Genetic and geographical integration for ruminant production under climate change with particular emphasis on Brazil

Conceita McManus* † | Helder Louvandini ‡ | Potira Hermuche ‡ |
Renato Fontes Guimarães † | Osmar Abilio Carvalho Junior † | Felipe Pimentel † |
Daniel Pimentel † | Samuel Rezende Paiva † | Vanessa Peripolli † |

*Corresponding author: conceita@unb.br

Abstract The use of georeferencing technologies and genetic information has increased to integrate management and planning of livestock production systems, predict adaptive capacities, and aid in developing strategies for national Animal Genetic Resource Conservation Programs. Researchers and farmers can use this information to define conservation objectives for individual breeds and examine environmental factors that affect extinction risk, such as disease threats. Molecular markers and geographic information come together in landscape genetics, a combination of landscape ecology and population genetics, to provide information on the interaction between landscape and evolutionary processes. Results reveal attributes that affect genetic adaptation to specific environmental stressors such as diseases, parasites, extreme heat, vegetation type, lack of water, or combinations. Recent preliminary studies in Brazil used these tools to identify the regional usage patterns for animal production based on environmental criteria and breed distribution data. The results have been used as a further criterion to optimise in situ, and ex situ conservation schemes and plan expansion and adaptation of production systems. The use of production environment descriptors and climatic and genetic information will help maintain animal production systems in a changing world.

Keywords: georeferencing, landscape genetics, molecular markers, stressful environments

1. Introduction

In general, stress is looked upon as a symptom resulting from exposure of an animal to a hostile environment (Stott 1981) and consists of external body forces that tend to displace homeostasis. Environmental forces continuously act upon animals that disrupt homeostasis, resulting in new adaptations that can be detrimental or advantageous to man’s interest. Measuring the magnitude of stress often means measuring the degree of adaptation. In animal production, selection pressures have been increased to improve productivity using modern animal breeding techniques and reproductive technologies, reducing fitness, disease resistance, and tolerance to rising environmental stresses such as higher temperatures or lower water availability. Seo et al (2010) found that climate variables are highly significant determinants of primary species choice after controlling for soils, geography, household characteristics, and country fixed effects. There are also questions related to food security, social and economic aspects, and environmental impacts such as the production and management of waste, greenhouse gas emissions and livestock-related land-use change, and large-scale public health and epidemiological investigations (Robinson et al 2014).

Changes in the geographic zones of suitability for many animal species are expected with climate change. Still, studies show gaps in our knowledge of essential scientific issues about climate change impacts and adaptation in livestock systems (Escarcha et al 2018). Some regions become more favourable for increased agricultural productivity, and others become less favourable. Environment, feed (quantity and quality), water availability, and various aspects linked to the animals themselves define an animal’s adaptation to stressful environments. How can we use geographical technologies with currently available information to predict regions where animal genetic resources should adapt, investigate regions of the genome to find adaptive
genes, and define environmental limits for farm animal genetic resources? Several studies show the expected impacts of climate change on livestock (Rojas-Downing et al 2017, McManus et al 2020a), but few look at integrating genetic and geographic information to mitigate and plan livestock expansion or management systems. According to Leempoel et al (2017), we can potentially identify the genetic bases of species adaptation to particular geographic conditions or climate change using Geographical Information Systems (GIS). Manel et al (2003) described how geographical and environmental features could facilitate genetic variation structure at the population and individual levels, with implications for ecology, evolution, and conservation biology. This approach differs from other genetic approaches, such as phylogeography (Marske 2016). It focuses on finer spatial and temporal scales, such as those with recent farm livestock migration.

Currently, it is recommended to seek the union of different disciplines (genetics, sociology, economics, and geography) for efficient characterisation of genetic resources (Yaro et al 2017). The use of global positioning systems (GPS) in conjunction with molecular tools has been suggested to provide a better description of the relationship between genetics and the environment (Groeneveld et al 2010). There is increased interest in using georeferencing technologies and genetic information to manage and plan livestock production systems and predict adaptive capacities. Results can be used as criteria to optimise in situ and ex situ conservation schemes and design expansion and adaptation of production systems (Hermuche et al 2013, Costa et al 2014). Understanding spatial genetic differences are essential for conservation programs to achieve adequate sampling and representation of populations. According to Herrero et al (2017), information about the global structure of agriculture and nutrient production and its diversity is essential to improve the present understanding of national food production patterns, agricultural livelihoods, food chains, and their linkages with land use and their associated ecosystem services. This information can promote healthy diets and ecosystems in the face of population growth, urbanisation, and climate change. Karesh et al (2012) also showed that geographical information could be helpful when treating animal diseases and zoonoses.

Studies with locally adapted breeds can help identify genes and metabolic pathways of interest in adaptation studies, as they have had sufficient time to be genetically adapted to the environment (Galal and Boyazoglu 2001). These adaptations arise through equilibrium between evolutionary (crossbreeding, artificial selection, and mainly genetic drift) and local environment forces (Cesconeto et al 2017). Greater integration of knowledge of population genetics and conservation of genetic resources is required within animal breeding programs. The long-term sustainability of animal genetic resource management activities must be a constant concern for commercial and locally adapted breeds. The involvement of farmers' associations and cooperatives (breeders) is indispensable for maintaining long-term policies involving genetic resources in a region (Leroy et al 2017). Therefore, there is a need to unite all agents in the production chain in the decision-making process regarding the management and use of genetic resources.

2. Biodiversity, environment and production environment descriptors

Agrobiodiversity is a critical component of global biodiversity. More than 75% of food worldwide is produced by a little more than 17 species of domestic plants (12) and animals (5). The manipulation and management of these resources are essential for food safety worldwide. On the other hand, the pressure to increase productivity and sustainability of production systems is intense and increasing. The planetary dimension of research problems (sustainable development, climate change, management of biodiversity, water quality and quantity, emerging diseases, and bioenergy), the need to increase food production as a response to the increase in human population and change in its distribution put the farm animals, their management and breeding problems in the centre of the debate in society (KC et al 2018, Berners-Lee et al 2018). Increasing the efficiency of livestock production systems will be necessary. Studies in farm animal systems worldwide have shown that there will be a need to substitute breeds and species in production systems over the next 30 years due to climate and market changes (Godber and Wall 2014, Gaughan and Cawdell-Smith 2015, Zhang et al 2017).

Changes in allele frequencies in specific loci depend on the selection pressures and objectives. A hitchhiking effect occurs whereby chromosome regions surrounding advantageous alleles are swept during this process (Fay and Wu 2000), which results in genomic regions with elevated homozygosity, known as selection signatures (Haasl and Paysseur 2016).

One of the responses to climate change is the use of locally adapted breeds in purebred or crossbred systems and the study of these breeds to elucidate physical or physiological traits and metabolic pathways that contribute to resilience, resistance, or tolerance under stressful conditions. Traditionally, domestic animal characterisation has been based on phenotypic characteristics (morphological and productive) resulting from the genotype x environment interaction. The genotypes subjected to specific environmental conditions could change (Mariante and Cavalcante 2006). Thus, the choice of individuals representing the breed or population to be preserved or studied only by its genotype resembles the pursuit of a moving target that changes according to the environment. One way to avoid this problem is by associating genetic data with phenotypic information (Egito et al 1999). Studies show links between Brazilian local adapted or naturalised breeds in terms of heat tolerance (McManus et al 2009a, McManus et al 2009b, Castanheira et al 2010), parasite resistance (Bricarello et al 2004, McManus et al 2010) as well as genes that are specific to locally adapted breeds (Silva et al 2011).

https://www.malque.pub/qjs/index.php/avr
Accurate descriptions of production environments are also important in this context. Escaracha et al (2018), looking at climate change and adaptation in livestock systems, highlight the need for understanding production systems and animal (and plant) adaptation capacities to ensure the future of livestock farming and secure food supply.

Georeferencing breed distribution systems allow information on Animal Genetic Resources (AnGR) to be linked to multiple databases. Thus, data related to AnGR can be used to analyse animal production systems and their evolution over time, correlated with changes in the environment (for example, weather or disease epidemiology) and where and how the AnGR can be used under what conditions. This also allows for the use of the information in landscape genetics studies, an interdisciplinary research area that combines population genetics, landscape ecology, and spatial statistics (Storfer et al 2007). This area of study aims to describe and explain how the landscape and its components affect the genetic variation of populations of animals and plants. These methods can increase our knowledge of how landscape heterogeneity influences the population structure, gene flow, and adaptation. It is important to emphasise that evolution and adaptation processes, drift, and dispersion occur geographically.

More recently, Resende et al (2021) have proposed the application of "enviromics" to breeding practice, by which the similarity among sites assessed on an "omics" scale of environmental attributes drives the prediction of unobserved genotype performances. They suggest that this will allow for (1) accurate matching of sites to their most appropriate genotypes, (2) better definition of breeding areas that have a high genetic correlation to ensure selection gains across environments, and (3) efficient determination of the best sites to carry out experiments for further analyses.

3. Geographical tools

The regions and countries predicted to prosper in a climate-changed world (Llewellyn, 2007) will tend to be those that recognise its importance and inexorability, foresee that there may be at least some implications for their industry (including farms), and take appropriate steps well in advance. O'Neill et al (2010) argue that the knowledge of interactions between genotype, environment, and management in the livestock systems is required to generate genotypes for efficient livestock production that are economically and environmentally sustainable. Farmers should reduce their reliance on infrastructure and veterinary products to alleviate environmental stress and more on the animal's ability to achieve fitness in a given production environment (Prasad and Sejian 2015). Aside from purely biological aspects, the systems that favour animal adaptation to environmental stress will require new methods of management and management of production resources. Therefore, the combination of technologies and human resources training to operationalise these are key factors to minimise the risks of new production forms.

Establishing an economically viable production system in a given region requires the choice of breeds or genetic groups suited to local environmental conditions (Daltro et al 2017). Gaining a more comprehensive understanding of the dynamics of adaptive genomic attributes across and among collections of organisms coexisting in the same ecosystems and coupling that data with geographic information technologies has led to the emergent field of landscape genomics (Pariset et al 2012). Landscape genomics is the spatially explicit study of geographic patterns of genome-wide genetic variation (Sork et al 2013). Numerous efforts have started to explore the use of environmental descriptors (e.g., Scherf 2008, Scholtz et al 2010, McManus et al 2021a) to better facilitate genetic evaluation with environmental interactions. The results may also reveal genetic adaptation to specific environmental stressors such as diseases, parasites and extreme heat, vegetation type, lack of water, or combinations.

Costa et al (2014) found that breeding values for type traits and milk production differed according to environmental factors such as rainfall and temperature, indicating that these factors influence the distribution of genetic values of dairy traits within Brazil, while McManus et al (2016) looked at the dynamics of cattle production in Brazil reflecting changes in environmental factors, such as pasture type, temperature, and humidity. McManus et al (2014b and b) found that breed distribution in sheep and goats was also linked to environmental factors. Carvalho et al (2012) investigated the use of Google Maps to locate rural properties, while Hermcuhe et al (2012, 2013a,b,c) studied the dynamics of sheep, cattle, and goats (McManus et al 2014a,b,c) production and their relation to environmental factors. The use of GIS was also used to predict expansion areas for the production and exportation of products and infrastructure (de Sá et al 2018). Disease risk has also been an essential question in GIS studies with livestock (Bajardi et al 2012).

Most landscape genetics analyses are based on genetic distance, and inaccurate estimates of allele frequencies can influence gene flow estimates (Storfer et al 2007, 2010). Although replication of sites or transects is favoured for the reasons above in landscape genomics studies, the balance between sample size and the number of locations depends on the type of downstream analysis. In landscape genomics studies, the detection power is limited by the total number of samples collected (Lotterhos and Whitlock, 2015). It is important to sample a sufficient number (for example, > 10) of individuals per location to generate accurate allele frequency estimates for analyses based on genetic differentiation estimates between populations. However, optimising the number of population pairs sampled (with smaller sample sizes per site) can be robust for detecting selection when the sampling sites represent a range of environmental variable values in the study area (De Mita et al 2013). Landscape genomics has emerged as a prominent framework for studying the genomic basis of local adaptation. Researchers use large genomic data sets to examine the genome for loci that exhibit selection signatures in heterogeneous environments.
(Haasl and Payseur 2016). These efforts have been very successful, for example, in identifying the genes underlying adaptation to hypoxia in high-altitude human populations (Beall 2007a, b, Simonson et al 2010), environmental responses in oak populations throughout climatic gradients (Sork et al 2016) and differences in growth response between salmon populations in response to geological conditions (Vincent et al 2013). Studies of biotic factors have also successfully identified local adaptation to characteristics of life history (Sun et al 2015), community composition (Harrison et al 2017), and disease prevalence (Leo et al 2016, Mackinnon et al 2016, Wenzel et al 2016). Landscape genomics has already helped dramatically deepen our understanding of adaptation’s genomic basis (Funk et al 2012, Shryock et al 2015).

Understanding spatial dependence and distribution of agricultural production factors is crucial for territorial planning and regional development. Therefore, the spatial-temporal analysis provides an opportunity to understand the factors that control agricultural development, apply resource application strategies, and reduce social and environmental impacts. In this regard, the adoption of Geographic Information Systems (GIS) and Spatial Statistics have helped establish spatial data processing procedures. The dynamics of crop and animal products can also be studied (McManus et al 2016) both alone and in association with each other (Maranhão et al 2019).

3.1. Spatial Statistics

Spatial statistics support the analysis of geographic data by techniques that describe the distribution of data in the geographic space (descriptive spatial statistics), analyse the spatial patterns of the data (spatial pattern analysis), identify and measure spatial relationships (spatial regression), and create a surface from sampled data (spatial interpolation, usually categorised as geostatistics).

Genetic processes often create spatial structures. Therefore, using spatial information in genetic data analysis, we can test spatial patterns, detect spatial arrangement scales, describe spatial genetic systems, and infer the patterns’ directionality (migrations). Several statistical measures can be used to test these structures.

i) Island/hierarchical island model.

In these models, there are patches of related genotypes. In the island model (Latter, 1973), a species is subdivided into several discrete finite populations, races or subspecies, between which some migration occurs. The hierarchical island model (Slakin and Voelm 1991), on the other hand, has populations within a neighbourhood that exchange migrants at a much higher rate than do populations in different neighbourhoods. Wright’s hierarchical F statistics can indicate the extent of gene flow within and between these.

ii) Isolation by distance (IBD): clines of genetic differentiation

According to Wright (1943), complete continuity of distribution, but interbreeding is restricted to small distances by only short-range means of dispersal, is opposite from the island model. Remote populations may become differentiated merely from isolation by distance. Wright used inbreeding coefficients to study these relationships.

iii) Inbreeding avoidance: repulsion structure

Crossing with related individuals can reduce individual fitness through the deleterious effects of inbreeding depression (Nelson-Flower et al 2012). Inbreeding avoidance is expected to have developed in most species, especially in cooperatively breeding species where individuals may delay dispersal until long after sexual maturity. Such potential mechanisms include sex-biased dispersal and avoidance of kin known through associative learning.

3.1.1. Spatial Autocorrelation

Spatial autocorrelation is defined, in general, as values of a variable that are not independent of the corresponding spatial locations. In genetics, genetic distance is correlated to spatial distance. There can be two types of spatial autocorrelation: positive: closer individuals are more similar than at random, negative: closer individuals are more dissimilar than at random. Therefore, we can compute spatial autocorrelation (Moran’s I, Mantel’s M) for different distance classes to determine at what distance differences between populations occur (Silva 2015), thereby directing conservation measures.

iv) Moran’s Index

Moran’s index is a measure of spatial autocorrelation. A correlation characterises spatial autocorrelation among nearby locations in space. It is more complicated than one-dimensional autocorrelation because spatial correlation is multi-dimensional (i.e., 2 or 3 dimensions of space) and multi-directional.

v) Mantel’s correlation

This is a statistical test of the correlation between two matrices. Geographical distances compared with genetic divergence or genetic distances, estimated by pairwise FST and related statistics, by the Mantel test is one of the most popular approaches to evaluate spatial processes driving population structure (Diniz-Filho et al 2013). For example, the correlation of
genetic and geographical distances using the Mantel test for sheep in Brazil was -0.0485. The probability of observing a correlation greater than or equal to observed was P = 0.873 and less than or equal to observed = 0.127. This is probably due to high individual variation within breeds and locations sampled. Pariset et al (2011) studied sheep in Albania, Greece and Italy, using mtDNA and SNP data, and found no geographic patterns. On the other hand, a Mantel test revealed a correlation between genetic and geographic distance.

**vi) Correlogram**

The correlation of a time series with its past and future values is called autocorrelation. It is also referred to as "lagged or series correlation". Positive autocorrelation indicates a specific form of "persistence", the tendency of a system to remain in the same state from one observation to the next. If a time series exhibits correlation, the samples' future values probabilistically depend on the current and past samples. Autocorrelations can be exploited in prediction as well as modelling time series. Canonical correspondence analysis (CCA) has been used as an alternative because it enables genetic diversity to be related to environmental factors and can test for environmental factors that contribute significantly to explaining genetic diversity variations (Angers et al 1999).

**vii) Multivariate Analysis and Spatial Patterns**

Principal Components Analysis (PCA) is widely used in remote sensing for multivariate data's underlying dimensions (Fung and Ledrew 1987). PCA reduces the number of variables of the complex data set into fewer variables but maintains maximum variance, whereby linearly transformed components from the original data are not correlated (Hermuche et al 2012). The first principal component (PC1) can be described as maximising the original data set's variance. PC2 contains other areas of significant change, with other PCs becoming increasingly less significant changes. Low-order principal components include a lot of noise (error) and are usually not used in the analysis. Usual multivariate analyses ignore spatial information and may reveal apparent spatial structures but overlook more subtle patterns. Therefore, there is a need for taking spatial information into account, which can be done using spatial PCA analyses.

Cavalli-Sforza et al (1994) stimulated principal component analysis (PCA) to population genetic variation. They collected count data for many genetic variants ("alleles") from population samples at many geographic locations, produced an allele-frequency map for each allele and a spatially interpolated map representing variation in allele frequency across space. PCA was then used to distil the many allele-frequency maps into a smaller number of "synthetic maps," which we refer to as PC maps for brevity.

Self Organising Features Map (SOFM) is a type of Artificial Neural Network (ANN) with non-supervised training, which may be described as a nonlinear, ordered, and smooth mapping of high-dimensional input data domains onto the elements of a regular, low-dimensional array (Kohonen, 1988).

Hermuche et al (2012) analysed sheep production data in Brazil over time using principal components and self-organising maps. This made it possible to identify municipalities with sheep production data constant over time. Two distinct regions of stability were identified – in the Northeast and South, due to the tradition of production in these two regions. The classification of municipalities from the Kohonen method showed the dynamics of production growth from the different maps generated. Thus, it can be inferred that the increase in the size of the maps increases the ability to detect changes in the most sensitive production regions. New municipalities were added, with a trend towards the inclusion of municipalities of the Midwest with incipient growth. The evaluation of the curves of the centroids of production of each group throughout the analysis indicates this production's behaviour. This type of information can help to guide political decisions as to infrastructure requirements for production systems (p.ex. roads, slaughterhouses).

It is also possible to carry out an Allele aggregation index for alleles (AAIA) tests the null hypothesis that each allele at a locus is distributed randomly across a landscape (i.e., no aggregation or genetic structure) relative to the aggregation of the actual organisms sampled for analysis purposes. For sheep in Brazil, this index averaged over all alleles (\(R_{\text{ave}}\) = 1.225, \(P \left( R_{\text{ave}} \leq \text{RND} \right) = 0.181\). This indicated that there are genetic clusters for this species, and so there are nonrandom patterns of the spatial distribution of genes in the landscape. This is useful when building conservation and sampling programs for farm animals (McManus et al 2021b).

More recently, mapping genetic marker variability in both locally adapted sheep (McManus et al 2020b), cattle (Silva 2015, Souza et al 2022, Costa et al 2020), and pigs (Cesconeto et al 2017) showed spatial patterns for marker distribution, which may lead to the identification of specific markers for heat tolerance and other adaptive traits. These are easily identified in the 3D surface spatialised genetic distances for sheep in Brazil (Figure 1) and can be superimposed on surface maps to see where the divisions occur and aid in conservation program collection strategies.

### 3.2. Inferring directionality

Directionality can be inferred from reduced variability as the population becomes distanced from its origin, usually due to bottlenecks (Hamblin et al 2006). Genetic drift can cause significant losses of genetic variation in small populations. Population bottlenecks occur when a population's size is reduced for at least one generation.
3.2.1. Bayesian clustering – assignment tests

Bayesian clustering approaches are derived from traditional assignment tests that use individual multi-locus genotypes to assign individuals to populations (François and Durand 2010). These authors alert that current models fall into two broad categories, those with (TESS (Durand et al 2009), BAPSS (Corander et al 2008), STRUCTURE (Pritchard et al 2000), or without (GENELAND (Guillot et al 2005)) admixture, which substantially differ in their prior distributions and background assumptions/models. Models without admixture are generally not robust to the inclusion of admixed individuals in the sample, thus providing an incorrect assessment of population genetic structure in many cases. In contrast, admixture models are robust to an absence of admixture in the sample.

3.2.2. Wombling or Monmonier algorithm

Wombling uses the multi-locus genotypes and spatial locations of individual samples to identify populations (Crida and Manel, 2007). It determines the number of populations present in the study area and their boundaries. Monmonier (Monmonier 1973) also uses allele frequency data from pre-defined sampling sites to divide the study area into groups. However, like Wombling, it utilises spatial data. The Monmonier algorithm does not include a mechanism for determining how many groups should be defined (Figure 2). Instead, the number of groups is an input to the algorithm, determining the boundary location(s) (Martein and Gregovich 2008). For example, Souza et al (2022) used this method to determine minimum sampling distances for native cattle breeds in Brazil.
3.3. Phylogeography

Studies in phylogeography look at historical processes responsible for the present-day distribution of native breeds of animals, using geographic distribution and population genetics (Avise et al 1987, Marske 2016). While most studies are carried out with wild species, they have more recently included domestic species. Most studies (>80%) also use mtDNA as this evolves rapidly in populations of higher animals and usually is transmitted maternally without intermolecular recombination (Avise 1998). Many studies have been carried out with sheep (Liu et al 2020 (Tibet), Campos et al 2020 (Mexico), Amane et al 2020 (Ethiopia), McManus et al 2010 (Brazil)) and goats (Al-Araimi et al 2017 (Arabian), Joshi et al 2004 (India)), worldwide, among many others. At least five large mtDNA haplotypes were found in the sheep (Meadows et al 2005) and six in goat species (Colli et al 2015). This information is used to infer the origin and possible climatic adaptations in breeds.

Other markers can also be used, such as microsatellites (Ciani et al 2013) in Italy, where a clear geographic cline was observed when investigating the genetic relationships among breeds. SNPs have also been used where Kijas et al (2009) found sheep are characterised by weak phylogeographic structure, overlapping genetic similarity, and generally low differentiation consistent with their short evolutionary history. However, the degree of population substructure was sufficient to cluster individuals based on geographic origin and known breed history.

3.4. Selection Signatures

Selection signatures of local adaptation can be examined using whole-genome data and enabling spatial association assessment using molecular markers. Candidate loci to adaptation are identified by assessing genome-environment associations (Stucki et al 2017). Measuring Local Indicators of Spatial Association (LISA) for these candidate loci makes it possible to detect whether similar genotypes tend to gather in space, which constitutes a useful indication of the possible kinship relationship between individuals (Cesconeto et al 2017).

Paim et al (2018) described the main statistical methods and software currently used to analyse genomic data and the identification of selection signatures. Cesconeto et al (2017) identified selection signatures in Brazilian swine and correlated these with the monthly variation of Brazilian environmental variables using BayeScan and SamBada software (Stucki et al 2017). Global spatial territory correlation for environmental variables corroborates this finding (average Moran's I = 0.89, range from 0.55 to 0.97). The distribution of alleles over the territory was not strongly correlated with the breed/genetic groups, probably due to crosbreeding.

4. Concluding remarks

According to Avise (1998), integration with geographic data need not be limited to genetic data but may include morphological and production traits. There are few studies in Brazil combining geographical and genetic information. These studies can help us better understand genetic x environment interactions and identify possible markers associated with adaptation to specific environments. While these studies are common in ecology studies, they are still livestock terms. Leempoel et al (2017) give some basic instructions on how these types of studies can be carried out effectively, looking at sampling, GIS systems, software, and datasets. Joost et al (2010) looked at methods to integrate different data sets and highlight problems related to interdisciplinary comparisons. This study area shows a promising future and can help build more consistent research, breeding, and political policies toward livestock farming worldwide.

Conflict of Interest

The authors declare that there is no conflict of interest.

Funding

Thanks are due to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (finance code 001).

References


