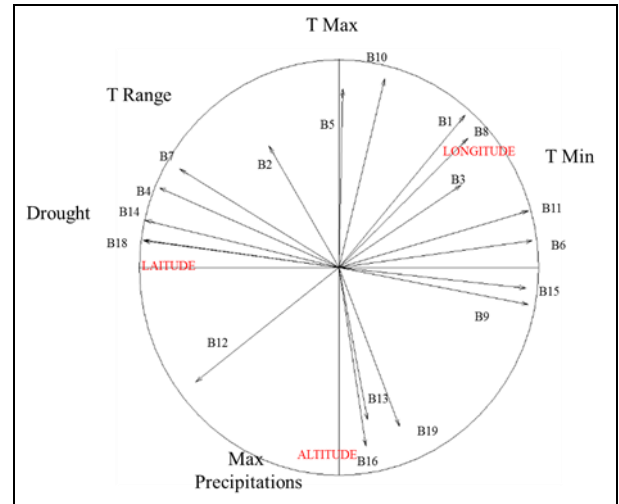


Figure 2 Correlation circle from PCA based on 19 bioclimatic variables

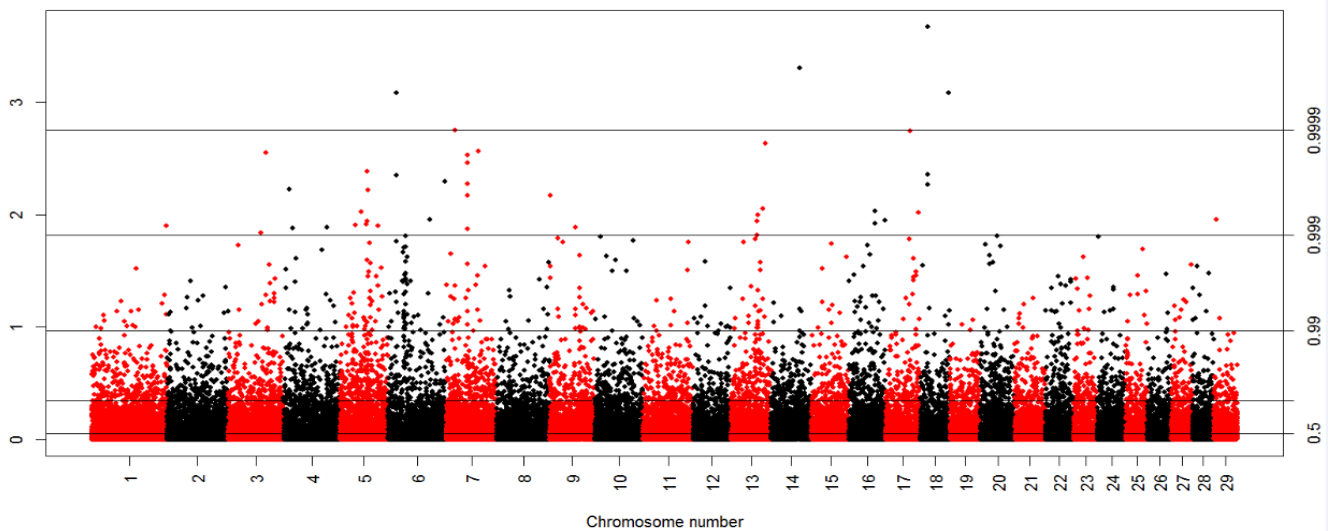


B1 = Annual Mean Temperature; B2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)); B3 = Isothermality (B2/B7) (* 100); B4 = Temperature Seasonality (standard deviation *100) ; B5 = Max Temperature of Warmest Month; B6 = Min Temperature of Coldest Month; B7 = Temperature Annual Range (B5-B6); B8 = Mean Temperature of Wettest Quarter; B9 = Mean Temperature of Driest Quarter; B10 = Mean Temperature of Warmest Quarter; B11 = Mean Temperature of Coldest Quarter; B12 = Annual Precipitation; B13 = Precipitation of Wettest Month; B14 = Precipitation of Driest Month; B15 = Precipitation Seasonality (Coefficient of Variation); B16 = Precipitation of Wettest Quarter; B17 = Precipitation of Driest Quarter; B18 = Precipitation of Warmest Quarter; B19 = Precipitation of Coldest Quarter

Results of PCA on bioclimatic variables are depicted in Figure 2. The two first axis explained 82.1% of the total inertia. The first axis was mainly correlated to variables related to minimal precipitation and temperatures and was linked to latitude; the second axis was correlated to values of maximum precipitation and temperatures and was linked to altitude.

The RV coefficient between molecular data and both geographic and climatic information was equal to 0.38, significantly greater than 0 (p-value=0.001), suggesting evidence of genetic adaptation to different climatic pressures of the studied breeds. Indeed Manhattan plot of the contributions (squared scores) of SNPs to the second component of the coinertia analysis (Figure 3) allowed to distinguish some genomic regions mostly contributing to the breeds' differentiation according to drought and latitude (first component of both the PCA on bioclimatic variables and coinertia analysis)

Figure 3 SNPs contribution (squared scores) to the first component of the coinertia Analysis



Conclusions

First results of combined genotypic and climatic data suggests that the 8 investigated cattle breeds from central Mediterranean area showed genetic differences which may have been shaped by different climatic conditions. Further analyses are on going to identify the genetic regions with a potential adaptive role. Adaptive ability of livestock to climate variations will indeed be a factor of first importance in the Mediterranean climate extending area. The GALIMED project, with its multidisciplinary approach, will contribute to understand the genetic basis of adaptation and to provide to farmers, technicians and industrial partners, a solid scientific foundation to reconsider their objectives and selection criteria and to improve their farming practices to prepare their livestock to the new environmental conditions.

References

- de Roos, A.P. et al. (2009) Reliability of genomic predictions across multiple populations. *Genetics* 183: 1545–1553
- Doledec, S., and D. Chessel (1994). Co-inertia analysis: an alternative method for studying species–environment relationships. *Freshwater Biology* 31:277–294.
- Dray S., Chessel D. and J. Thioulouse (2003): Co-inertia analysis and the linking of ecological data tables. *Ecology*, 84, 3078-3089
- Jombart T. and Ahmed I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*
- Purcell, S. et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.
- Tenesa A., Navarro P., Hayes B.J. et al., (2007) Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 17: 520-526

Acknowledgment

This work was partially funded by “INRA metaprogramm ACCAF, “Misura 214 PSR 2007-2013 RAS” RisGenSar project and “PON02_00451_3133441”. The authors are very grateful to the director of the Alberese ANABIC Selection Center of Maremmana breed to have been actively promoting this project