Universitat Autònoma de Barcelona

GENOMIC EVALUATIONS OF BEEF CATTLE BREEDS IN TROPICAL CLIMATES

Lino César Ramírez Ayala

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons: http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons: http://es.creativecommons.org/blog/licencias/

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license: https://creativecommons.org/licenses/?lang=en

UNIVERSITAT AUTÒNOMA DE BARCELONA

Departament de Ciència Animal i dels Aliments, Facultat de Veterinària

CENTRE DE RECERCA EN AGRIGENÒMICA

Departament de Genètica Animal

GENOMIC EVALUATIONS OF BEEF CATTLE BREEDS IN TROPICAL CLIMATES

Lino César Ramírez Ayala

Doctoral thesis to obtain the PhD degree in Animal Production of the Universitat Autònoma de Barcelona, July 2020

Supervisors

Dr. Miguel Pérez Enciso **Dr. Yuliaxis Ramayo Caldas**

El **Dr. Miguel Pérez Enciso**, investigador ICREA del Departament de Ciència Animal i dels Aliments de la Universitat Autònoma de Barcelona (UAB), i el **Dr. Yuliaxis Ramayo Caldas**, investigador de l'Institut de Recerca i Tecnologia Agroalimentàries (IRTA),

fan constar,

que **Lino César Ramírez Ayala** ha realitzat sota la seva direcció el treball de recerca

"Genomic evaluations of beef cattle breeds in tropical climates"

i certifiquen

que aquest treball s'ha dut a terme al Departament de Ciència Animal i dels Aliments de la Facultat de Veterinària de la Universitat Autònoma de Barcelona i a la unitat de Genètica Animal del Centre de Recerca en Agrigenòmica (CRAG)

considerant

que la memòria resultant és apta per optar al grau de Doctor en Producció animal per la Universitat Autònoma de Barcelona,

i perquè quedi constància, signen aquest document a

Bellaterra, a 24 de juliol de 2020.

Dr. Miguel Pérez Enciso Lino César Ramírez Ayala Dr. Yuliaxis Ramayo Caldas

Cover designed by Dr. Jordi Leno Colorado.

Lino César Ramírez Ayala was financially supported by “Don Carlos Antonio López” National Grant Program (BECAL) from Paraguay.

In memory of my father, Aldo Luis (1948 – 1994)
Dedicated to my mother Zulmina Edit, my sisters Mirta Edith, Liz Teresita, Lourdes Angelina, my brother Víctor Aldo, my nieces, Jazmín de María, Gema Teresita, my nephews, José Daniel, Juan Miguel, my brother-in-law Luis Fernando and especially to my great loves Mónica Carolina and Sol Abigail

“Reiko va'erã remanótarõ ĝuarãicha,

ha remba'apo va'erã remano'ỹtarõ ĝuarãicha”

— Ñe'ẽnga Guaraníme

CONTENT

SUMMARY

The largest number of cattle heads is concentrated in the so-called tropical and subtropical regions. These are characterized by special and specific environments, so understanding how cattle adapt to this type of ecosystem to survive and produce is important. For many years, several strategies have been implemented to genetically improve cattle in the tropics in order to find the best techniques that efficiently combine adaptation and production. The main objective of this thesis is to evaluate the effects of adaptation and selection in beef cattle in tropical climates, as well as their potential genetic gains through different strategies of genetic improvement.

In a first study, we used the whole-genome sequences of a total of 12 samples from the Chacuba population (CHCU) and 60 samples from six other breeds of taurine, zebu and cross-breeds to estimate the genetic diversity, structure and precise ancestral origin of the CHCU animals. Although these animals were assumed to be a closed population, closeup analysis indicates limited introgression of *Bos indicus*, probably due to a single, non-continuous introgression event. The extended haplotype homozygosity test (EHH) was used to identify regions that may have played an important role in adaptation to tropical conditions. Regions with a high percentage of zebu were enriched in possible selective events, but only slightly and adaptation cannot be explained by the influence of zebu breeds alone. EHHs revealed signs of possible adaptation that included genes involved in thermogenesis (*UCP1*, *DIO2* and *ACSL1*) and hair development (*BMPR1A*, *CDSN* and *FGF7*). We also identified variants within these genes that may have a functional impact and could thus explain some of the phenotypic differences observed between the CHCU and the French Charolais breed.

In a second study, we developed a flexible generic advance simulator, called SeqBreed, in order to optimize genomic prediction (GP) or genome-wide association studies, incorporating some of the most popular GP methods such as genomic best linear unbiased prediction (GBLUP), single-step GBLUP, pedigree-based BLUP, and mass selection and including several visualization tools.

Finally, in a third study, we used the SeqBreed software to compare three *Bos indicus* x *Bos taurus* cross-breeding programs: F1, grading up or backcrossing and rotational crosses. We simulated, using real SNP data of zebus and taurines, phenotypes of three traits of utmost importance in terms of productivity, mainly in tropical production systems based on cattle breeding: shear force, growth and tolerance. The accuracy of the prediction was compared between three 50k chips that differed in the way the SNPs were chosen: (i) randomly, (ii) with a minimal difference in allele frequency between the breeds, and (iii) with a minimal difference in allele frequency between the breeds as long as the allele frequency was greater than 0.09 on the *Bos taurus*. We found that the rotational crossing system is the optimal one in terms of predictive accuracy, and that the selection of markers based on the differences in allele frequency between the races does not compensate and is even detrimental

.

RESUMEN

La mayor cantidad del censo ganadero se encuentra concentrado en regiones tropicales y subtropicales. Éstas se caracterizan por presentar ambientes especiales y específicos, por lo que comprender cómo el ganado vacuno se adapta a ese tipo de ecosistema para sobrevivir y a la vez producir es muy importante. Desde hace muchos años se vienen utilizando diversas estrategias que buscan mejorar genéticamente a los vacunos con el fin de encontrar las mejores técnicas que combinen de manera eficiente la adaptación y la producción.

El objetivo principal de esta tesis es evaluar los efectos de la adaptación y la selección en el ganado bovino de carne en los climas tropicales, ya sea adaptado, introducido o mejorado, así como sus posibles ganancias genéticas a través de diferentes estrategias de mejoramiento genético.

En un primer trabajo, utilizamos la secuencia del genoma completo de un total de 12 muestras de la población Chacuba (CHCU) y 60 muestras de otras seis razas de origen taurino, cebú y cruzados para estimar la diversidad genética, la estructura y el origen ancestral preciso de los animales CHCU. A pesar de que se suponía que estos animales eran una población cerrada, el análisis con el software ADMIXTURE indica una introgresión limitada procedente de *Bos indicus*, probablemente debido a un evento de introgresión único y no continuo. El test de extensión de homocigosis de haplotipos (EHH) se utilizó para identificar las regiones que pueden haber tenido un papel importante en la adaptación a las condiciones tropicales. Las regiones con un alto porcentaje de cebú están enriquecidas en posibles eventos selectivos, pero sólo ligeramente y la adaptación no puede explicarse sólo por la influencia de las razas cebúes. El test EHH revela señales de posible adaptación en genes implicados en la termogénesis (*UCP1*, *DIO2* y *ACSL1*) y el desarrollo del pelo (*BMPR1A*, *CDSN* y *FGF7*). También identificamos variantes dentro de estos genes que pueden tener un impacto funcional y que podrían explicar así algunas de las diferencias fenotípicas observadas entre el CHCU y la raza Charolais francesa.

En un segundo trabajo, desarrollamos un programa de simulación, que denominamos SeqBreed, con el fin de optimizar la predicción genómica (GP) o los estudios de asociación de todo el genoma, incorporando algunos de los métodos más populares de GP como pueden ser la mejor predicción genómica lineal insesgada (GBLUP), GBLUP de una sola etapa (*single step*), BLUP basado en el pedigrí y selección masal e incluye varias herramientas de visualización.

Por último, utilizamos el software SeqBreed para comparar tres estrategias de cruzamiento entre *Bos indicus* x *Bos taurus*: F1, clasificación o retrocruza y cruces rotativos. Simulamos, utilizando datos reales de SNP de cebúes y taurinos, fenotipos de tres rasgos de suma importancia en términos de productividad, principalmente en los sistemas de producción tropical: fuerza de cizallamiento (*shear force*), crecimiento y tolerancia a parásitos y calor. Se comparó la precisión de la predicción entre tres chips de 50k que diferían en la forma de elegir los SNP: (i) aleatoriamente, (ii) con una diferencia mínima en la frecuencia de los alelos entre las razas, y (iii) con una diferencia mínima en la frecuencia de los alelos entre las razas siempre que la frecuencia de los alelos fuera superior a 0,09 en el *Bos taurus*. Encontramos que el sistema de cruce rotacional es el óptimo en términos de precisión de predicción, y que la selección de marcadores basados en las diferencias de frecuencia alélica entre las razas no compensa e incluso es perjudicial.

LIST OF FIGURES

GENERAL INTRODUCTION

PAPER I

PAPER II

PAPER III

ANNEXES: Paper I

ANNEXES: Paper III

LIST OF TABLES

LIST OF PUBLICATIONS

The present thesis is based on the work contained in the list of articles below:

- Paper I. **Ramírez-Ayala, L.C.**, Rocha, D., Ramos-Onsins, S.E., Leno-Colorado, J., Charles, M., Bouchez, O., Rodríguez-Valera, Y., Pérez-Enciso, M., Ramayo-Caldas, Y. (2020) 'Whole-genome sequencing reveals insights into adaptation of French Charolais cattle to Cuban tropical conditions'. *Genetics Selection Evolution* (submitted).

- Paper II. Miguel Pérez-Enciso, **Lino C. Ramírez-Ayala**, Laura M. Zingaretti. (2020) 'SeqBreed: a python tool to evaluate genomic prediction in complex scenarios' *Genetics Selection Evolution* **52**, 7.

- Paper III. **Ramírez-Ayala, L.C.**, Razmkabir, M., Leno-Colorado, J., Zingaretti, L.M., Ramayo-Caldas, Y., Pérez-Enciso, M. (2020). 'Evaluation genomic prediction strategies in crossbred *indicine* x taurine cattle programs'. (in preparation).

ABBREVIATIONS

θ	Watterson's Nucleotide Variability
ACSL1	Long-chain Acyl-CoA Synthetase 1
ANG	Angus
A.R.P.	Asociación Rural del Paraguay
BLUP	Best Linear Unbiased Prediction
BMPR1A	Bone Morphogenetic Protein Receptor Type 1A
BRG	Brangus
BRH	Brahman
BRM	Brahman
BSP	Brangus + Senepol
CDSN	Corneodesmosin Precursor
CHCA	Canadian Charolais
CHCU	Chacuba
CHFR	French Charolais
DGRP	Drosophila Genome Reference Panel
DIO2	Deiodinase Iodothyronine Type II
DM	Dry Matter
EHH	Extended Haplotype Homozygosity
EBV	Estimate Breeding Value
FGF7	Fibroblast Growth Factor 7
FST	Differentiation Values
GBV	Genomic Breeding Values
GBLUP	Genomic Best Linear Unbiased Prediction
GEBV	Estimated Genomic Breeding Value
GFRA2	GDNF Family Receptor Alpha 2
GP	Genomic Prediction
GS	Genomic Selection
GWAS	Genome-wide Association Studies
HFD	Hereford
IND	*B. indicus*

LD	Linkage Disequilibrium
LIMS	Limousin
MAF	Minimum Allele Frequency
MAS	Marker Assisted Selection
NEL	Nelore
PA	Prediction Accuracy
PAR	Pseudo-autosomal Region
PCA	Principal Component Analysis
pSBVB	Polyploid Sequence Based Virtual Breeding
QTL	Quantitative Trait Loci
QTN	Quantitative Trait Nucleotides
RGU	Red Angus
SNP	Single Nucleotide Polymorphism
ssGBLUP	single-step GBLUP
THI	Temperature and Humidity Index
TXL	Texas Longhorn
SBVB	Sequence-Based Virtual Breeding
SEN	Senepol
SGT	Santa Gerdtrudis
UCP1	Uncoupling Protein 1

Chapter 1

General Introduction

General Introduction

1.1. Genetic basis for adaptation to heat and tropical climates

Almost half of the world's population lives in tropical and subtropical climates. These climates are characterized by high temperatures throughout the year, persistence of ecto and endoparasites, and periods of food scarcity due to seasonal patterns of rainfall and, consequently, seasonal variation in pasture quality and availability (Burrow, 2001; Carvalheiro *et al.*, 2019). All of these characteristics place a heavy burden on livestock, especially for species that originated in more temperate climates, such as dairy or beef cattle. The current climate change will exacerbate these problems; however, climate change also provides an opportunity to study the genetic basis of adaptation to heat or others challenges faced under tropical conditions.

1.1.1. Heat stress

Heat stress is the sum of external forces acting on livestock that cause a rise in body temperature with a consequent animal response. It is one of the most damaging factors contributing to reduced growth, production, reproductive output, milk quantity and quality, and natural immunity, making animals more vulnerable to disease (Berihulay *et al.*, 2019). Livestock responses to this stress vary according to breed, production level, feed quantity and quality, and overall health status. This phenomenon occurs when the animal tries to dissipate heat unsuccessfully or is overwhelmed, and as a result performance or health is affected.

Heat also tends to reduce livestock performance by inhibiting appetite (Polsky & von Keyserlingk, 2017). Feed intake by the animal is directly related to all aspects of energy metabolism, with heat release for maintenance, activities and production. Total heat production in the animal depends, in part, on dry matter intake. High feed consumption increases the metabolic rate and water intake, resulting in higher effort for heat loss. When the ambient temperature increases above the thermoneutral zone, defined as the thermal environment where the animal experiences optimal health and maximum productivity (Ames, 1980), the metabolic rate also rises as a result of the increment in body temperature. To balance heat production with heat loss, the animal reduces its feed intake. This causes a

reduction in the metabolic rate and consecuently, decreases its mantainance. In addition to ambient temperature, humidity and air movement also contribute to the occurrence of heat stress in animals, being more severe when ambient temperature together with relative humidity are higher than 78°F (25.5ºC) (Habeeb *et al.*, 2018). A part from these daytime conditions, nocturnal conditions, such as minimum wind speed, minimum solar radiation, and minimum temperature and humidity index (THI), also influence heat stress, because livestock can often dissipate considerable heat during the night if temperatures are low (Mader *et al.*, 2006). THI is an index that measures ambient temperature, relative humidity, and evaporation rate together. This allows an assessment of an environment's potential for producing heat stress in farm animals (Dikmen & Hansen, 2009; Aggarwal & Upadhyay, 2013). Habeeb *et al.* (2018), studying these parameters in dairy cattle, identified that when THI exceeds 72, cows are likely to begin to experience heat stress, affecting their calving rates. When THI exceeds 78, milk production will be seriously affected, and above 82, very significant losses are likely to occur.

Heat stress also affects reproduction in both sexes. In males, Lees *et al.* (2019), mention several findings that studied how heat load, either through scrotal isolation or whole-body exposure, adversely affects spermatogenesis and/or viability of sperm. Additional studies have evaluated the effect of scrotal temperature and body temperature in Wagyu bulls. Scrotal temperatures were remotely monitored while the bulls were placed through a series of heat load regimes (Wallage, Gaughan, *et al.*, 2017; Wallage, Johnston, *et al.*, 2017). These studies highlighted that mechanisms that were thought to maintain scrotal temperature begin to break down during periods of heat loading. Meanwhile, in females, heat stress can affect tissue and organ function and alter the formation of proteins and hormones, which in turn can lead to low fertility by affecting the synthesis of proteins and hormones associated with the reproductive organs. The heat generated by cows producing 30 kg/day of milk is double than the maintenance heat production of non-lactating cows and high-yielding cows (producing 55 kg/day of milk) is approximately three times higher than maintenance heat production. It has been shown that body temperatures of high-yielding dairy cows in a humid region start to increase exponentially when air temperatures reach 26-27°C. Therefore, even a small increase in air temperature of the order of 1-2°C can induce severe hyperthermia in dairy cows (Wolfenson & Roth, 2019).

1.1.1. Persistence of parasites

Tropical climates also face other challenges as the persistence of parasites (Scasta, 2015). Livestock production losses due to external parasites have long been a major concern for beef cattle producers in tropical and subtropical regions. They can be grouped into two general categories according to their place of action, internal and external. Reduced voluntary feed intake is a common phenomenon observed in parasitic infections. Gastrointestinal parasites normally reduce the availability of nutrients to the host animal, both by reducing voluntary feed intake and by decreasing the efficiency of absorbed nutrients (Coop & Kyriazakis, 1999). Forbes *et al.* (2000), found that heifers naturally infected with parasites during grazing grazed for a shorter time per day, i.e. 105 minutes less, compared to animals treated with anthelmintic, ten weeks after participation in grazing. In addition, voluntary feed intake showed that infected animals consumed 0.78kg less of dry matter (DM) grass when compared to non-infected animals. Interestingly, these changes were not evident until two weeks after pasture participation. This may be because the heifers in this study were assumed to be naturally infected with parasites through grass intake and therefore needed to consume sufficient amounts of grass before clinical signs could appear. Reduced grazing time is assumed to be associated with reduced appetite and this in turn is common with parasitic infections.

Among the ectoparasites we can mention the tick, which has limited the expansion of production of pedigree and mongrel cattle such as those of British origin. The use of these breeds would be a strategy to improve meat quality through zebu populations in cross-breeding and synthetic breed generation systems. The feasibility of this alternative is largely limited by the higher susceptibility of *Bos taurus* (from now on *B. taurus*) tick infestation compared to *Bos indicus* (from now on *B. indicus*), while the *indicine* x *taurine* cross has an intermediate susceptibility (Frisch, 1999). It has been verified that tick resistance has a genetic basis, which implies the viability of improving this trait with specific breeding programs (Sollero *et al.*, 2014). Therefore, knowledge on the resistance level in each breed, among different breeds and individuals crossed becomes an important alternative to improve the production of beef cattle.

1.1.2. *Bos taurus* x *Bos indicus*

Climate change will mainly affect cattle raised outdoors, as most beef cattle. Beef cattle breeds can be divided into temperate zone bulls of European origin, African taurine breeds and Indian zebu breeds (Figure.1.1) and various mixed populations (Pitt *et al.*, 2019; Paim *et al.*, 2020).

Figure 1.1. *B. taurus* Angus cow (left) and *Bos B. indicus* Brahman bull (right). Photo credits: Sociedad de Criadores de Angus del Uruguay website (https://angusuruguay.com/fotos/) and adapted from Sutarno & Setyawan (2016), respectively.

Zebu cattle have developed specific attributes and functions to adapt to a wide range of environmental constraints in tropical climates, such as modified skin, hair, sweat, endocrine profile, fat restricted to specific parts of the body, increased heat flow from the body to the skin and reduced metabolic rate. Negative physiological effects on cow productivity, such as high respiration rates and increased blood flow, are more noticeable in *taurine* cattle than in the *indicine*, so the consequences of heat stress exposure in milk and meat production are less pronounced in the latter. This explains the significant variability between *B. indicus* and *B. taurus* cattle in their responses to heat stress, which should be explored to better understand the genetic basis of heat adaptation in cattle (Morenikeji *et al.*, 2020). *Indicine* crosses with *taurine* cattle show a wide range of intermediate types suggesting that these traits have a multifactorial genetic basis (Porto-Neto *et al.*, 2018; Tesema *et al.*, 2019). In recent years, research has been conducted on the molecular genetic basis of a wide range of traits of interest in livestock production. This has included studies to detect Quantitative Trait Loci (QTL), which are loci that explain some of the variation in traits of interest, develop

increasingly dense genome maps as well as studies on the association between molecular genetic markers and traits of interest (Haskell *et al.*, 2014).

Figure 1.2 shows that, like many other traits, heat and parasite resistance are highly polygenic traits and may be amenable to standard genetic improvement procedures. The main problem, however, would be to obtain reliable phenotypic measurements.

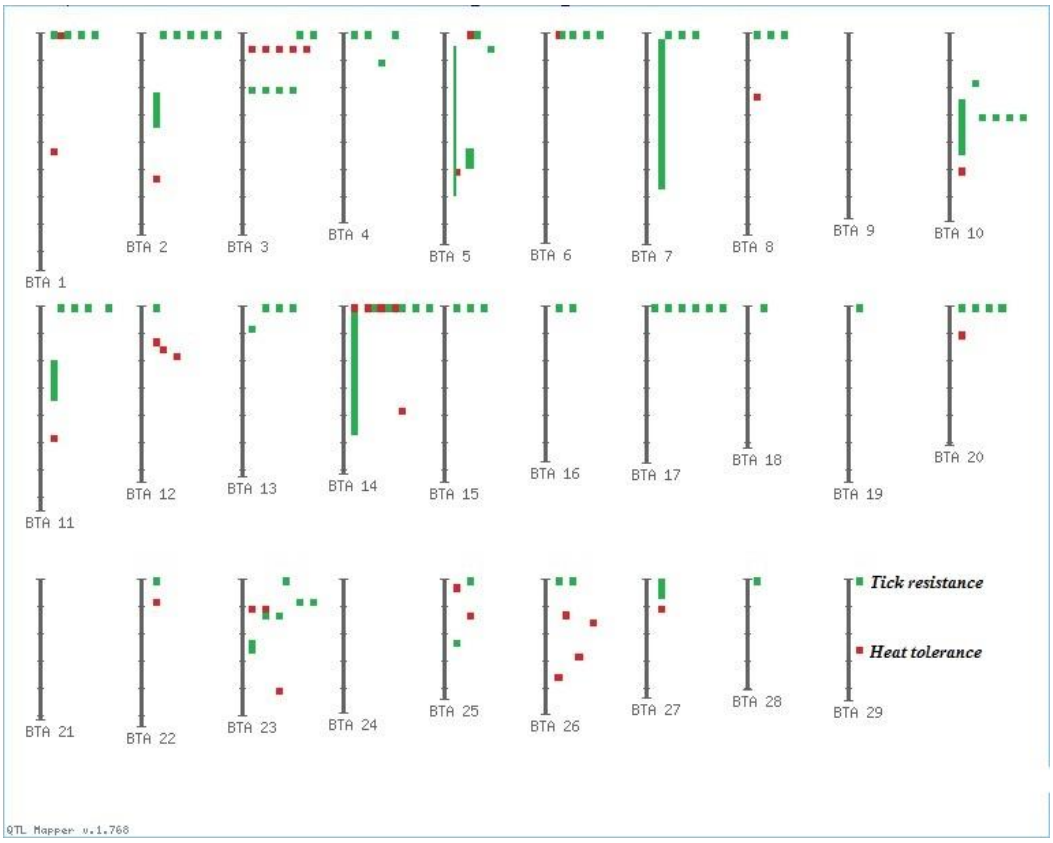

Figure 1.2. Graphical representation of QTL mapped on the cattle autosome for tick resistance (green) and heat tolerance (red) traits.

1.2. Genetic improvement of cattle in tropical climates

Genetic improvement results from selecting above-average candidates as parents for the next generation. This improvement accumulates through breed replacement, crossbreeding, and selection within breeds, with the latter resulting in cumulative genetic progress. Genetic improvement has been increasingly based on sophisticated statistical methods, including mixed model methodology, to provide accurate estimates of individual breeding values

(EBVs). Dramatic genetic improvements have been achieved in several species by combining within-breed selection with reproductive technologies (such as artificial insemination and embryo transfer) to more effectively disseminate elite genomes (Georges *et al.*, 2019).

The diversity of ecological zones and production systems that characterize tropical environments infers that no single breeding system or genotype will be fully efficient for all situations, nor its effects on the interaction between the genotype and the environment. With this situation, cattle breeding strategies change over time in an attempt to optimize production levels. Breeders often face the challenge of identifying and developing genotypes or genetic combinations that maintain a balance between production and adaptation (Setshwaelo *et al.*, 1990). Using genetic variation for adaptation to tropical climates among and within breeds is one option for making genetic improvements in livestock, thus seeking a sustainable solution to constant climate change. Breeding and selection are two tools available to breeders to explore genetic variation between and within breeds, respectively (Ghebrewold, 2018). These strategies help to increase the frequency of favourable genes that enable livestock to produce efficiently and thrive in difficult environments.

Crossbreeding is a breeding strategy used to exploit the genetic variation that exists between breeds or lines. By implementing a breeding program, combining two or more breeds, breeders can increase production levels in two ways, heterosis (or hybrid vigor) and complementarity. Heterosis is the result of non-additive genetic effects and is defined as the percentage of superiority expressed in a trait by the cross progeny over the average of the parent breeds in the cross (Wakchaure *et al.*, 2015). Heterosis tends to be highest for traits with the lowest heritability. Since heritability estimates of adaptive traits, such as heat tolerance and resistance to both endo and ectoparasites, are characterized as low to moderate ranging between 0.05 – 0.4, crossbreeding could be used to improve these traits to achieve and benefit from heterosis (Burrow & Henshall, 2014; Ghebrewold, 2018). Heterosis can be parental (maternal or paternal) referring to the performance of the animals as parents and individual heterosis referring to the non-parental performance of the individual. Complementarity is the result of additive genetic effects when two or more traits complement each other, using two or more breeds it achieves a higher frequency of the desired genes between crosses than it would have been found within a single breed (Mishra *et al.*, 2017). To

design an effective breeding strategy, it is important to understand how heterosis between traits varies depending on the combination of breeds and environmental conditions, which is particularly important in tropical systems that encompass a wide variety of breeds and environment conditions (Bunning *et al.*, 2019).

Crossings in the tropics depend on many factors, especially environmental ones, as well as on the objectives pursued in terms of production. Crossing strategies can be classified into the following three broad categories:

1.2.1. Breed replacement strategies

Perhaps the most used method for a long time is the so-called grading up. This type of crossing refers to the breeding strategy in which a single sire breed (or cross) is continuously used as the parent type, like the sire through several generations (Table 1). Koufariotis *et al.* (2018) mention that the formation of the Brahman breed involved two different stages of "grading up", the first in the United States and the second in Australia. This predominantly involved the crossing of *B. taurus* females with Brahman bulls. This is genomically reflected in the introgression of large segments of *B. taurus* into the Brahman genome on the X chromosome.

Generation	Sires	Dams	Progeny	% A in progeny	% N in progeny
0	*Bos taurus* (A)	*Bos indicus* (N)	1/2A - 1/2N	50	50
1	A	1/2A - 1/2N	3/4A - 1/4N	75	25
2	**A**	3/4A - 1/4N	7/8A - 1/8N	87.5	12.5
3	**A**	7/8A - 1/8N	15/16A - 1/16N	93.75	6.25
4	**A**	15/16A - 1/16N	31/32A - 1/32N	96.9	3.1
5	**A**	31/32A - 1/32N	63/64A - 1/64N	98.4	1.6
6	**A**	63/64A - 1/64N	127/128A - 1/128N	99.2	0.8

Table 1.1. Concept of grading up. Repeatedly breeding successive crossbred generations back to the same breed (*B. taurus* breed A, in this example) over several generations produces animals that are essentially purebreds. Adapted from *Beef Cattle Production System* (p.47), by Herring (2014), Texas, USA: CABI. Copyright (2014) by Andy D. Herring.

1.2.2. Establishment of stable crossing systems

Two breed cross is the process where two pure breeds are mated together and their cross progeny are used in the breeding program (Figure 1.3). The two breed cross system produces a first cross progeny, or F1. In this system, the progeny resulting from the crossing of two breeds are usually sold for slaughter or to another commercial breeder. The system is most useful for situations where the females of a specific breed are well adapted to a particular environment (Mishra *et al.*, 2017). This type of cross is where the highest percentage of heterosis is exploited.

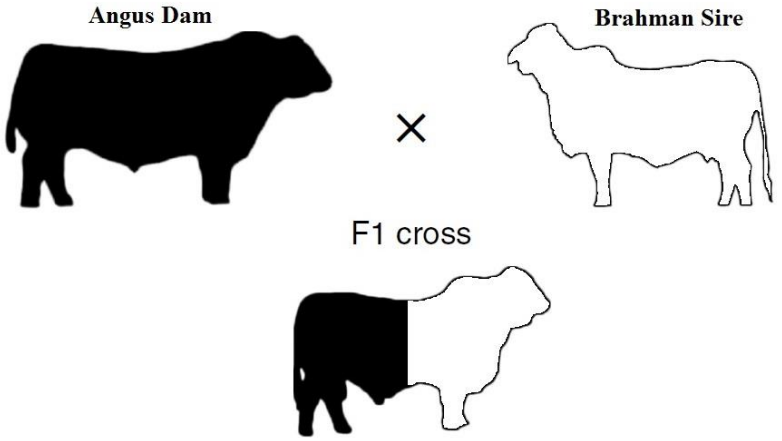

Figure 1.3. Typical cross between two pure breeds of different subspecies. Adapted from (Koren *et al.*, 2018).

1.2.3. Rotational crossing or criss-crossing

This crossing system follows the same principle as those mentioned above, that is, the initial generation begins with a *taurine* breed and another *indicine*, but the difference is that the resulting replacement heifers in each generation are mated with bulls of the opposite breed to their sire (Figure 1.4). The results of rotational crosses show that high levels of hybrid vigor can be maintained in advanced generations (Neufeld-Arce, 2006) and reap the important other half of the advantage of the cross, the complementarity of the differences of the breeds involved (Gosey, 2005).

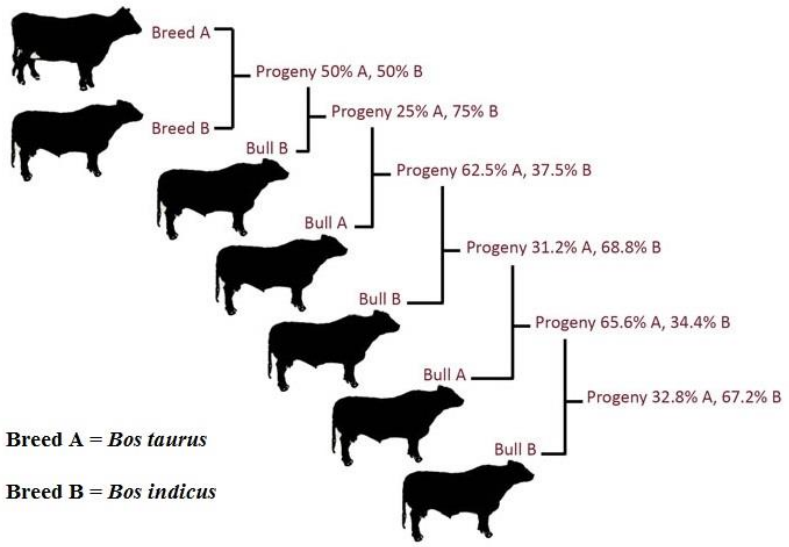

Figure 1.4. Rotational crossbreeding. Starting at 50/50%, the rotation stabilizes at 65/35% or 35/65%, giving 65% from the last sire line used. Adapted from Breeding for profit (p. 11), compiled by Bertram *et al.* (2002), Queensland, Australia: Copyright (2002) by Department of Primary Industries.

1.2.4. Formation of synthetic or composite populations

There are many examples of the formation of synthetic or composite breeds suited for the tropics, such as Brangus, Braford, Santa Gertrudis, among others. Their development through hybridization of the two main subspecies, *B. taurus* and *B. indicus*, was intended to combine environmental adaptability and optimal production yield (Paim *et al.*, 2020), as well as to produce heterosis without continuous crosses. Galukande *et al.* (2013), mention that the formation of these breeds can be in two ways. The first and simplest form involves two parental breeds being crossed to produce the F1 generation. Selected F1 individuals are crossed with each other to produce the F2 generation. This process is repeated in the following generations (Figure 1.5). The second program involves three breeds, for example, it could produce a synthetic with 25 percent zebu genes, 25 percent from some *taurine* breed and 50 percent *B. taurus* genes, from another breed.

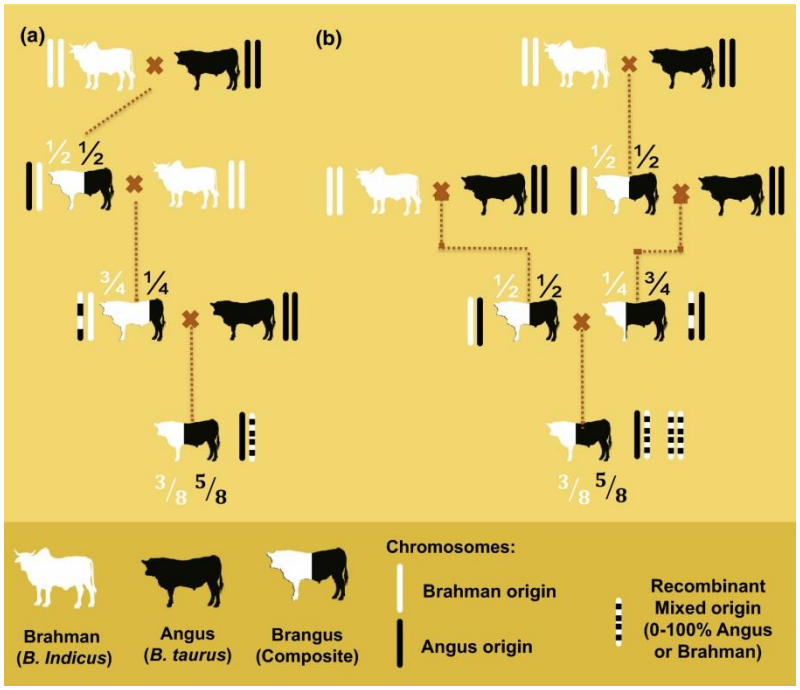

Figure 1.5. Scheme of formation of the Brangus breed, a synthetic breed composed of 5/8 Angus and 3/8 Brahman. Adapted from (Paim *et al.*, 2020).

1.3. Situation in Cuba

Cuba has breeding programs in main livestock species, with more or less development depending on the country's priorities, genetic material and interest contribution to food production. The programs are designed as a pyramid shape, with the upper stratum comprising the genetic nuclei, in the middle the multiplier herds and at their base the commercial herds. Pure breeds are mostly used in crossbreeding programs.

The beginning of genetic improvement in Cuban cattle rearing dates back to the 60´s. Through artificial insemination, different breeds specialized in milk production (Holstein, Jersey, and Brown Swiss), were mated with zebu cows in order to study the performance of crossed F1 females, under the same conditions. These crossing resulted in the current breeds: Siboney from Cuba (5/8 Holstein 3/8 Zebu) and Mambí from Cuba (3/4 Holstein 1/4 Zebu). Between 1981 and 1991, animal husbandry switched its interest from meat production (which accounted for 15% of the total cattle production) to milk production. Starting in the 1990s, the insemination started with more rudimentary genotypes like Creole, Milking Zebu (3/4 Zebu

1/4 Holstein) and zebu, until reaching the current genotypes, a herd under insemination, with the best participation of Siboney. About 50% of all the females from milking breeds included on the genetic herds are Siboney and Siboney crossbred, 15% are Mambí and Mambí crossbred and only about 5 % are Holstein (Hernández *et al.*, 2015).

According to the *'Informe de país sobre la situación nacional de los recursos zoogenéticos en animales de granja'* (Fernández, 2003), ten dairy and meat cattle breeds and buffaloes have been adequately characterized in Cuba. This characterization has mainly consisted in the calculation of statistical parameters of the most important traits from an economic point of view. In addition, estimates of genetic parameters and genetic values have been made by the best linear unbiased prediction (BLUP) Animal Model method, which has been limited in breeds with a small effective number of animals. This model performs an independent evaluation of each breed, disregarding the fact that populations are multibreed because there is more than one breed in the same herd. On the other hand, breeds obtained from crossing have common ancestors with the Holstein breed, so the animals are related in the pedigree. In addition, studies carried out by Acosta *et al.* (2013) reported that Mambí from Cuba and Siboney from Cuba breeds are genetically related. In this regard, the breeds that have been used the most specifically in dairy cattle have been Siboney, dairy zebu and Holstein, the first two for their hardiness and adaptability, and Holstein for its use on zebu for crossbreeding. As for meat production breeds, the most commonly used have been the zebu and the Creole. Neither breed is used as pure. Dairy breeds with well-defined breeding schemes for milk production are Holstein, Siboney and Mambí. For meat production: zebu, Charolais and Santa Gertrudis. The Creole breed has also been through improvement schemes, mainly for meat production.

1.4. Situation in Paraguay

Paraguay is responsible for almost 4% of world beef exports and ranks eighth among the top ten exporters in the world. Most of the beef breeds that have been commercially exploited in the country were imported. Imports were made for breeding purpose to improve and cross local breeds with selected breeds for their meat production. For example, in the beginning of the 1950s, the Inter-American Service for Agricultural Cooperation (STICA) introduced zebu breeders with a strong influence of the Nelore breed, shortly after artificial insemination

started at the Barrerito Experimental Station, and in 1953 the Artificial Insemination Centre was set up with specimens mainly selected for meat production. Since then, the introduction of genetic material is constant and it achived with the live specimens, frozen semen and frozen embryos (Ministerio de Agricultura y Ganadería, 2004; Asociación Rural del Paraguay, 2010).

It had been generally assumed that the local breed performance was poor and not worth to devote efforts in its improvement, if they were subject to selection. Yet, crossbreeding produced a spectacular result in the first generation, which further accentuated the desire to absorb it with the breeding breed. The import of pure animals from other breeds to cross them with the Creole or native breed was intensified, and became the strategy for the improvement of production and productivity, using the mechanism of absorbent crossing to replace one breed by another. As the crossing scheme was not accompanied by selection, after several generations the crossbred diluted its hybrid vigor and resulted in lack of adaptation to the climate, the indexes of production fell again, generating conflicts between the genetic variability and the maximization of response to selection in productive aspects.

Other commercially exploited breeds are in the same situation as creole, without plans to characterize the biotypes of the crosses. To date, there are no census records by breed, other than those of the Genealogical Registry of the Asociación Rural del Paraguay (A.R.P.). Because alternative crossing systems (criss-cross) between European species and zebu, the herd is maintained in terms of increased productivity. The assessment of the benefits of the specific crossing systems used (absorption, alternative, terminal or complementary) has become very difficult and they are not counted as a whole, although there are some individual data that reflect the estimated trends. In this segment of production, the introduction of genetic material is continuous in all breeds, in the form of live specimens or frozen semen and/or embryos. In the group that uses its own generated resources, commercial bulls and wombs, almost all are not subjected to genetic evaluation processes, which is detrimental for the maintenance of biotypes and therefore to biodiversity. The ease of access to imported genetic material, strategically used, explains the little interest in developing their own schemes of breeders produces as it requires that consume time and capital. Most of the commercial breeds that make up the genetic material of the country, with the exception of

creoles, have continuous introduction of foreign genetic material (Ministerio de Agricultura y Ganadería, 2004; Köbrich *et al.*, 2018; Nin-Pratt *et al.*, 2019).

1.5. Genomic selection

Most economic traits are polygenic (i.e. influenced by many genes) and tracking a few through DNA markers will only explain a small proportion of the genetic variance. In addition, there are individual genes with small effects, so large amounts of data are needed to accurately estimate their effects (Goddard & Hayes, 2007). Before the massive use of genetic markers, BLUP tool was used to directly predict the genetic values of the animals. Overall effects were treated as random effects with a variance structure defined by parentage based on pedigree data (Henderson, 1984). In the 1990s, several efforts were devoted to identify polymorphic chromosome segments (QTLs) that impacted on quantitative traits. To identify candidate genes affecting traits of interest, the QTLs mapped were linked to genetic markers chosen to represent the dispersed coverage of the entire genome. It was also a prerequisite for the implementation of marker assisted selection (MAS).

Weller *et al.* (1990) proposed special designs to increase the power of QTL detection in dairy cows, in particular the daughter-granddaughter design. This scheme traces the transmission of genetic markers in families comprising a sire and his proven offspring, which already have precise phenotypes (adjusted mean yield of a large number of daughters). Due to their highly polymorphic capacity, microsatellites were initially used as markers, but their genotyping raised costs and QTL detection studies were limited to the use of only a few hundred microsatellites per animal. Thanks to the availability of high-density marker coverage, QTL detections have improved considerably (Weller *et al.*, 2013).

Single nucleotide polymorphism (SNP) polymorphisms helped to provide sufficient genomic information to be used for EBV. Meuwissen *et al.* (2001), inspired by the MAS selection process, proposed to integrate all available genetic markers simultaneously as random effects and evade the over-adjustment of the MAS model, which they called genomic selection. In this way, the effects of all the markers were added together to estimate the total genetic value of an individual, assuming a normal distribution with variations after a certain previous distribution. The basic principle of genomic selection is that SNPs are in linkage

disequilibrium (LD) with QTLs in the genome (Mrode *et al.*, 2019). The use of markers such as SNPs allows the indirect identification of all QTLs in the genome by mapping the chromosome segments defined by adjacent SNPs.

Genomic selection involves estimating the effects of SNPs in a reference population, consisting of animals with phenotypic and genotypic records. The estimated genomic breeding value (GEBV) of candidates for selection without phenotypes is then predicted. Different approaches have been proposed for the efficient application of genomic selection. One of the most used algorithms is marker-based kinship (VanRaden, 2008). This approach, known as genomic BLUP or genomic best linear unbiased prediction (GBLUP), has become the practical method for genomic selection (Hayes *et al.*, 2009). In addition, a single-step GBLUP approach (ssGBLUP) was developed for situations where the simultaneous use of marker-based kinship and pedigree-based kinship is advantageous (Legarra *et al.*, 2009; Aguilar *et al.*, 2010; Christensen & Lund, 2010). Genomic selection can be conceptualized as a two-step process. The first step consists on estimating the effects of markers in a training population containing animals with phenotypic (measurement of economically important traits) and genotypic (information on the SNP of each animal) data. The second step is to estimate their genomic breeding values (GBV) by adding up the marker effects of the estimated SNP from the training population.

Hayes *et al.* (2013) mention that one of the main advantages of genomic selection over traditional selection (based only on pedigree and phenotype), is the accurate selection of animals in their early life stages, based on their genomic predictions. This becomes very useful for traits that are difficult or expensive to measure, such as fertility, disease resistance and feed conversion, to cite some outstanding examples. Several studies have documented the benefits of using this selection system. In dairy farming, genomic selection has almost replaced traditional selection based on progeny testing (de Koning, 2016). In beef cattle, increasing the accuracy of genetic predictions for traits of interest remains a major challenge (Kuehn *et al.*, 2011). Compared to dairy cattle, genomic selection in beef cattle is at an early stage. Beef production is expected to benefit from this approach identifing genetically superior animals earlier and more accurately, and select for traits of interest such as meat quality and feed efficiency that are difficult and costly to measure (Albuquerque *et al.*, 2017).

Several challenges still need to be overcomed for their application to be considered successful. One is the necessity of a founder population that relates genomic profile to phenotypic performance. Another challenge is the breed specificity, which will require more dense marker panels. Nevertheless, genomic selection increases genetic progress, brings greater intensity and precision in selection, reduces the generation interval and allows to genetically select for the trait of interest (Miller, 2010).

1.6. Simulation software

In recent years, there has been an increase in the use of simulation tools for the analysis of complex systems. This software allows complex dynamic events to be modelled, which are not possible using algebraic methods due to its versatility and the possibility of experimenting with different options. In addition to the ease of interpretation of results without previous knowledge, simulation has become important when evaluating alternative breeding schemes, allowing a wide range of hypotheses to be explored at no cost (Zingaretti *et al.*, 2019). This is particularly effective in the case of genomic selection, as it remains an expensive technology that requires adaptation to the specific breeding scenario.

Simulation studies have made important contributions to the advancement of animal and plant breeding (Daetwyler *et al.*, 2013; Liu *et al.*, 2019). Since many breeding programs now incorporate genomic information that is costly to obtain, simulation is both useful and necessary to compare, at low cost, the potential of different methods of analysis to increase the accuracy of the EBVs and to compare the alternative structures of breeding programs. In addition, simulation can be used to test and compare software packages. Recently, many alternative simulation strategies have been applied in livestock. These strategies use different forms of data simulation, have distributions of the effects of QTL, and present different relationship structures. This complicates the comparison of results and conclusions drawn from different studies (Hickey & Gorjanc, 2012).

Both real and simulated genomic data have been used in genomic prediction studies to investigate aspects such as the power of different methods of analysis, to compare alternative genomic breeding programs and to explore the dynamics of genomic selection. While actual data offer the advantage of reflecting real life complexity, the cost may limit the available

data to a single, possibly non-random, sample from one population with a small sample size. On the other hand, simulated data allow the researcher to explore important aspects, such as the genetic architecture of the trait, the number of markers for analysis, the degree of relationship between the formation and prediction populations, thus allowing some sources of variability to be assessed, such as drift, which cannot be assessed with most real data. The usefulness of simulation lies in the rapid re-testing of a wide range of hypotheses at low cost, such as the initial feasibility of genomic selection or the impact of the reference population size (Daetwyler *et al.*, 2013).

Genetic data simulation can be grouped into three categories of algorithms: backward-time, forward-time, and resampling approaches. The first, also known as coalescent simulation, simulates the ancestral sequences of individuals up to the current generation and the most common recent ancestor, introducing mutations into the generated genealogy. It has been widely used in population genetics due to its speed but presents difficulties concerning selection, the most relevant parameter in animal breeding. One of the most recent softwares of this category is the one developed by Kelleher *et al.* (2016), called msprime. The efficiency of this program, according to its own authors, lies in the ability to simulate much larger sample sizes, in addition to simulating realistic scenarios including recombination models on whole chromosomes without resorting to approximations. As for the data structure, which is known as a succinct tree sequence that reduces the amount of space needed to store the simulations and eliminates the significant overhead of loading and analyzing large volumes of text to analyze the simulated data, msprime uses it to represent the simulation results in a concise manner so that they can be processed efficiently.

The forward-time simulation methods are designed to start from an initial population and follow its evolution under various genetic models, over several generations, with samples usually taken from the last or one of the last generations. The final population is reached when the stop criterion is met (e.g. a certain number of generations). In this way all the ancestral information is obtained and the properties of the population are observed in any generation (Yuan *et al.*, 2012). In practice, in animal breeding the most popular approach is the forward simulation. simuPOP (Peng & Kimmel, 2005) and QMSim (Sargolzaei & Schenkel, 2009), are two well-known software applications based on this simulation approach. The first allows

the generation of large multi-generational populations with complex diseases according to the specified demographic and evolutionary properties and allele frequencies of the current generation. Arbitrary study designs, determination methods and genetic mapping methods can be applied to the simulated populations. The second was designed to simulate large genomes and complex pedigree structures, imitating cattle populations. It is basically a family-based simulator, which can also take into account predefined evolutionary characteristics, such as ligation imbalance, mutation, bottlenecks and expansions. Recently, following the gene drop approach, Zingaretti *et al.* (2019), developed a versatile simulation tool called Polyploid Sequence Based Virtual Breeding (pSBVB), an extension of the Sequence-Based Virtual Breeding software (SBVB, Pérez-Enciso *et al.*, 2017), which allows the evaluation of genomic selection strategies in polyploid species, such as variable SNP density, genetic architecture or population size, as well as the optimization of experimental designs for association studies.

Resampling is another simulation strategy that is usually based on existing data sets. Unlike the previous ones, which need to manage complex scenarios of evolution and demography, resampling does not require this type of modelling. It uses random sampling with replacement (e.g. bootstrapping) to generate simulated data from a given dataset. The sample size is theoretically unlimited, while the length of the simulated genome is usually limited to that of the existing genome. Here, the simulated data usually follows the allele frequency and linkage disequilibrium patterns in the observed chromosome population. However, there will still be some deviations between simulated and actual data in respect to allele frequency and linkage imbalance, which depends largely on the size of the simulated population (Yuan *et al.*, 2012). An example in this category is the software developed by Wright *et al.* (2007), called HAP-SAMPLE, which is particularly useful studying haplotype-phenotype association methods, whose power depends largely on the duration and specificity of the associated haplotypes and the frequencies of risk alleles. In other words, it simulates SNP data with realistic patterns of linkage disequilibrium and allele frequencies for disease association studies by resampling haplotypes of chromosome lengths.

Chapter 2

Objectives

Objectives

The main objective of the thesis is to evaluate the effects of adaptation and selection in beef cattle in tropical climates, whether adapted, introduced or improved, as well as their potential genetic gains through different mechanisms of genetic improvement.

This general objective has been developed into the following specific objectives:

1. To estimate the genetic diversity, ancestral origin and putative *Bos indicus* hybridization in a Charolais Cuban population adapted to the tropical climate.
2. Identify selective sweeps related to adaptation in Cuban Charolais.
3. To develop a forward simulation tool to explore genomic selection in generic scenarios.
4. Evaluate genetic progress in various selection schemes with *Bos taurus* x *Bos indicus* crosses and propose several criteria to improve genomic evaluation.
5. To explore opportunities for improving production and adaptation traits in the most widely used beef breeds in tropical climates through genomic selection.

Chapter 3

Papers and studies

Paper I

WHOLE-GENOME SEQUENCING REVEALS INSIGHTS INTO ADAPTATION OF FRENCH CHAROLAIS CATTLE TO CUBAN TROPICAL CONDITIONS

Ramírez-Ayala, L.C.[1], Rocha, D.[2], Ramos-Onsins, S.E.[1], Leno-Colorado, J.[1], Charles, M.[2,3], Bouchez, O.[4], Rodríguez-Valera, Y.[5], Pérez-Enciso, M.[1,6] Ramayo-Caldas, Y.[2,7]*

[1]Plant and Animal Genomics, Centre de Recerca en Agrigenòmica (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, Spain.

[2] Université Paris-Saclay, INRAE, AgroParisTech, GABI, F-78350, Jouy-en-Josas, France.

[3]INRAE, SIGENAE, F-78350, Jouy-en-Josas, France.

[4]INRAE, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France.

[5] Facultad de Ciencias Agropecuarias, Universidad de Granma, Granma, Cuba.

[6] Institut Català de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

[7] Animal Breeding and Genetics Program, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, Caldes de Montbui, 08140, Spain.

*Corresponding author

Genetics Selection Evolution (submitted)

Abstract

Background

Early in the 20th century, Cuban farmers imported Charolais cattle directly from France. These animals (CHCU population) are now adapted to rough environmental tropical conditions that include long periods of drought and food shortage, all accompanied by extreme temperatures that European taurine cattle can hardly endure.

Results

In this study we used whole-genome sequence from a total of 12 CHCU samples and 60 samples from six other *taurine*, *B. indicus* and crossed breeds to estimate the genetic diversity, structure and precise ancestral origin of the CHCU animals. Despite these animals were supposed to be a closed population, admixture analysis indicates a limited *B. indicus* introgression, likely due to a single and not continuous introgression event. The Extended Haplotype Homozygosity (EHH) test was used to identify regions that may have had an important role in the adaptation to tropical conditions. Regions with high percentage of *B. indicus* were enriched in putative selective events, but only mildly and adaptation cannot be explained by the *B. indicus* influence alone. For example, EHH reveals signals of potential adaptation in genome windows that include genes involved in thermogenesis (*UCP1, DIO2 and ACSL1*) and hair development (*BMPR1A, CDSN and FGF7*). We also identify variants within these genes without *B. indicus* influence that may have a functional impact and that could thus explain some of the phenotypic differences observed between CHCU and French animals.

Conclusions

In resume, our results based on whole-genome data confirm that CHCU samples were closely related to Charolais from France (CHFR) and Canada, but also revealed a limited *B. indicus* introgression into CHCU. We observed signals of recent adaptation to tropical conditions between CHCU and CHFR founder populations, which were largely independent of the *B. indicus* introgression. We reported candidate genes as well as variants that may have a functional impact and could explain some of the phenotypic differences seen between CHCU

and CHFR animals. However, experimental work is needed to further infer the involvement of these interesting genes in CHCU adaptation.

Introduction

Climate change and global warming is one of the main challenges being faced by Agriculture and Livestock husbandry. In this scenario, it is fundamental to assess which are the mechanisms whereby animals adapt to high temperatures. Hot temperature conditions of tropical climates today may resemble those that could be faced in the future in temperate climates by animals raised outdoors, such as most beef cattle and small ruminants. Therefore, animals currently living in hot climates but with an European or temperate climate origin can provide clues into the genetic mechanisms of adaptation to raising temperatures (Naves *et al.*, 2015).

Cattle breeds can be divided between temperate *taurine* breeds *B. taurus*, of European origin, and Indian zebu breeds *B. indicus*. Both subspecies diverged ~ 250,000 years (Upadhyay *et al.*, 2019). *B. indicus* produces less meat and of lower quality but is much better adapted to heat and parasites than *taurines*. For that reason, they were imported to tropical American regions starting mid-19th century (Ajmone-Marsan *et al.*, 2010; Koufariotis *et al.*, 2018). In those climates, they have largely replaced the primigenious cattle imported by the first Spanish and Portuguese settlers (Ginja *et al.*, 2019). Numerous hybrid populations between *B. taurus* and *B. indicus* also coexist with pure zebu, such as Brangus, Texas Longhorn, Santa Gertrudis among others. These mixed breeds exhibit good resistance to parasites and heat and also produce carcasses of much higher quality than pure zebu.

At the beginning of the 20th century, Cuban farmers imported Charolais animals from France, a breed now known as 'Chacuba' (CHCU). This population has adapted to breeding conditions in Cuban tropical environment in ~20 generations of breeding. In such a short period of time, evident phenotypic differences between original French Charolais (CHFR) and its Cuban counterpart have appeared. CHCU animals are smaller than French Charolais, with weights of 34 vs 46 kg at birth and 290 vs 493 kg for heifer's weight at 18 months (Renand *et al.*, 1997; Rodriguez-Valera *et al.*, 2018). Also, CHCU animals are hairless and produce carcasses with lower grade and higher fat content than CHFR (Renand *et al.*, 1997). Although

CHCU is thought to be a closed population with no records of interbreeding, Ribas (1981) reported the presence of a specific blood group zebu allele (U'_1), albeit at a very low frequency. More recently, Rodriguez-Valera *et al.* (2018) employed the Illumina's bovine 50k BeadChip to report the genetic structure and putative ancestral origin of this population. These authors report that CHCU clusters with *taurine* breeds. Nevertheless, despite the short period of time spanned since importation, a noticeable differentiation (*Fst*=0.049) between CHFR and the CHCU was also observed.

Since array genotype data are biased and offer low resolution, here we obtained whole-genome sequence data from 72 animals, including 12 CHCU animals, to provide an unbiased estimation of the population structure and fine map regions that may have played a role in the adaptation to tropical conditions of the Chacuba population.

Methods

Samples

A total of 72 whole-genome sequences from *taurine*, *indicine* and crossbred cattle were analyzed (Boussaha *et al.*, 2015). Since the importation, CHCU animals have been maintained under pedigree control at "Manuel Fajardo" genetic center located in Jiguani (Granma Province). Therefore, the relationship between samples can be accurately tracked and we used this information to select the 12 unrelated CHCU samples that were sequenced for this work. We also used 15 French Charolais (CHFR) (Letaief *et al.*, 2017), 6 Limousin from France (LIMS) (Guillocheau *et al.*, 2019) and sequences from 39 additional individuals that were downloaded from SRA (Table S1): Canadian Charolais (CHCA, n= 15), 5 Canadian Limousin, Brangus (BRG, n=5), Brahman (BRM, n=10) and Texas Longhorn (TXL, n=4) breeds. Brahman is a pure *B. indicus* breed whereas Brangus and Texas Longhorn are admixed breeds between *B. indicus* and *B. taurus*.

Bioinformatic analysis

The whole dataset was mapped against the Bovine reference assembly (UMD3.1.1) using BWA v. 0.7.12-r1039 (Li & Durbin, 2009). PCR duplicates were removed using Picard MarkDuplicates (v2.18.9) and realigned around indels with the GATK IndelRealigner tool (Mckenna *et al.*, 2010). For each individual, the SNP calling was done with SAMtools

mpileup and bcftools call (v. 0.1.19-96b5f2294a) using the following parameters: minimum and maximum depths between 5x and twice the average sample's depth; the minimum SNP quality was 10 in each sample; minimum mapping quality and minimum base quality of 20. Afterwards, we merged individual gVCF files into a multi-individual VCF file, with all the SNPs from the 72 samples. For this purpose, we followed a two-step approach as detailed elsewhere (Pérez-Enciso *et al.*, 2017), available at https://github.com/miguelperezenciso/NGSpipeline. In brief, to identify whether a position is equal to the reference, polymorphic or missing, we first generated a fasta file from the gVCF file for each individual and generated a multi-individual VCF file using the individual file. Once the multiple sample file was obtained, we discarded the SNPs with > 20% of missing data of the samples in all populations. Finally, we imputed the missing genotypes and inferred phases with Beagle 4.1 (Browning & Browning, 2016).

The identification of genetic variants altering transcription factor binding sites (TFBSs) were predicted with a custom script using the transcription factor binding site models from the JASPAR (JASPAR CORE 2018 collection, Sandeli *et al.*, 2004), HOCOMOCO (version v10, Khamis *et al.*, 2018) and TRANSFAC (version v3.2 public, Knüppel *et al.*, 1994) databases. These databases contain curated set of transcription factor binding models represented as Position Weight Matrices (PWMs), derived from published collections of experimentally defined eukaryotes TFBSs. Only vertebrate PWMs were downloaded and used in our study. Finally, the identification of microRNA binding sites was done using TargetScan (release 7.2) (Agarwal *et al.*, 2015)

Population genomics

We estimated Watterson's (1975) nucleotide variability (θ) and differentiation values (F_{ST}, Hudson *et al.*, 1992) between populations with mstatspop (v. 0.1beta, https://github.com/CRAGENOMICA/mstatspop) in consecutive non-overlapping windows of 30 kb size. This software implements algorithms that allow for missing data (Ferretti *et al.*, 2012). Extended Haplotype Homozygosity (EHH)-derived statistics (*Rsb*, Tang *et al.*, 2007) was computed between CHCU and CHFR for each SNP (https://github.com/CRAGENOMICA/Tang_Rsb). As putative selective events, we retained the windows with 'permutation p-value' < 0.05 among the 2000 windows containing the

largest Rsb SNPs. The 'permutation p-value' was obtained by randomly shuffling the CHCU and CHFR samples and running the Rsb algorithm for the whole genome. The process was repeated 100 times and we computed the number of times the observed Rsb statistics was larger than the values obtained from permutation. In so doing we obtained a 'permutation p-value' for each SNP. This process aims at correcting for different levels of linkage disequilibrium along the genome that may locally inflate Rsb values.

Admixture

The software ADMIXTURE v1.3.0 (Alexander *et al.*, 2009) was run in an unsupervised manner with a number of clusters k=2 using CHCU, BRM, CHFR and CHCA genotypes. The program was run with either all SNPs or pruned data to remove SNPs in strong disequilibrium, but results were undistinguishable. To get a more precise map of potential admixture, we employed ELAI software (Guan, 2014). This is a partially supervised algorithm that requires data from the putative founder populations (CHFR and BRM) and the potentially admixed population (CHCU). Therefore, ELAI was run using only CHCU, CHFR and BRM genotypes. ELAI implements a two-layer hidden Markov model and was run using the recommended default parameters, which included removal of SNPs with minimum allele frequency (MAF) < 0.01.

Results

Population structure and the impact of *indicine* introgression on CHCU

Average depth across breeds varied between 8.7 (LIMS) and 12.5 (CHCU, Table 1). We found 42,144,809 SNPs; of those, 14,929,949 were exclusive of the *indicine* breed (BRM); 6,839,436 were found only in *taurines* and 1,176,249 were exclusive of CHCU. As expected (The Bovine HapMap Consortium, 2009), *B. indicus* samples were more variable than *B. taurus*, BRM nucleotide variability was 0.0035 per base pair while the observed nucleotide variability is halved in LIMS and European Charolais (Table 1). The genome diversity of hybrid breeds TXL and BRG was intermediate between *B. indicus* and *B. taurus*, and that of CHCU was similar to that in TXL (Table 1). Interestingly then, despite the small number of founders of CHCU, its variability was larger than in its ancestral population CHFR.

Breed	Country	No. of Samples	Variability Wattersons	Fst vs. CHCU	% Missing	Mean Depth
CHCU	CU	12	0.0021	-	0.03	12.5
CHFR	FR	15	0.0015	0.05	0.06	11.4
CHCA	CA	15	0.0018	0.04	0.13	10.4
LIMS	FR / CA	11	0.0016	0.06	0.21	8.7
TXL	US	4	0.0023	0.07	0.08	10.3
BRG	US	5	0.0028	0.08	0.05	11.5
BRM	US / AU	10	0.0035	0.31	0.11	10.8

Table 1. Description of the breeds analyzed

In line with previous observations (Lin *et al.*, 2010; Kasarapu *et al.*, 2017), the PCA plot shows a clear separation between *indicine* and *taurine* breeds, the latter breeds being tightly clustered (Figure 1A). As expected, samples from admixed breeds Brangus and Texas Longhorn, were positioned towards the *indicine* cluster but much closer to the *taurine* than to the *indicine* clusters, because the percentage of *indicine* is less than 50% (11% in the case of TXL, (McTavish *et al.*, 2013)). As for CHCU, it was positioned nearby Limousin and Canadian Charolais, but separate from original French Charolais. Note a CHCU individual seems to be an outlier so we inspected its genotype heterozygosity patterns (11%), but we could not find any anomalous deviation. That individual did not look as outlier when only BRM, CHCU and CHFR are represented in a separate PCA plot (Figure 1B). In terms of Fst, the closest populations to CHCU were the Canadian (0.04) and French Charolais (0.05) (Table 1). The plot of Fst values across 1Mb windows with CHCU exhibited a modal value near zero with *taurine* breeds, whereas that vs. the *indicine* population was clearly distinct (Figure 1C).

In a previous work based on the Illumina 50k Chip genotype data, we reported a putative introgression of *B. indicus* into CHCU, although of limited extent (Rodriguez-Valera *et al.*, 2018). This could explain why CHCU samples clusters distinctly from CHFR, despite the short period since their original importation from France. Here we confirm this introgression with better resolution using whole-genome data. As shown in Figure 1D, an unsupervised ADMIXTURE analysis with k=2 clusters clearly separate *taurine* CHCA and CHFR breeds from *indicine* Brahman, while a small introgression of *B. indicus* into CHCU is revealed. This is reflected also in the PCA showing only BRM, CHFR and CHCU breeds (Figure 1B).

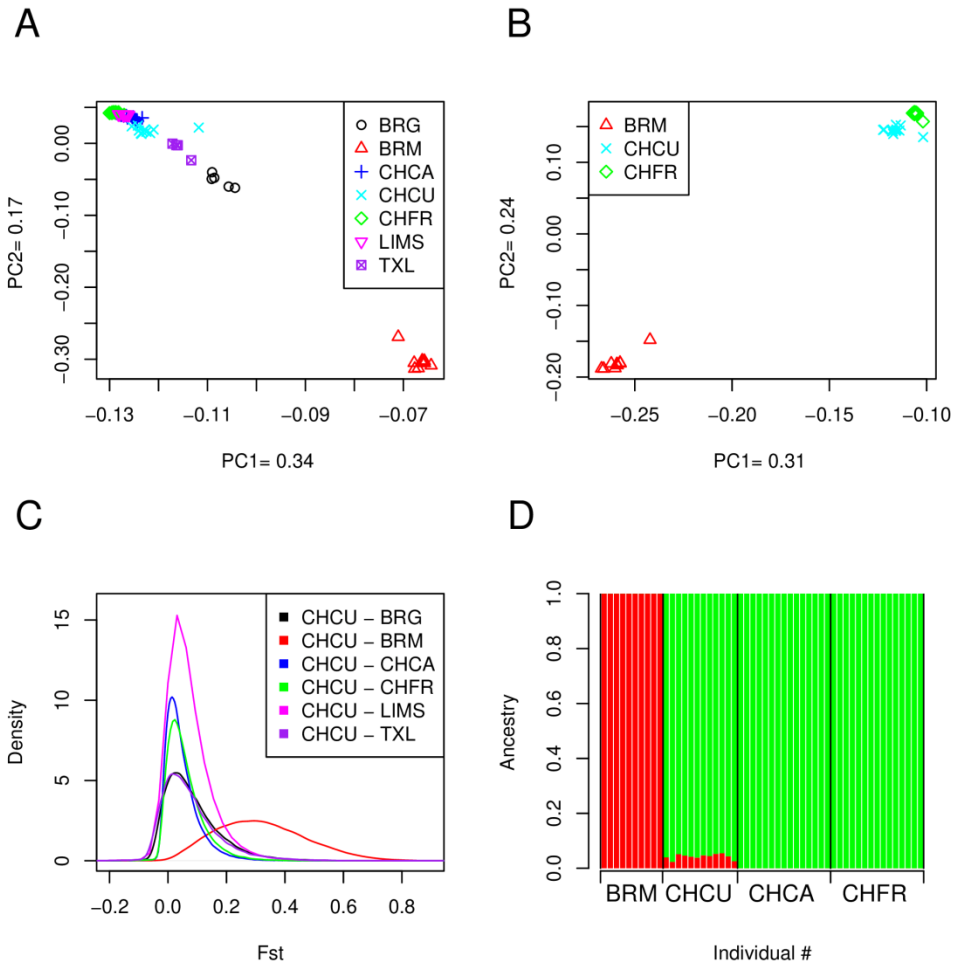

Figure 1: Population structure**. (A)** Principal component analysis using all samples. Individuals are grouped in *B. taurus, B. indicus* and Hybrid. Black: Brangus, red: Brahman, blue: Canadian Charolais, cyan: Cuban Charolais, green: French Charolais, magenta: Limousin, purple: Texas Longhorn; **(B)** Principal component analysis using all samples. Red: Brahman, cyan: Cuban Charolais, green: French Charolais; **(C)** Fst between CHCU and others breeds; **(D)** Results of Admixture analyses with 2 ancestral populations (k=2) the color correspond to: Red: Brahman (*B. indicus*) and green: French and Canadian Charolais (*B. taurus*).

To evaluate in detail the extent and impact of *B. indicus* introgression into CHCU, we run ELAI software (Guan, 2014). ELAI provides a map, for each hybrid individual, of the chance at each SNP to descend from one of the two putative founder populations. ELAI's results show that *B. indicus* introgression was not homogeneous, neither across individuals nor across chromosomes, suggesting a recent and limited introgression (Figure 2, and Figure S1). We

can expect an initial imbalance in *B. indicus* composition across CHCU animals if only a few founder CHCU animals were crossed with *B. indicus*. This imbalance is to be smoothed out in future generations as recombination events increase. For the same reason, it is expected that correlation in *indicine* percentage between individuals across windows will increase with time, assuming *indicine* introgression was a single and not continuous event. Figure S2 shows the relation between BRM content and Fst between CHCU and CHFR across windows. A larger *indicine* component was associated with a higher differentiation between Charolais, but only mildly (correlation = 0.19). This suggests that divergence between founder (CHFR) and derived (CHCU) populations is not likely due to *indicine* introgression alone.

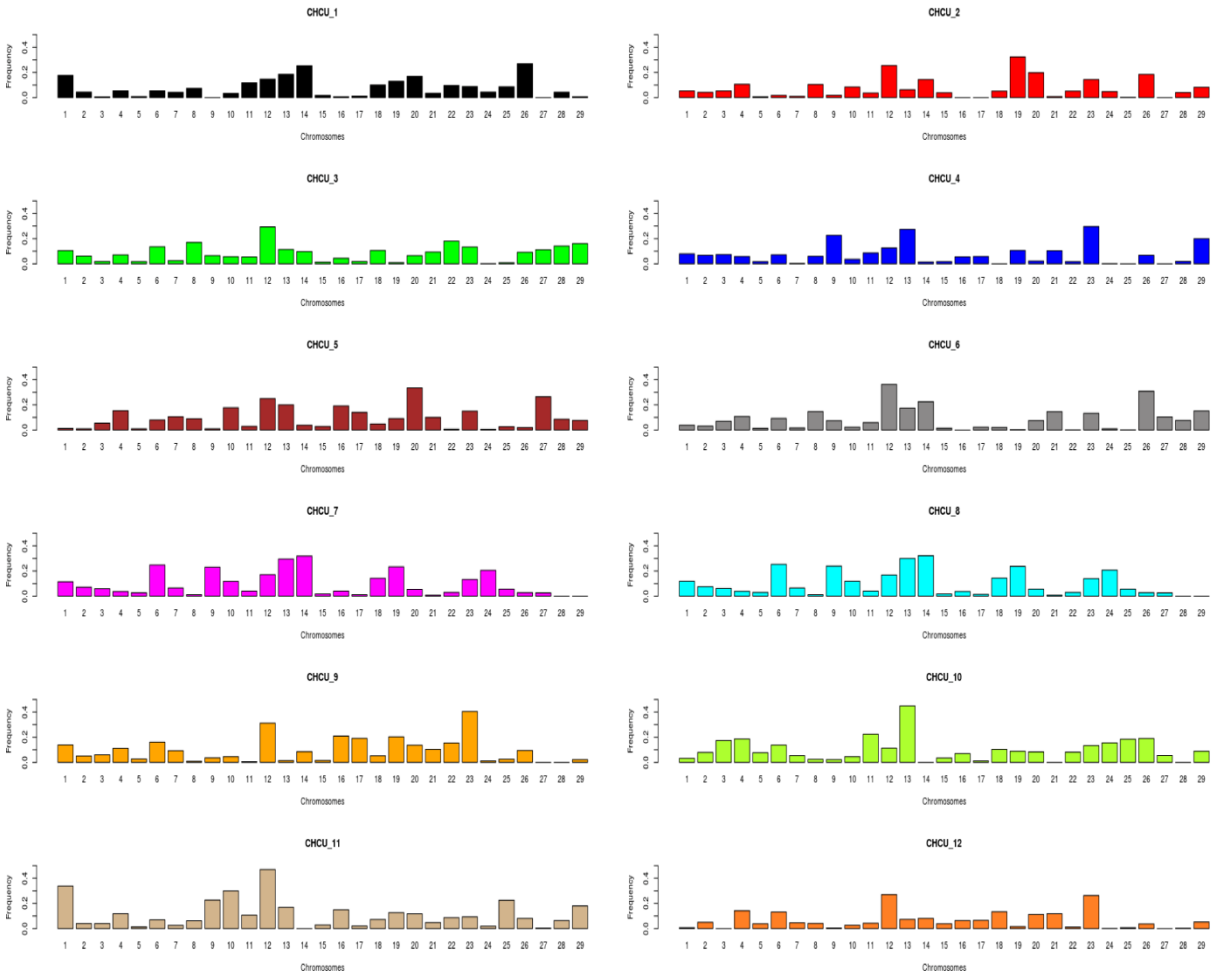

Figure 2. Frequency of *B. indicus* component per CHCU sample and chromosome inferred from ELAI software

Putative Selective sweeps between Cuban and French Charolais

We considered as candidate chromosomal intervals to be further inspected the 2,000 regions with largest Rsb and a 'permutation p-value' < 0.05, with the result of 1,817 of 30 kb windows selected (Table S2). We found 68 regions out of 104 reported by Rodriguez-Valera *et al.* (2018) also identified here. Note that the detection of putative selection sweeps in that paper was based on SNP Chip data and windows were much larger (~700 kb vs. 30 kb here). Moreover, in Rodriguez-Valera *et al*. (2018) we annotated both positive and negative extreme Rsb values without consider a second filter based on the permutated p-value. Here instead we only focused in positive values, since we want to detect regions where the putative selective pressure is specific to the Cuban population, i.e., where CHCU disequilibrium is larger than in CHFR. It is tempting to hypothesize that the detected *B. indicus* introgression is related to adaptation in CHCU. If this were the case, genome regions of high *B. indicus* percentage should be enriched in selective signals. Figure 3 shows the distribution of *B. indicus* percentage for putative selective windows and the rest of the genome. The average percentages of Brahman within selected and control windows were 0.22% and 0.17%, respectively. The difference, although small, was significant (P < 5e-09) according to a Wilcoxon rank test.

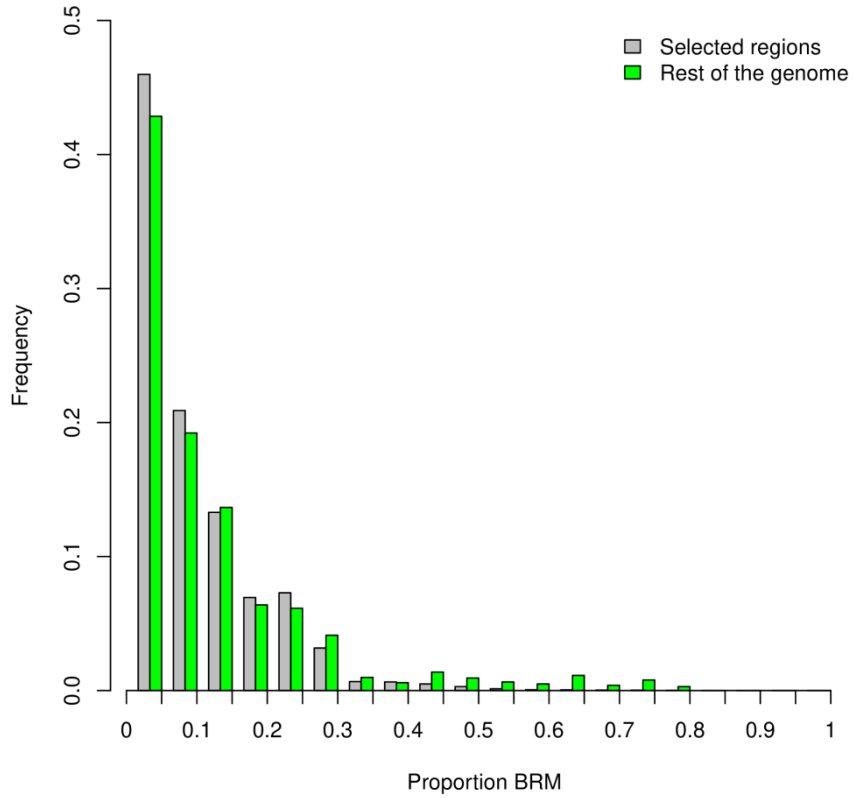

Figure 3: Percentage of BRM in windows containing putative selective events vs. rest of the windows, inferred from ELAI software.

We identified 887 genes within the selected windows. Among those genes, we found three genes (*UCP1*, *DIO2* and *ACSL1*) involved in thermogenesis as well as three other candidate genes involved in hair development (*BMPR1A*, *CDSN* and *FGF7*). We also identified variants within these genes that may have a functional impact and that could hence explain some of the phenotypic differences seen between CHCU and CHFR animals. For example, we found two deleterious missense variations within *UCP1* (rs443726914 and rs715309385) and 77 upstream SNPs, including four fixed (rs438305189, rs211174809, rs209939359 and rs211622720) for the alternative alleles in CHCU. Interestingly, these four upstream SNPs potentially alter the binding sites of several transcription factors (see Table S3). There were 986 SNPs within *BMPR1A*, a candidate gene for hair development, 14 SNPs were located within the 3'UTR and 104 variants in upstream regions, including three (rs381889584, rs384649723, rs210640548) fixed for the alternative allele in CHCU. Another candidate gene

of hair development was the corneodesmosin gene (*CDSN*). From 466 SNPs within our bovine samples, we identified one SNP potentially impacting splicing of *CDSN* (rs462034580), 194 located in upstream and 3 missense deleterious SNPs (rs209222317, rs434552200 and rs479537418). Interestingly, one of the *CDSN* deleterious mutations (rs434552200) is nearly fixed for the alternative allele in CHCU, with a frequency of 0.994. Further experimental analysis of the impact of these variants is needed to see if may explains some of the phenotypic differences found between CHCU and CHFR.

Discussion

Creole cattle refer to those descendants of first European animals that have adapted to local conditions in the American continent, typically living in tropical environments. However, as (Burgos-Paz *et al.*, 2013) showed for 'creole' pigs, the origin of these animals is usually mixed and little trace of original founders can usually be recovered. There are very few instances where the origin of extant animals can be tracked accurately and, for that reason, CHCU is a unique population where pedigree records have been maintained for most of the time and where isolation is, in principle, guaranteed.

Despite this, here we confirm previous results based on SNP Chip data (Rodríguez-Valera *et al.*, 2018) which suggested that CHCU was crossed with *B. indicus* animals. Compared with the previous report (Rodríguez-Valera *et al.*, 2018), our results on whole-genome data are not slanted by the SNP ascertainment bias, allowing us to report a most accurate trace of the impact of *indicine* introgression on CHCU. Futhermore, another advantage is the possibility to estimate and contrast the nucleotide diversity between breeds, but also to fine mapping the putative selective sweeps between Cuban and French Charolais. Our results suggest that *B. indicus* introgression was more likely a recent single event of limited extent. We estimate that the *B. indicus* content in CHCU is in the order of 4-8%, depending on the method used and on genome region or individual considered (Figures 1D, 2, Figure S2). Overall, this *B. indicus* component had a measurable impact in increasing CHCU nucleotide variability compared to the European breed, more than offsetting the effect of the founder bottleneck. Interestingly, the *B. indicus* component is associated with adaptation (Figure 3) but by no means can *B. indicus* alleles be considered the main drivers of selection. This leads us to conclude that,

likely, most adaptation events that have occurred in Chacuba are mainly due to changes in allele frequencies that were already present in French Charolais, i.e., soft sweeps.

We observed that 68 regions out of 104 reported by Rodríguez-Valera *et al.* (2018) were confirmed by our results based on whole-genome data. Oberved differences in the number de genomic intervals between studies can be explained by the fact that in Rodríguez-Valera *et al.* (2018) we annotated both positive and negative extreme Rsb values without considering a second filter based on the permutated p-value. Here, instead, we only focused in positive and significant (permutated p-value < 0.05) intervals, since we want to detect regions where the putative selective pressure is specific to the Cuban population. Moreover, please note that the detection of putative selection sweeps in Rodríguez-Valera *et al.* (2018) was based on SNP chip data, and the size of the genomic intervals were much larger (~700 kb vs. 30 kb here). Therefore, our results provided a better resolution and accurate identification of putative selective sweeps between CHCU and CHFR, wich in turn facilitates the identification of genetic variants within candidate genes related to the adaptation to tropical conditions.

The precise dissection of all signals of adaptation is beyond the scope of this paper, and here we focus on genes and phenotypes that look of particular interest, that is, thermotolerance and hair development. Both traits seem pivotal to the adaptation to tropical conditions and may also explain some phenotypic differences between the CHFR and CHCU breeds. As for heat tolerance, we detected *UCP1*, *DIO2* and *ACSL1* as genes putatively under selection. The mitochondrial uncoupling protein 1 (*UCP1*, BTA17: 17,467,450-17,473,822) is predominantly expressed in brown adipose tissue (BAT) and plays important roles in regulating body temperature, metabolic rate and controlling energy expenditure via both non-shivering thermogenesis and diet-induced thermogenesis (Chouchani *et al.*, 2019). The *UCP1* gene generates heat by uncoupling the electron transport during ATP synthesis in the inner membrane of BAT mitochondria (Nicholls *et al.*, 1978). *UCP1*-dependent non-shivering thermogenesis is necessary for thermoregulation as demonstrated by the cold intolerance of *Ucp1* knockout mice during long cold exposure (Enerbäck *et al.*, 1997). The *UCP1* gene is also indispensable for body temperature maintenance in non-cold circumstances (Tsubota *et al.*, 2019). Variations in bovine *UCP1* have been associated to milk yield, milk fat percentage and milk protein percentage (Zhou *et al.*, 2017), but so far as we know not to thermotolerance.

We can hypothesize that genetic variations within *UCP1* might partly explain the heat tolerance seen in CHCU. These variants should have a negative impact on *UCP1* by reducing its activity or expression.

DIO2, another thermogenic gene located within the BTA10: 92,624,410-92,632,934 interval, encodes iodothyronine deiodinase 2, an enzyme that catalase the conversion of thyroid hormone T_4 to the T_3 active form (Gereben *et al.*, 2008). This enzyme is highly expressed in BAT and mice, where *DIO2* inactivation exhibit impaired thermogenesis, leading to hypothermia during cold exposure (de Jesus *et al.*, 2001). In addition, mice with targeted disruption of *Dio2* have cold-induced overexpression of *Ucp1* (Christoffolete *et al.*, 2004), indicating that thyroid hormone T_3 activates/enhances the transcription of *UCP1*. Furthermore, heat stress performed on C2C12 cells have shown that *DIO2* is down regulated after 6 and 8 days of hyperthermia (Tang *et al.*, 2018). *DIO2* might therefore be also indispensable for body temperature maintenance in non-cold circumstances. No association with productive traits of genetic variants in bovine *DIO2* have been reported sor far. Similarly to *UCP1*, it might be possible that *DIO2* has been under differential selection in CHCU and CHRF populations and we expect that variants *DIO2* would reduce its activity or expression. Among the 162 variants mapping in *DIO2*, we found 37 upstream SNPs but none was fixed in CHCU. Finally, the *ACSL1* is required for cold thermogenesis (Ellis *et al.*, 2010) and encodes acyl-CoA synthetase long chain family member 1. Adipose-specific *Acsl1*$^{-/-}$ mice had greater fat mass when fed a high-fat diet but were also markedly cold intolerant, compared to control mice.

BMPR1A is among the candidates involved in hair development. *BMPR1A* is located within region BTA28: 41,817,915-41,875,990 and encodes the bone morphogenetic protein receptor 1A. Signalling via this receptor is required for the proper differentiation and proliferation of postnatal hair follicles from hair matrix precursor cells. The conditional inactivation of *Bmpr1A* in mouse ventral limb ectoderm and its derivatives (epidermis and hair follicles) results in a lack of external hair (Andl *et al.*, 2004; Kwan *et al.*, 2004; Yuhki *et al.*, 2004). Moreover, Ge *et al.*, (2019) have shown that over-expression of microRNA *miR-29a/b1* induces in mice a short-hair phenotype and eventual hair loss by repressing bone morphogenetic protein signalling through *Bmpr1a* (Ge *et al.*, 2019). Using TargetScan, we

discovered binding sites for several microRNAs in the 3'UTR of bovine *BMPR1A*, including for *miR-29b*. However, we did not find any SNP disrupting the binding site of this microRNA. Further work is required to investigate if some other microRNAs regulateof *BMPR1A* and if SNPs in 3'UTR alter their binding sites. In addition, it is possible that some SNPs in upstream might impact the binding of transcription factors. As previously mentioned, the *CDSN* is also related to hair development. Corneodesmosin is a glycoprotein and an extracellular component of corneodesmosomes (a modified desmosome found in the uppermost layers of the epidermis) and is expressed in the inner root sheath of hair follicles (Simon *et al.*, 1997). Mutations in *CDSN* in human are associated with hypotrichosis simplex, a scalp-specific hair loss (Levy-Nissenbaum *et al.*, 2003) and targeted inactivations of *Cdsn* in mice showed rapid hair loss, confirming the essential role of *Cdsn* for maintaining the architecture of the hair follicle (Matsumoto *et al.*, 2008; Leclerc *et al.*, 2009). Finally, we identified the fibroblast growth factor 7 gene (*FGF7*), within a region under selection (BTA10: 61,005,653-61,068,184). *FGF7*, also known as keratinocyte growth factor (*KGF*), is a member of the FGF family and is expressed in hair follicles (Rosenquist & Martin, 1996). Mice lacking *Fgf7* display a perturbation in the direction of the hairs within the fur coat, giving the hair coat an unkempt appearance. This hair defect seems to be restricted to the cells giving rise to the hair shaft (Guo *et al*., 1996). Newborn transgenic mice expressing human *FGF7/KGF* exhibit wrinkled skin and a reduced hair follicle density (Guo et al., 1993). In addition, administration of human recombinant *FGF7/KGF* induces dose-dependent hair growth over most of the body of nude mice (Danilenko *et al*., 1995). These findings indicate that *FGF7* is an important endogenous mediator of hair follicle growth, development and differentiation.

Conclusions

In summary, our results based on whole-genome data confirm that CHCU samples were closely related to Charolais from France and Canada, but also revealed a limited *B. indicus* introgression into Chacuba. We observed signals of recent adaptation to tropical conditions between CHCU and CHFR founder populations, which were largely independent of the *B. indicus* introgression. Some of those regions harbor genes involved in thermogenesis (*UCP1*, *DIO2* and *ACSL1*) and hair development (*BMPR1A*, *CDSN* and *FGF7*) and variants within these genes may have a functional impact that could explain some of the phenotypic

differences seen between CHCU and CHFR animals. However, experimental work is needed to further infer the involvement of these interesting genes in CHCU adaptation.

List of abbreviations

CHCU: Cuban Charolais

CHFR: French Charolais

CHCA: Canadian Charolais

LIMS: Limousin

BRG: Brangus

BRM: Brahman

TXL: Texas Longhorn

θ: Watterson's nucleotide variability

F_{ST} : differentiation values

EHH: Extended Haplotype Homozygosity

Declarations

Ethics approval

Not applicable.

Consent for publication

Not applicable.

Competing Interests

The authors declare no competing interests.

Funding

MPE and SERO were funded by MINECO grant AGL2016-78709-R and from the EU through the BFU2016-77236-P (MINECO/AEI/FEDER, EU) and the "Centro de Excelencia Severo Ochoa 2016-2019" award SEV-2015-0533. LCRA is funded by "Don Carlos Antonio López" Graduate program (BECAL) from Paraguay. YRC was funded by the European Union H2020 Research and Innovation programme under Marie Skłodowska-Curie grant (P-Sphere) agreement No 6655919.

Authors' contributions

MPE, DR and YRC, designed the study. YRC and YRV performed the sampling. LCRA, SERO, JLA and DR analyzed the data. LCRA, MPE, SERO, DR and YRC interpreted the results and wrote the manuscript. All the authors read and approved the final version of the manuscript.

Acknowledgements

We thank to the *Empresa Genética y Cría* Manuel Fajardo and farmers of *San José del Retiro* for logistic support and help during the sampling.

References

Agarwal V, Bell GW, Nam JW, Bartel DP (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**: e05005.

Ajmone-Marsan P, Garcia JF, Lenstra JA (2010). On the origin of cattle: How aurochs became domestic and colonized the world. *Evolutionary Anthropology* **19**: 148–157.

Alexander DH, Novembre J, Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**: 1655–1664.

Andl T, Ahn K, Kairo A, Chu EY, Wine-Lee L, Reddy ST, *et al.* (2004). Epithelial *BMPR1A* regulates differentiation and proliferation in postnatal hair follicles and is essential for tooth development. *Development* **131**: 2257–2268.

Boussaha M, Esquerré D, Barbieri J, Djari A, Pinton A, Letaief R, *et al.* (2015). Genome-wide study of structural variants in bovine Holstein, Montbéliarde and Normande dairy breeds. *PLoS ONE* **10**: e0135931.

Browning BL, Browning SR (2016). Genotype Imputation with Millions of Reference Samples. *American Journal of Human Genetics* **98**: 116–126.

Burgos-Paz W, Souza CA, Megens HJ, Ramayo-Caldas Y, Melo M, Lemús-Flores C, *et al.* (2013). Porcine colonization of the Americas: A 60k SNP story. *Heredity* **110**: 321–330.

Chouchani ET, Kazak L, Spiegelman BM (2019). New Advances in Adaptive Thermogenesis: *UCP1* and Beyond. *Cell Metabolism* **29**: 27–37.

Christoffolete MA, Linardi CCG, De Jesus L, Ebina KN, Carvalho SD, Ribeiro MO, *et al.* (2004). Mice with Targeted Disruption of the *DIO2* Gene Have Cold-Induced Overexpression of the Uncoupling Protein 1 Gene but Fail to Increase Brown Adipose Tissue Lipogenesis and Adaptive Thermogenesis. *Diabetes* **53**: 577–584.

Danilenko DM, Ring BD, Yanagihara D, Benson W, Wiemann B, Starnes CO, *et al.* (1995). Keratinocyte growth factor is an important endogenous mediator of hair follicle growth, development, and differentiation: Normalization of the nu/nu follicular differentiation defect amelioration of chemotherapy-induced alopecia. *American Journal of Pathology* **147**: 145–154.

Ellis JM, Li LO, Wu PC, Koves TR, Ilkayeva O, Stevens RD, *et al.* (2010). Adipose Acyl-CoA synthetase-1 directs fatty acids toward β-oxidation and is required for cold thermogenesis. *Cell Metabolism* **12**: 53–64.

Enerbäck S, Jacobsson A, Simpson EM, Guerra C, Yamashita H, Harper ME, *et al.* (1997). Mice lacking mitochondrial uncoupling protein are cold-sensitive but not obese. *Nature* **387**: 90–94.

Ferretti L, Raineri E, Ramos-Onsins S (2012). Neutrality tests for sequences with missing data. *Genetics* **191**: 1397–1401.

Ge M, Liu C, Li L, Lan M, Yu Y, Gu L, *et al.* (2019). miR-29a/b1 Inhibits Hair Follicle Stem Cell Lineage Progression by Spatiotemporally Suppressing WNT and BMP Signaling. *Cell Reports* **29**: 2489-2504.e4.

Gereben B, Zavacki AM, Ribich S, Kim BW, Huang SA, Simonides WS, *et al.* (2008). Cellular and molecular basis of deiodinase-regulated thyroid hormone signaling. *Endocrine Reviews* **29**: 898–938.

Ginja C, Gama LT, Cortés O, Burriel IM, Vega-Pla JL, Penedo C, *et al.* (2019). The genetic ancestry of American Creole cattle inferred from uniparental and autosomal genetic markers. *Scientific Reports* **9**: 11486.

Guan Y (2014). Detecting structure of haplotypes and local ancestry. *Genetics* **196**: 625–642.

Guillocheau GM, El Hou A, Meersseman C, Esquerré D, Rebours E, Letaief R, *et al.* (2019). Survey of allele specific expression in bovine muscle. *Scientific Reports* **9**: 4297.

Guo L, Degenstein L, Fuchs E (1996). Keratinocyte growth factor is required for hair development but not for wound healing. *Genes and Development* **10**: 165–175.

Guo L, Yu QC, Fuchs E (1993). Targeting expression of keratinocyte growth factor to keratinocytes elicits striking changes in epithelial differentiation in transgenic mice. *The EMBO Journal* **12**: 973–986.

Hudson RR, Slatkin M, Maddison WP (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.

de Jesus LA, Carvalho SD, Ribeiro MO, Schneider M, Kim SW, Harney JW, *et al.* (2001). The type 2 iodothyronine deiodinase is essential for adaptive thermogenesis in brown adipose tissue. *Journal of Clinical Investigation* **108**: 1379–1385.

Kasarapu P, Porto-Neto LR, Fortes MRS, Lehnert SA, Mudadu MA, Coutinho L, *et al.* (2017). The *Bos taurus-Bos indicus* balance in fertility and milk related genes.

PLoS ONE **12**: e0181930.

Khamis AM, Motwalli O, Oliva R, Jankovic BR, Medvedeva YA, Ashoor H, *et al.* (2018). A novel method for improved accuracy of transcription factor binding site prediction. *Nucleic Acids Research* **46**: e72.

Knüppel R, Dietze P, Lehnberg W, Frech K, Wingender E (1994). TRANSFAC Retrieval Program: A Network Model Database of Eukaryotic Transcription Regulating Sequences and Proteins. *Journal of Computational Biology* **1**: 191–198.

Koufariotis L, Hayes BJ, Kelly M, Burns BM, Lyons R, Stothard P, *et al.* (2018). Sequencing the mosaic genome of Brahman cattle identifies historic and recent introgression including polled. *Scientific Reports* **8**: 17761.

Kwan KM, Li AG, Wang XJ, Wurst W, Behringer RR (2004). Essential roles of BMPR-IA signaling in differentiation and growth of hair follicles and in skin tumorigenesis. *Genesis* **39**: 10–25.

Leclerc EA, Huchenq A, Mattiuzzo NR, Metzger D, Chambon P, Ghyselinck NB, *et al.* (2009). Corneodesmosin gene ablation induces lethal skinbarrier disruption and hair-follicle degeneration related to desmosome dysfunction. *Journal of Cell Science* **122**: 2699–2709.

Letaief R, Rebours E, Grohs C, Meersseman C, Fritz S, Trouilh L, *et al.* (2017). Identification of copy number variation in French dairy and beef breeds using next-generation sequencing. *Genetics, selection, evolution* **49**: 77.

Levy-Nissenbaum E, Betz RC, Frydman M, Simon M, Lahat H, Bakhan T, *et al.* (2003). Hypotrichosis simplex of the scalp is associated with nonsense mutations in *CDSN* encoding corneodesmosin. *Nature Genetics* **34**: 151–153.

Li H, Durbin R (2009). Making the Leap: Maq to BWA. *Mass Genomics* **25**: 1754–1760.

Lin BZ, Sasazaki S, Mannen H (2010). Genetic diversity and structure in *Bos taurus* and *Bos indicus* populations analyzed by snp markers. *Animal Science Journal* **81**: 281–289.

Matsumoto M, Zhou Y, Matsuo S, Nakanishi H, Hirose K, Oura H, *et al.* (2008). Targeted deletion of the murine corneodesmosin gene delineates its essential role in skin and hair physiology. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 6720–6724.

Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, *et al.* (2010). The Genome Analysis Toolkit : A MapReduce framework for analyzing

next-generation DNA sequencing data. *Genome Research* **20**: 1297–1303.

McTavish EJ, Decker JE, Schnabel RD, Taylor JF, Hillis DM (2013). New World cattle show ancestry from multiple independent domestication events. *Proceedings of the National Academy of Sciences of the United States of America* **110**: E1398-E1406.

Naves M, Flori L, Thevenon S, Gautier M (2015). Adaptation of tropical cattle breeds to their environment, in the perspective of climatic change. In: Livret du Workshop Recolad 2015. INRA. Paris: INRA, Résumé, p. 38. Workshop Recolad 2015, Paris, France, 11 February 2015/12 February 2015.

Nicholls DG, Bernson VS, Heaton GM (1978). The identification of the component in the inner membrane of brown adipose tissue mitochondria responsible for regulating energy dissipation. *Experientia Supplementum* **32**: 89–93.

Pérez-Enciso M, De Los Campos G, Hudson N, Kijas J, Reverter A (2017). The 'heritability' of domestication and its functional partitioning in the pig. *Heredity (Edinb)* **118**: 160–168.

Renand G, Menendez-Buxadera A, Krauss D, Menissier F (1997). Comparaison expérimentale du Charolais Cubain sélectionné en conditions tropicales avec du Charolais Français. *Rencontres autour des Recherches sur les Ruminants* **4**: 230.

Ribas M (1981). Gene frequencies in the blood group systems of the Cuban Charolais. *Annales de Génétique et de Sélection Animale* **13**: 293-300.

Rodriguez-Valera Y, Renand G, Naves M, Fonseca-Jiménez Y, Moreno-Probance TI, Ramos-Onsins S, *et al.* (2018). Genetic diversity and selection signatures of the beef 'Charolais de Cuba' breed. *Scientific Reports* **8**: 11005.

Rosenquist TA, Martin GR (1996). Fibroblast Growth Factor Signalling in the Hair Growth Cycle: Expression of the Fibroblast Growth Factor Receptor and Ligand Genes in the Murine Hair Follicle. *Development Dynamics* **205**: 379–386.

Sandeli A, Alkema W, Engström P, Wasserman WW, Lenhard B (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Research* **32**: D91-D94.

Simon M, Montézin M, Guerrin M, Durieux JJ, Serre G (1997). Characterization and purification of human corneodesmosin, an epidermal basic glycoprotein associated with corneocyte-specific modified desmosomes. *Journal of Biological Chemistry* **272**: 31770–31776.

Tang J, He A, Yan H, Jia G, Liu G, Chen X, *et al.* (2018). Damage to the myogenic differentiation of C2C12 cells by heat stress is associated with up-regulation of several selenoproteins. *Scientific Reports* **8**: 10601.

Tang K, Thornton KR, Stoneking M (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biology* **5**: e171.

The Bovine HapMap Consortium (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**: 528–532.

Tsubota A, Okamatsu-Ogura Y, Bariuan JV, Mae J, Matsuoka S, Nio-Kobayashi J, *et al.* (2019). Role of brown adipose tissue in body temperature control during the early postnatal period in syrian hamsters and mice. *Journal of Veterinary Medical Science* **81**: 1461–1467.

Upadhyay M, Bortoluzzi C, Barbato M, Ajmone-Marsan P, Colli L, Ginja C, *et al.* (2019). Deciphering the patterns of genetic admixture and diversity in southern European cattle using genome-wide SNPs. *Evolutionary Applications* **12**: 951–963.

Watterson GA (1975). On the Number of Segregating Sites in Genetical Models without Recombination. *Theoretical Population Biology* **7**: 256–276.

Yuhki M, Yamada M, Kawano M, Iwasato T, Itohara S, Yoshida H, *et al.* (2004). *BMPR1A* signaling is necessary for hair follicle cycling and hair shaft differentiation in mice. *Development* **131**: 1825–1833.

Zhou H, Cheng L, Byun SO, Aizimu W, Hodge MC, Edwards GR, *et al.* (2017). Haplotypic variation in the *UCP1* gene is associated with milk traits in dairy cows. *Journal of Dairy Research* **84**: 68–75.

SUPPLEMENTARY MATERIAL

Figure S1. Proportion of *B. indicus* per CHCU samples

Figure S2. Relationship between proportion of *B. indicus* in CHCU and Fst CHCU-CHFR, each dot corresponds to a 30kb window

Table S1. Description of the samples employed in the study

Table S2. Description of the intervals identified as selective sweeps between Cuban and French Charolais

Table S3. Putative regulatory SNPs (rSNPs) located upstream of *UCP1* gene

Paper II

SeqBreed: a python tool to evaluate genomic prediction in complex scenarios

Miguel Pérez-Enciso[1,2*], Lino C. Ramírez-Ayala[1], Laura M. Zingaretti[1,3]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, 08193 Bellaterra, Barcelona, Spain

[2]ICREA, Passeig de Lluís Companys 23, 08010 Barcelona, Spain

[3]Universidad Nacional de Villa María, IAPBCyA- IAPCH Villa María, Córdoba, Argentina

*Corresponding author

Genet Sel Evol ***52***, *7 (2020)*

https://doi.org/10.1186/s12711-020-0530-2

Abstract

Background

Genomic prediction (GP) is a method whereby DNA polymorphism information is used to predict breeding values for complex traits. Although GP can significantly enhance predictive accuracy, it can be expensive and difficult to implement. To help design optimum breeding programs and experiments, including genome-wide association studies and genomic selection experiments, we have developed SeqBreed, a generic and flexible forward simulator programmed in python3.

Results

SeqBreed accommodates sex and mitochondrion chromosomes as well as autopolyploidy. It can simulate any number of complex phenotypes that are determined by any number of causal loci. SeqBreed implements several GP methods, including genomic best linear unbiased prediction (GBLUP), single-step GBLUP, pedigree-based BLUP, and mass selection. We illustrate its functionality with drosophila genome reference panel (DGRP) sequence data and with tetraploid potato genotype data.

Conclusions

SeqBreed is a flexible and easy to use tool that can be used to optimize GP or genome-wide association studies. It incorporates some of the most popular GP methods and includes several visualization tools. Code is open and can be freely modified. Software, documentation, and examples are available at https://github.com/miguelperezenciso/SeqBreed.

Background

Genomic prediction (GP) is a method whereby DNA polymorphism information is used to predict the breeding value of individuals for complex traits. The availability of high-throughput single nucleotide polymorphism (SNP) genotyping in a cost-effective manner has led GP to become a standard tool in the analysis and improvement of complex traits (Meuwissen *et al.*, 2013). GP has revolutionized breeding programs in plants and animals and, today, GP methods are also widely used in human genetics or ecology. Nevertheless, GP

is more expensive than traditional pedigree – based breeding. GP can be difficult to implement in practical scenarios, due in part to the difficulty of optimizing genotyping strategies and to uncertainty about the genetic basis of complex traits. Thus, it is highly advisable to evaluate its potential advantages and expected performance in advance. GP accuracy depends on a large number of factors. Several can be controlled by the practitioner, to some extent, such as the number of SNPs, number of individuals, selection intensity, and the evaluation method. Other factors cannot be modified, such as linkage disequilibrium and are even unknown (genetic architecture). Although several approximations of the accuracy of GP have been developed, e.g. (Daetwyler *et al.*, 2008; Goddard, 2009), it remains difficult to analytically assess the influence of these factors in practical scenarios across generations. For this purpose, stochastic computer simulation is the most reliable option. Although critical factors such as the detailed genetic architecture of complex traits are unknown, the main genetic parameters are reasonably well known for most complex traits, such as heritability and the distribution of genetic effects, which can be approximated by a gamma distribution (Hayes & Goddard, 2001; Eyre-Walker & Keightley, 2007; Caballero *et al.*, 2015). Thus, a simulation study can be performed to evaluate the effect of the number of causal loci quantitative trait nucleotides (QTN) and of their location to assess the robustness of predictions.

Here, we present a versatile python3 forward simulation tool, SeqBreed, to evaluate GP performance in generic scenarios and with any genetic architecture (i.e., number of QTN, their effects and location, and the number of traits). The purpose of SeqBreed is to generate phenotype and genotype data of individuals under different (genomic) selection strategies. SeqBreed is inspired by a previous pSBVB fortran software program (Zingaretti *et al.*, 2019), but the code has been rewritten in python3 and many new options have been added. Python can be much slower than compiled languages, but is much easier and friendlier to use, allowing direct interaction with the user to, e.g., make plots or control selection and breeding decisions. In addition, many libraries in python, such as ‘numpy’ (https://numpy.org/) or ‘pandas’ (https://pandas.pydata.org/), are wrappers on compiled languages, such that careful programming significantly alleviates the limited speed of native python. Thus, SeqBreed is much more versatile than pSBVB and incorporates many new options, such as genome-wide association studies (GWAS) and principal component analysis (PCA). Most importantly, it allows automatic implementation of standard genomic selection procedures. Usage details and

the main features of SeqBreed are described in the following and in the accompanying GitHub site https://github.com/miguelperezenciso/SeqBreed.

Implementation

Outline

Broadly, SeqBreed takes genotype / sequence data from a founder population and simulates phenotypes according to a predetermined genetic architecture. Offspring genomes and phenotypes can be simulated under selection or random drift. By default, selection is simulated across a predetermined number of generations and selection intensities. SeqBreed offers extensive flexibility to the user. For example, accuracy of GP with several SNP arrays can be simultaneously compared using the same data; offspring of specific pairs of parents can be generated; and dihaploid offspring can be simulated. SeqBreed can be run using scripts or interactively, where the user can, say, obtain plots for each generation or generate genotype data of a given set of individuals. Examples of the program's usage are in the GitHub's jupyter notebook https://github.com/miguelperezenciso/SeqBreed/blob/master/SeqBreed_tutorial.ipynb and in the python script https://github.com/miguelperezenciso/SeqBreed/blob/master/main.py.

SeqBreed is programmed in python3 using an object-oriented paradigm. The generic SeqBreed flowchart is visualized in Fig. 1.

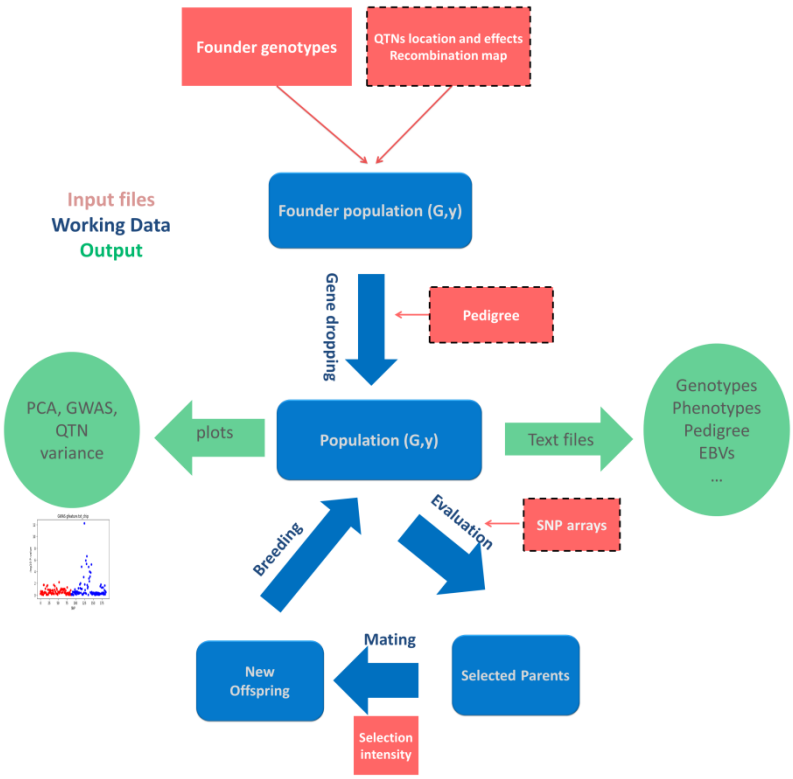

Figure 1. Outline of the SeqBreed pipeline. Inputs are shown in red squares, dashed border rectangles represent optional input, internal data are in blue rounded squares, main operations are indicated in blue, and outputs are in green circles; G and y refer to genotypes and phenotypes, respectively. The program starts with an optional gene dropping step following an input pedigree. No selection is performed at this stage. The bottom loop represents selection, where new offspring are generated based on the genotypes of selected parents. A list of SNPs in the genotyping array must be determined when using GBLUP and BLUP. A new cycle starts when these new offspring are added to the existing population. Plots can be performed at several stages.

As input, SeqBreed minimally requires a genotype file from the founder base population in vcf (Li *et al.*, 2009) or plink-like format (Chang *et al.*, 2015). A typical SeqBreed run consists of the following steps:

1. Upload founder sequence genotypes in vcf or plink format. The program automatically determines ploidy and the number of chromosomes and SNPs.
2. Specify genome characteristics. Sex-linked SNPs and/or recombination rates can be

specified.

3. Specify desired heritabilities and causal SNPs (QTN) and their effects for every trait. Environmental variances are inferred given founder genomes, QTN effects and heritabilities.
4. Offspring genomes and phenotypes are simulated by gene-dropping along a predetermined pedigree or by implementing selection.

PCA plots or GWAS options are also implemented. The main python classes are:

- Population: This class contains the main attributes for running selection experiments and is a container for Individual objects. It includes methods to add new individuals generated by mating two parents or randomly shuffling founder genomes in order to increase the number of base population animals (see (Pérez-Enciso *et al.*, 2017)). It also prints basic population data and summary plots.
- Individual: It allows generation, manipulation, and printing of individuals' genotypes and phenotypes. Internally, an individual's genome is represented by contiguous non recombining blocks rather than by the list of all SNP alleles, which allows dramatic savings in memory and increases in efficiency (see Figure 1 in Pérez-Enciso et al. (Pérez-Enciso *et al.*, 2017)).
- Genome: All genome characteristics are stored and can be accessed by methods in this class. It specifies ploidy, number and class of chromosomes, and recombination rates or SNP positions.
- GFounder: SeqBreed requires as minimum input the genotypes of the so-called 'founder population', which comprises the parents of the rest of individuals to be generated. This class stores these genotypes and automatically retrieves main genome features such as SNP positions, number of chromosomes, etc. Initial genotypes can be filtered by minimum allele frequency (MAF).
- QTN: This class determines the genetic architecture for each trait simulated. It has methods to determine the environmental variance given a desired heritability, and to plot variance components for QTN. In its current version, SeqBreed allows for dominance and additive actions, but not epistasis.
- Chip: This class is basically a container for the list of SNPs that are included in a genotyping array. It allows easy comparison of different genotyping strategies in genomic selection.

Specifying genome features and genetic architecture

By default, SeqBreed assumes that all loci are autosomal and a recombination rate of 1 cM = 1 Mb throughout the genome. It includes options to specify sex or mitochondrial chromosomes, and local and sex specific recombination maps. A pseudo-autosomal region (PAR) is not accommodated for sex chromosomes, i.e., the whole Y chromosome is assumed to be non-recombining. A mitochondrial chromosome is a non-recombining chromosome that is transmitted maternally. SeqBreed allows for autopolyploidy of any level, which is automatically detected from vcf files. Accurate modeling of meiosis in polyploids is notoriously difficult (Baduel *et al.*, 2018; Jighly *et al.*, 2018) and SeqBreed implements a simplified algorithm:

1. For each chromosome id, homologs are randomly paired.
2. Within each pair of homologs, cross-over events are generated as for diploids, i.e., no interaction between homolog chromosomes is modeled and the number of cross-over events is simulated following a Poisson distribution with a rate equal to chromosome length in Morgans.
3. Sex chromosomes are modeled with a maximum ploidy of 2.

Therefore, our algorithm does not fully model the interaction of preferential pairing of homologous chromosomes and double reduction arising from multivalent formation (Voorrips & Maliepaard, 2012). For the purposes of this software (i.e. comparison of GP strategies over a limited number of generations), it is unlikely that this approximation has a dramatic effect.

SeqBreed allows the simulation of any number of phenotypic traits, regardless of ploidy. For each trait, broad heritability must be specified. There are three options to specify the number of QTN and their effects (https://github.com/miguelperezenciso/SeqBreed#3-specifying-genetic-architecture): (i) a random number of QTN positions are sampled genome-wide and additive effects are sampled from a gamma distribution Γ (shape = 0.2 and scale = 5), as suggested by Caballero et al. (Caballero *et al.*, 2015); (ii) the positions of the QTN are specified in a file and additive effects are sampled from a gamma distribution; and (iii) QTN positions and additive and dominant effects for each trait are specified in an external file. By default, QTN are not removed from the sequence data to perform genomic evaluation. To remove QTN from evaluation, a SNP chip can be defined that excludes the QTN. Options (i) and (ii) can only be used with one trait and without dominance. SeqBreed adjusts the environmental variance $Var(e)$ to retrieve the desired broad-sense heritabilities (H^2) from

$Var(e) = Var(g) \times (1 - \mathrm{H}^2)/\mathrm{H}^2$, where $Var(g)$ is the variance of the genotypic values of individuals in the founder population. The genotypic value for individual i is defined as:

$$g_i = \sum_{j=1}^{nQTN} \gamma_{ij}\, a_j + \sum_{j=1}^{nQTN} \delta_{ij}\, d_j,$$

where $nQTN$ is the number of QTN, a_j is the additive effect of the j-th QTN, that is, half the expected difference between homozygous genotypes, with γ_{ij} taking the values -1, 0 and 1 for homozygous, heterozygous, and alternative homozygous genotypes, respectively, d_j is the dominance effect of the j-th QTN, with δ_{ij} taking the value 1 if the genotype is heterozygous and 0 otherwise. In the case of polyploids:

$$g_i = \sum_{j=1}^{nQTN} \eta_{ij}\, a_j + \sum_{j=1}^{nQTN} \varphi_{ij}\, d_j,$$

where η_{ij} is the number of copies of the alternative allele (coded as 1) minus half the ploidy for the j-th QTN and the i-th individual, and a_j is therefore the expected change in phenotype per copy of allele '1' at the j-th QTN. In polyploids, technically as many dominance coefficients as ploidy levels (h) minus two can be defined, which is not practical. As in pSBVB (Zingaretti *et al.*, 2019), we define φ_{ij} as the minimum number of copies of allele '1' such that the expected phenotype is d (see Figure 1 in Zingaretti et al. (Zingaretti *et al.*, 2019)). SeqBreed uses $\varphi_{ij} = 1$, that is, all heterozygous individuals have the same genotype value as the complete homozygous '1'. SeqBreed computes genotypic values for each individual and simulate phenotypes from $y_i = \mu + g_i + e_i$, where μ is a constant and e is a normal deviate $e \sim N\left(0, Var(e)\right)$.

For multiple traits, the user needs to specify additive and dominant QTN effects separately for each trait. This is done via an external text file, where additive and dominant QTN effects are specified for each trait (option 3 in https://github.com/miguelperezenciso/SeqBreed#3-specifying-genetic-architecture). There is no specific assumption on genetic correlations between traits. To simulate no pleiotropy, QTN for each trait have zero effects for all other traits. Note that this does not prevent a non-zero genetic correlation arising from linkage disequilibrium. The program does not automatically adjust desired genetic correlations and, thus, different QTN values may need to be tested to fit desired correlations. Environmental

correlations are always zero.

It is typically difficult to find real sequence data to generate a reasonably sized founder population. To accommodate this, SeqBreed can generate 'dummy' founder individuals by randomly combining recombinant haplotypes. This can be done in two ways, either by generating a random pedigree and simulating a new founder individual by gene-dropping along this pedigree, or by directly simulating a number of recombining breakpoints and assigning random founder genotypes to each block between recombination breakpoints (https://github.com/miguelperezenciso/SeqBreed/blob/master/README.md#breeding-population).

Gene dropping and selection implementation

Seqbreed can be run along a predetermined pedigree or using a combination of options (several examples are provided in the GitHub site). It is also possible to generate new individuals interactively, including dihaploids. To speed up computations and avoid unnecessary memory usage, only recombination breaks and ancestor haplotype ids are stored for each individual (see Figure 1 in (Pérez-Enciso *et al.*, 2000)).

SeqBreed allows computing estimated breeding values using several GP methods. It also allows several lists of SNPs (SNP chips) to be defined, such that GP performance can be easily compared across chips. From a methodological point of view, most GP implementations are based on penalized linear methods (e.g., de los Campos et al., [14]). SeqBreed includes some of the most popular GP options, including pedigree BLUP (Henderson, 1984), GBLUP (VanRaden, 2008), and single-step GBLUP (Legarra *et al.*, 2009). Only single trait GP algorithms are implemented so far. Mass selection is also implemented. For GBLUP and single-step GBLUP, the genomic relationship matrix $\mathbf{G}$ is obtained using VanRaden (VanRaden, 2008) as:

$$\mathbf{G} = \frac{\mathbf{X}\mathbf{X}'}{2\sum_{j=1}^{nSNP} p_j(1-p_j)},$$

where $\mathbf{X}$ is a $N \times n$SNP matrix containing genotypes (coded 0,1,2 deviated from the mean) for SNPs on the chip, p_j is the allele frequency of the j-th SNP, N is the number of individuals and nSNP is the number of SNPs. To avoid potential singularity problems, diagonal elements

of $\mathbf{G}$ are multiplied by 1.05. SeqBreed requires that heritabilities to be used in BLUP or GBLUP are provided (i.e., they are not estimated). The program allows the incorporation of other custom GP methods based on a user python function or by exporting SNP data and phenotypes from SeqBreed, running a genetic evaluation, externally and then importing the resulting estimated breeding values.

Selection can be automatically configured and run, as documented in the GitHub examples (https://github.com/miguelperezenciso/SeqBreed). Running a selection scheme requires specifying the number of generations, the numbers of females and males to be selected, and the number of offspring per female. SeqBreed splits the selection process in three steps, which allows a fine control over the breeding program. First, breeding values are predicted using the chosen evaluation method and marker information. By default, the data from all individuals across the current and previous generations are used, but this can be changed by specifying the subset of individuals to be used. Second, a function is used to generate offspring from selected parents. This function requires specifying the candidates for selection (allowing for continuous or discrete generations), selection intensity, family size, and either assortative or random mating between selected parents. Hierarchical mating between females and males is employed by default (https://github.com/miguelperezenciso/SeqBreed#7-implementing-selection). Assortative and random mating schemes are implemented; more sophisticated mating schemes, such as based on optimal contributions (Sonesson and Meuwissen, 2000; Sánchez *et al.*, 2003), have to be specified manually by modifying the function 'ReturnNewPed' in the selection module (https://github.com/miguelperezenciso/SeqBreed/blob/master/src/selection.py).

Visualization

A novel feature of SeqBreed, as compared to our previous software pSBVB, is the capability of graphical outputs. Figure 2 illustrates some of the plots that can be performed automatically. Figure 2a shows the results of the QTN.plot() function, which plots the individual QTN variance as a function of MAF, the histogram of QTN variances, and the cumulative variance when QTN are sorted by MAF. This is performed for each phenotype and for both additive and dominance variances, based on allele substitution effects $\alpha = a + d(1-2p)$, where p is the minimum allele frequency, and assuming complete equilibrium. In

addition, PCA plots using all sequence or custom defined SNP sets (Fig. 2b) are available, as well as GWAS plots showing p-values or false discovery rate (FDR) values (Fig. 2c). Genotype and phenotype data can also be exported in text files.

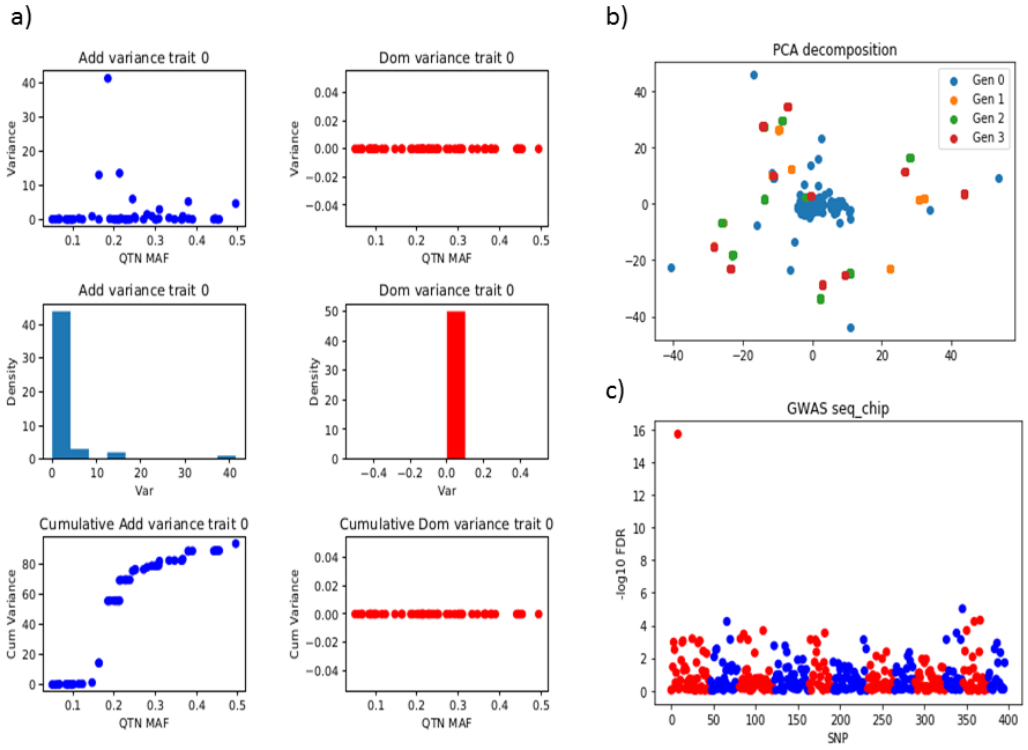

Figure 2. Example plots produced by SeqBreed. **(a**) Contribution of each QTN to total variance; **(b**) Principal component analysis plot; individuals of different generations are in different colors; **(c)** Genome wide association study showing false discovery rate values (-log10 scale)

Usage and examples

The basic functioning of SeqBreed is illustrated by the main.py script that is available at https://github.com/miguelperezenciso/SeqBreed/blob/master/main.py. This script, or its equivalent jupyter notebook (SeqBreed_tutorial.ipynb), shows the basic commands to run SeqBreed and import the required modules. First, SeqBreed modules are imported as:

```
from SeqBreed import genome as gg
from SeqBreed.selection import selection as sel
```

Founder population genotypes are uploaded from the vcf file, using the command:

```
gbase = gg.GFounder(vcfFile=vcffile, snpFile=seqfile)
```

which generates a GFounder object that contains founder genotypes, vcffile is the file containing genotypes and seqfile is generated by the program and contains information about SNP positions, which are used in the next step.

Next, the main genome features are specified. The following command creates a Genome object that assumes that the 'X' chromosome is the sex X chromosome, while SNPs on the chromosome named 'MT' are mitochondrial.

```
# seqfile is a file obtained in previous step
gfeatures = gg.Genome(snpFile=seqfile, mapFile=mapfile, ploidy=gbase.ploidy, XChr='X', MTChr='MT')
```

Genetic architecture can be specified in different ways. The simplest is to generate nqtn QTN randomly distributed along the genome with effects sampled from a gamma distribution, where h2 is the desired heritability.

```
# 10 QTNs are simulated, h2 of the trait is 0.7
qtn = gg.QTNs(h2=[0.7], genome=gfeatures, nqtn=10)
# environmental variances are computed
qtn.get_var(gfeatures, gbase)
```

Selection is implemented in cycles, the number of generations, the numbers of males and females selected, and family size must be specified. The following is an example with GBLUP selection, random mating, and continuous generations.

```
ngen = 5          # no. of selection generations
nsel = [5, 10]    # no. of males and females selected
noffspring = 10 # no. offspring per female

# selection cycles
for t in range(ngen):
   # STEP 0: generate marker data for evaluation, stored in X matrix
   # pop is a Population object containing individual genomes and phenotypes
   # chip0 is a Chip object containing SNPs to be used in genomic evaluation
   X = gg.do_X(pop.inds, gfeatures, gbase, chip0)

   # STEP 1: estimate breeding values using criterion GBLUP, assuming h2=0.3
   # criterion can take values 'random', 'phenotype', 'blup', 'gblup' or 'sstep'
   sel.doEbv(pop, criterion='gblup', X=X, h2=0.3, nh=gfeatures.ploidy)

   # STEP 2: pedigree with offspring of selected individuals
   # mating can be ‘assortative’ or ‘random’
   # generation indicates that individuals from generation onwards are considered
as selection candidates
   newPed = sel.ReturnNewPed(pop, nsel, famsize=noffspring, mating='random',
generation=0)

   # STEP 3: generates new offspring (this function adds QTN genotypes, true bvs
and phenotypes)
   # New individuals are added to current Population
   pop.addPed(newPed, gfeatures, qtn, gbase)
```

We illustrate the software with sequence data from the Drosophila genome reference panel (DGRP, (Huang *et al.*, 2014)), parsed and filtered as explained in (Forneris *et al.*, 2017), and genotype data from tetraploid potato (Enciso-Rodriguez *et al.*, 2018), parsed as described in (Zingaretti *et al.*, 2019). Data and scripts are in https://github.com/miguelperezenciso/SeqBreed/tree/master/DGRP and in https://github.com/miguelperezenciso/SeqBreed/tree/master/POTATO for the Drosophila and potato examples, respectively. The DGRP scripts illustrate the specific recombination map of Drosophila, where males do not recombine, as shown in the ‘dgrp.map’ file. The example provided in GitHub consists of a small experiment to compare genomic and mass selection. Plots in the jupyter notebook are implemented to track phenotypic changes by generation. The potato scripts illustrate how to generate an F2 cross between extreme lines and to perform a GWAS experiment in polyploids. GWAS results using PCA corrected phenotypes are also shown.

Conclusions and future developments

Several other programs have been developed for similar purposes as SeqBreed, including our own pSBVB (Zingaretti *et al.*, 2019), AlphaSim (Faux *et al.*, 2016) and its successor AlphaSimR (https://alphagenes.roslin.ed.ac.uk/wp/software-2/alphasimr/), PedigreeSim (Voorrips & Maliepaard, 2012), simuPOP (Peng & Kimmel, 2005), and QMSim (Sargolzaei & Schenkel, 2009). However, SeqBreed offers a unique combination of features for simulation of GP of complex traits, including built-in implementation of several GP methods, the possibility of simulating polyploid genomes, and several options to specify QTN and SNP arrays. It also allows new individuals to be generated interactively and provides graphical plots of results. It is easy to use, easy to install, and software options are illustrated with several examples in the GitHub site. Given the interactive nature of python and its graphical features, SeqBreed is especially suited for educational purposes. However, for large-scale simulations SeqBreed will not be as efficient as some fortran counterparts such as AlphaSim or pSBVB.

Note that SeqBreed was designed to evaluate the performance of GP or GWAS over a short time horizon, i.e., new mutations are not generated. SeqBreed is not designed to investigate the long-term effects of demography or selection on DNA variability because new mutations are not generated. For these purposes, Slim (Messer, 2013) or similar tools are more appropriate. To investigate realistic scenarios, the recommended input for SeqBreed is real sequence data.

Plans for further development of SeqBreed include additional features to generalize available genetic architectures (e.g., imprinting, epistasis), integration with machine-learning tools (scikit, keras) for genetic evaluation, development of an educational tool with an html-based interface, and improving output and plotting features.

Declarations

Availability of data and materials

https://github.com/miguelperezenciso/SeqBreed

Availability and requirements

- Project name: SeqBreed
- Project home page: https://github.com/miguelperezenciso/SeqBreed
- Operating systems: Tested in linux and mac. It should also run in windows python.
- Programming language: Python.
- License: GNU GPLv3
- Any restrictions to use by non-academics: None.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Funding

This work was supported by a PhD grant from the Ministry of Economy and Science (MINECO, Spain) to LMZ, by MINECO grant AGL2016-78709-R and from the EU through the BFU2016-77236-P (MINECO/AEI/FEDER, EU) to MPE and the “Centro de Excelencia Severo Ochoa 2016-2019” award SEV-2015-0533. LCRA is funded by "Don Carlos Antonio López" Graduate program (BECAL) from Paraguay.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

MPE conceived research. MPE and LMZ wrote software and documentation. MPE, LMZ and LCRA tested and validated the program.

Reference

Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L (2018). The 'Polyploid Hop': Shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Frontiers in Ecology and Evolution* **6**: 117.

Caballero A, Tenesa A, Keightley PD (2015). The nature of genetic variation for complex traits revealed by GWAS and Regional Heritability Mapping analyses. *Genetics* **201**: 1601–1613.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4**: 7.

Daetwyler HD, Villanueva B, Woolliams JA (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* **3**: e3395.

Enciso-Rodriguez F, Douches D, Lopez-Cruz M, Coombs J, de los Campos G (2018). Genomic selection for late blight and common scab resistance in tetraploid potato (Solanum tuberosum). *G3 (Bethesda)* **8**: 2471–2481.

Eyre-Walker A, Keightley PD (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics* **8**: 610–618.

Faux AM, Gorjanc G, Gaynor RC, Battagin M, Edwards SM, Wilson DL, *et al.* (2016). AlphaSim: Software for Breeding Program Simulation. *The Plant Genome* **9**: 10.3835/plantgenome2016.02.0013.

Forneris NS, Vitezica ZG, Legarra A, Pérez-Enciso M (2017). Influence of epistasis on response to genomic selection using complete sequence data. *Genetics Selection Evolution* **49**: 66.

Goddard M (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* **136**: 245–257.

Hayes B, Goddard ME (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution* **33**: 209–229.

Henderson CR (1984). *Applications of linear models in animal breeding*. Guelph: University of Guelph.

Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, *et al.* (2014). Natural Variation in Genome Architecture Among 205 *Drosophila melanogaster* Genetic Reference Panel Lines. *Genome Research* **24**: 1193–1208.

Jighly A, Lin Z, Forster JW, Spangenberg GC, Hayes BJ, Daetwyler HD (2018). Insights into population genetics and evolution of polyploids and their ancestors. *Molecular Ecology Resources* **18**: 1157–1172.

Legarra A, Aguilar I, Misztal I (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* **92**: 4656–4663.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* **193**: 327–345.

Messer PW (2013). SLiM: Simulating evolution with selection and linkage. *Genetics* **194**: 1037–1039.

Meuwissen T, Hayes B, Goddard M (2013). Accelerating Improvement of Livestock with Genomic Selection. *Annual Review of Animal Biosciences* **1**: 221–237.

Peng B, Kimmel M (2005). simuPOP: A forward-time population genetics simulation environment. *Bioinformatics* **21**: 3686–3687.

Pérez-Enciso M, Forneris N, de los Campos G, Legarra A (2017). Evaluating sequence-based genomic prediction with an efficient new simulator. *Genetics* **205**: 939–953.

Pérez-Enciso M, Varona L, Rothschild MF (2000). Computation of identity by descent probabilities conditional on DNA markers via a Monte Carlo Markov Chain method. *Genetics Selection Evolution* **32**: 467–482.

Sánchez L, Bijma P, Woolliams JA (2003). Minimizing inbreeding by managing genetic contributions across generations. *Genetics* **164**: 1589–1595.

Sargolzaei M, Schenkel FS (2009). QMSim: A large-scale genome simulator for livestock. *Bioinformatics* **25**: 680–681.

Sonesson AK, Meuwissen THE (2000). Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genetics Selection Evolution* **32**: 231–248.

VanRaden PM (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**: 4414–4423.

Voorrips RE, Maliepaard CA (2012). The simulation of meiosis in diploid and tetraploid organisms using various genetic models. *BMC Bioinformatics* **13**: 248.

Zingaretti ML, Monfort A, Pérez-Enciso M (2019). pSBVB: A versatile simulation tool to evaluate genomic selection in polyploid species. *G3 (Bethesda)* **9**: 327–334.

Figures

Figure 1 Outline of the SeqBreed pipeline.

Inputs are shown in red squares, dashed border rectangles represent optional input, internal data are in blue rounded squares, main operations are indicated in blue, and outputs are in green circles; G and y refer to genotypes and phenotypes, respectively. The program starts with an optional gene dropping step following an input pedigree. No selection is performed at this stage. The bottom loop represents selection, where new offspring are generated based on the genotypes of selected parents. A list of SNPs in the genotyping array must be determined when using GBLUP and BLUP. A new cycle starts when these new offspring are added to the existing population. Plots can be performed at several stages.

Figure 2 Example plots produced by SeqBreed.

(a) Contribution of each QTN to total variance.

Top, individual QTN variances as a function of minimum allele frequency (MAF); middle, histogram of QTN variances; bottom, cumulative variance when QTN are sorted by MAF. In blue, additive variances; in red, dominance variances. The figure shows a fully additive phenotype, thus dominance variances are zero.

(b) Principal component analysis plot; individuals of different generations are in different colors.

(c) Genome wide association study showing false discovery rate values (-log10 scale).

SNPs on different chromosome are represented in alternate colors.

Paper III

EVALUATING GENOMIC PREDICTION STRATEGIES IN CROSSBRED *INDICINE* X *TAURINE* CATTLE PROGRAMS

Ramírez-Ayala, *et al.*

Manuscript in preparation

Summary

Although genomic selection has been mainly applied to dairy cattle, it also presents potentially important advantages in beef cattle, including crossbred breeding programs in the tropics. Here, we compare three crossbred *B. indicus* x *B. taurus* breeding schemes: F1, Grading up and Rotational crosses. We simulated, using real SNP data from *B. indicus* and *B. taurus*, three phenotypic traits of utmost importance in terms of productivity, mainly in tropical production systems based on beef cattle raising: shear force, growth and tolerance. We compared predictive accuracy between three 50k chips that differed in how SNPs were chosen: (i) randomly, (ii) with a maximun difference in allele frequency between breeds, and (iii) with a maximum difference in allele frequency between breeds provided allele frequency was larger than 0.09 in *B. taurus*. Our results suggest that the Rotational crossing system is optimum in terms of predictive accuracy, and that selecting markers based on allele frequency differences between breeds does not pay off and is even detrimental.

Introduction

Worldwide population has been growing rapidly in the last decades, with the corresponding increase in food demand. Beef production would need to increase from 60 million to 130 million tons by 2050 to feed a growing world population, and 70% of this production increase is expected from beef industries located in subtropical and tropical regions of the world (Cooke *et al.*, 2020). The improvement in environmental efficiency of beef production systems seems to be, at least for the foreseeable future, part of the solution for the issue of global food security. *B. indicus*-influenced cattle predominate in these regions, but are often managed using practices developed for *B. taurus* breeds reared in temperate climates.

B. taurus and *B. indicus* are different subspecies, and diverge in social and biological functions due to selection pressure caused by complex evolutionary and domestication processes (De Faria, M. *et al.*, 2019; Cooke, R.F. *et al*., 2020). Zebu breeds are better adapted to subtropical conditions than European breeds. However, their response capacity for meat production is low. On the other hand, European breeds have a high response capacity for meat production, but tropical conditions do not allow them to express their genetic potential. Due to these limitations of the zebu and European breeds, the most appropriate alternative to increase meat production in these environments is to use cattle that have the optimal genetic

composition, resulting from crossbreeding between these breeds. Crossing has some desirable consequences, especially in commercial livestock production; such as obtaining heterosis or hybrid vigor and the combination of two or more desirable characteristics (complementarity) in the commercial specimen. Consequently, systematic crossbreeding methods improve meat production. It is important to consider that the most important traits in commercial meat production are: adaptability to the environment, fertility, calving ease, maternal ability, weight gain efficiency rate, carcass merit, and longevity. For the breeding farm the most important thing is reproduction, while for the fattening farmers it is the efficiency of feed conversion and the quality of the carcass. The importance of crossbreeding lies in hybrid vigor or heterosis, which is the difference in phenotype between the mean of crossbreds and their purebred parents. Heterosis is greater the less genetically related the parents are. The maximum heterozygosis is achieved in the first crossing; therefore, the maximum hybrid vigor is obtained in the F1. Generally, the maternal line is characterized by environmental adaptation, maternal ability and low maintenance requirements, while the paternal line will transmit growth and meat quality to the beef (Bunning et al., 2019).

Genomic selection requires of large datasets to be effective. For that reason, it has been mainly implemented in 'simple' scenarios such as dairy cattle, where one single breed is used worldwide. In many other circumstances, though, evaluation is required across breeds or in crossbred populations. So far, however, prediction accuracy has not been large. Several reasons may explain this: First, GxE interaction, as a result of different environments, the genetic basis may also change. Second, even in the absence of proper GxE, different allele frequencies can make it that the variance explained by each locus may differ largely between breeds. The importance of each factor needs to be further evaluated.

The objective of this work is to compare several crossbred schemes and alternative SNP selection criteria in order to maximize genomic prediction (GP) performance. In particular, we will investigate whether choosing SNPs among those with similar frequency in each breed can improve predictions.

Materials and Methods

Genotype data

We obtained high density genotyping data from the WIDDE public database (http://widde.toulouse.inra.fr). Genotypes from Angus (N=42), Brangus (N=12), Brahman (N=46), Hereford (N=35), Nelore (N=31), Red Angus (N=10), Senepol (N=12) and Santa Gertrudis (N=32) were downloaded because they are the most widely used breeds in subtropical climate intensive production systems. Only autosomal SNPs with a maximum missing rate of 25% and a minimum allele frequency of 1% were retained. Individuals with over 5% missing genotypes were also removed. Missing SNPs were imputed with BEAGLE 4.1 (Browning & Browning, 2016). Principal Component Analysis (PCA) was used to cluster individuals and remove outliers (Figure S1.A). Eight outlier Hereford samples were so removed. Brahman and Nelore individuals were clustered in a generic *B. indicus* population; similarly, red and standard Angus animals were merged, as were Brangus and Senepol, abbreviated as BSP. Plink v1.9 software (Purcell *et al.*, 2007) was used to calculate allele frequencies for each population and genetic differentiation (Fst) indices between them. After SNP and sample filtering, 721,136 SNPs from 220 samples were used (Table 1).

Breed	Code	No. of samples	New Code	No. of samples after SNP quality control
Angus	ANG	42	ANG	52
Red Angus	RGU	10		
Brangus	BRG	12	BSP	24
Senepol	SEN	12		
Hereford	HFD	35	HFD	27
Brahman	BRM	46	IND	77
Nelore	NEL	31		
Santa Gertrudis	**SGT**	32	SGT	32

Table 1. Summary of analysed samples

Genetic architectures

Beef breeding in temperate climates mainly targets growth rate (O'Neill *et al.*, 2010) and meat quality (Schutt *et al.*, 2009). For tropical climates, parasite and heat resistance are also relevant since they impose the main limits to animals' performance. To represent these related characteristics, we aimed at simulating body weight gain, shear force (meat quality) and generic tolerance to heat and parasites. To generate 'realistic' genetic architectures, we downloaded QTL regions for each of these three sets of phenotypes from the QTL database (Table S1) and we selected SNPs within these QTL regions from the HD bovine array as putative causal SNPs for each of the phenotypes. Among all potential candidate SNPs, 200 SNPs were selected as causal for each phenotype. Causal SNPs were sampled among those of the 10% highest average Fst within regions smaller than 1MB around the QTL positions.

Genetic effects were generated from a gamma distribution Γ(shape = 0.2 and scale = 5) (Caballero *et al.*, 2015), multiplied by the sign of the difference between the mean of the allele frequencies of the two pure bred *taurine* breeds (ANG and HFD) and the allele frequency of the zebu population (IND). In agreement with the literature (Kim *et al.*, 2003; Machado *et al.*, 2010; Elzo *et al.*, 2012; Boonkum and Duangjinda, 2015), complete dominance was assumed for tolerance and additivity in the two remaining phenotypes. Heritability values (h^2) were taken from the literature (Hewetson, 1972; Watterson, 1975; Burrow, 2001, 2012; Burrow *et al.*, 2001; Henshall, 2004; Alencar *et al.*, 2005; Cucco *et al.*, 2010; Su *et al.*, 2010; Berry and Crowley, 2013), and were shear force h2 = 0.3, growth h2 = 0.24 and tolerance h2 = 0.4.

Crossbred programs and genotyping strategies evaluated

We compared three of the most widely used crosses in tropical climates: F1, Grading up and Rotational crosses. The terminal or F1 system consists in crossing pure zebu with *taurine* animals, which maximizes individual heterosis in the cross, and combines the desirable effects of the two breeds for direct and maternal genetic effects. The Grading Up system is a backcross that consists of the replacement of an initial F1 population through the systematic crossing the female offspring with taurine bulls. The final objective of this type of cross is the substitution of animals with low productivity by individuals with better performance while retaining resistant animals. Finally, the two-breed Rotational cross or criss-cross is a relatively

simple and popular form of crossbreeding. In this system, two breeds are crossed and the resulting female offspring are kept as replacements and crossed back to one of the breeds. In following generations, females are bred to the alternate breed of their sire. This system would require a minimum of two breeding pastures (if only natural service is used), one for each breed of sire, and cows need to be identified by breed of sire. Over several generations, 67% of the maximum amount of heterosis is realized. Additionally, there are a large number of heifers from which replacements may be selected.

In each of the three schemes, 400 animals were simulated. One generation for the terminal cross, four for the other two schemes, distributed with 100 offspring each. Base population individual genotypes (ANG=52 and IND=77) were downloaded and processed as described. We assessed genomic prediction (GP) performance by cross-validation using GBLUP (VanRaden, 2008). In the F1 cross, phenotypes from 200 animals were removed and correlation between predicted and actual phenotypes for each of the traits were compared. We will refer to this correlation as 'predictive ability' (PA). In either the Grading or Rotational crosses, PA for phenotypes from the last two generations (N=200) was computed. Three SNP chips were compared:

1. Chip 1: 50 k SNPs randomly chosen, distributed evenly across the genome.
2. Chip 2: 50 k SNPs randomly chosen among those with a maximum allele difference 0.09 between Angus and *B. indicus*.
3. Chip 3: 50 k SNPs randomly chosen among those with a maximum allele difference 0.09 between Angus and *B. indicus* frequencies and maximum allele frequency (MAF) of the Angus population of $0.2 < MAF < 0.8$.

SNPs were selected from those available at the HD chip.

SeqBreed was used to implement the simulation and GP described above. SeqBreed (https://github.com/miguelperezenciso/SeqBreed) is inspired by previous pSBVB fortran software, but the whole code has been rewritten in python3 and many new options have been added. It is more versatile than pSBVB and incorporates options such as Genome Wide Association Studies (GWAS) or Principal Component Analysis (PCA). Importantly, it allows

automatic implementation of standard genomic selection procedures, such as GBLUP. The pipelines and data used are in github (https://github.com/miguelperezenciso/SeqBreed).

Results

Principal components analysis of populations

The genetic structure across and within breeds was assessed using PCA. We conducted two analyses, first considering all the breeds initially selected and then the "retained" populations only. For all the breeds analyzed (Figure S1.A), PC1 accounts for 29% of the total variation. It separates zebu population (Brahman and Nelore) from the *taurine* breed (Angus, Hereford and Red Angus). Also, we do find the hybrid breeds are closer to *taurine* than to *indicine* breeds. This is because they have a higher percentage of European blood. It is also important to mention that Angus (ANG) and Red Angus (RGU) breeds form a single group, as well as the Brahman (BRM) and Nelore (Nel) zebus. PC2, which accounts for 5% of the total variation, separates *indicine* animals as well as the Hereford breed, from the other animals. The second PCA implemented for the "retained" populations only, reveals a clear separation between the two subspecies, *B. taurus* and *B. indicus* (Figure S1.B).

Predictive accuracy

Figure 1 shows PA across chips and breeding schemes for each phenotype. Overall, Chip 1 (50k random SNPs) and Rotational cross were the best strategies, but there were differences depending on the specific trait. For tolerance, PA was markedly lower and there were not that large differences among genotyping or breeding strategies.

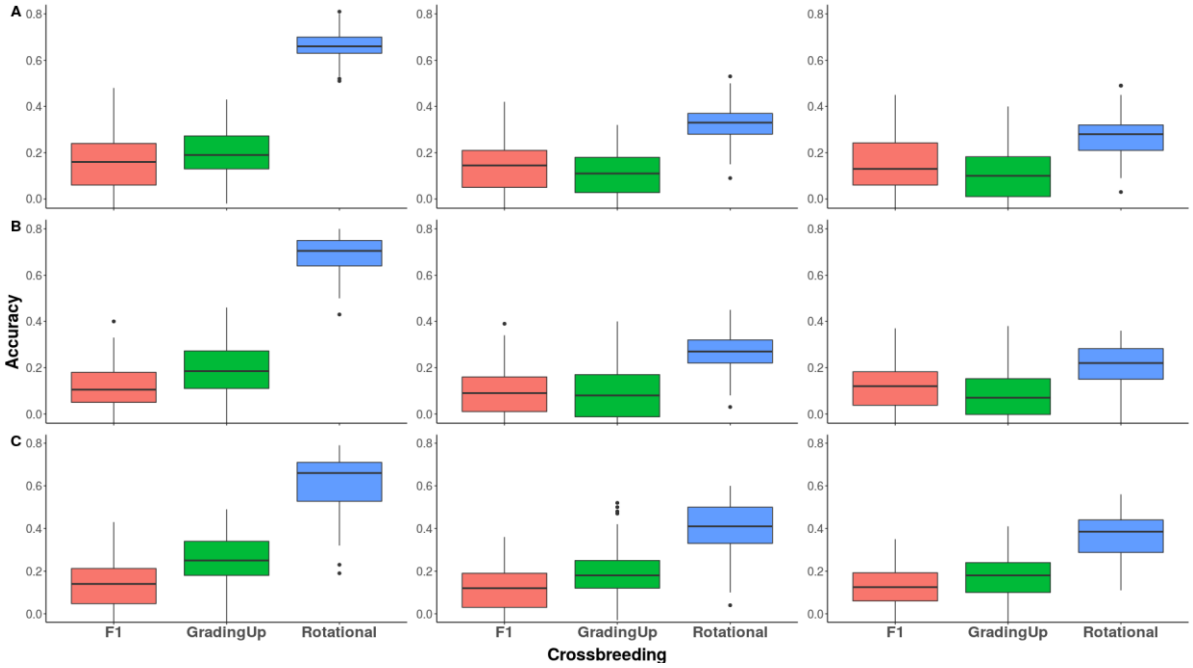

Figure 1. Correlations according to the crossbreeding scheme. **(A).** Shear, **(B).** Weight**, (C).** Tolerance vs. all chips (each column is a chip). Red boxes indicate F1 or terminal crossing scheme, green ones Grading Up and blue ones Rotational, for all cases.

The global effect of the chips on PA can be seen in Figure 2. Chip 1 (50k random) showed the highest PA (0.26), followed by Chip 2 (0.18) and Chip 3 (0.17). We clearly observe that choosing SNPs at random is the best strategy and that fixing a restriction on Fst or allele frequency does not help. As for the breeding scheme, the advantage of the Rotational cross is again clear on average (Figure 3). The Rotational scheme showed a higher correlation (0.38) with respect to the other two systems, F1 (0.12) and Grading Up (0.16).

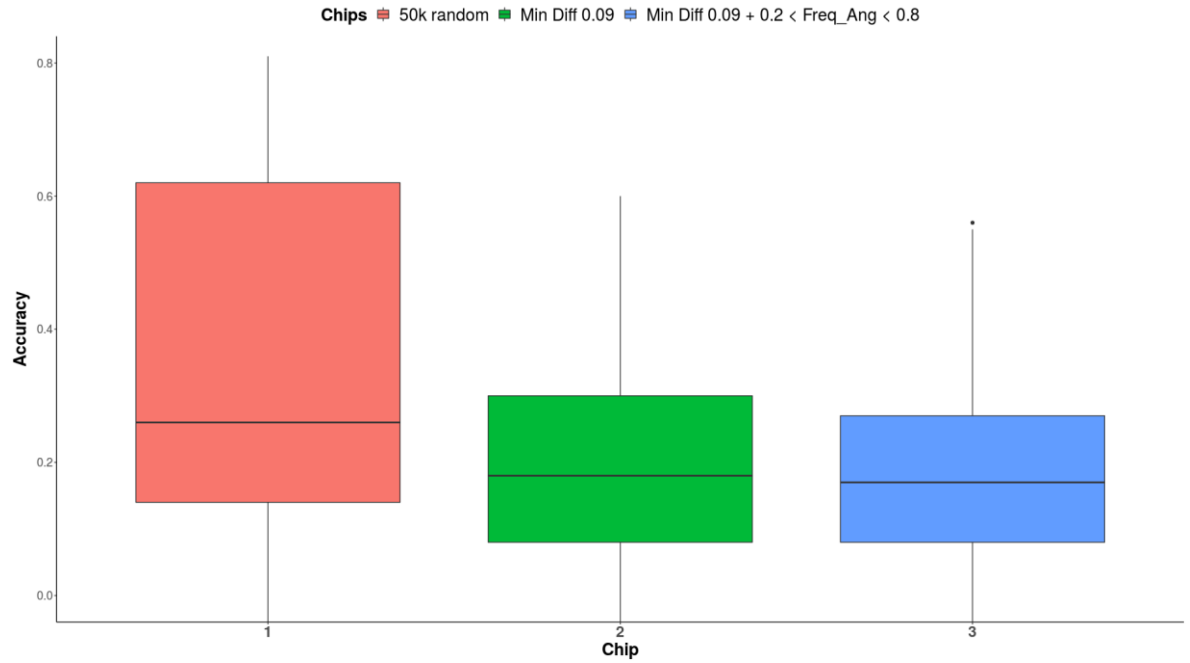

Figure 2. Global effect of the chips on the phenotypes.

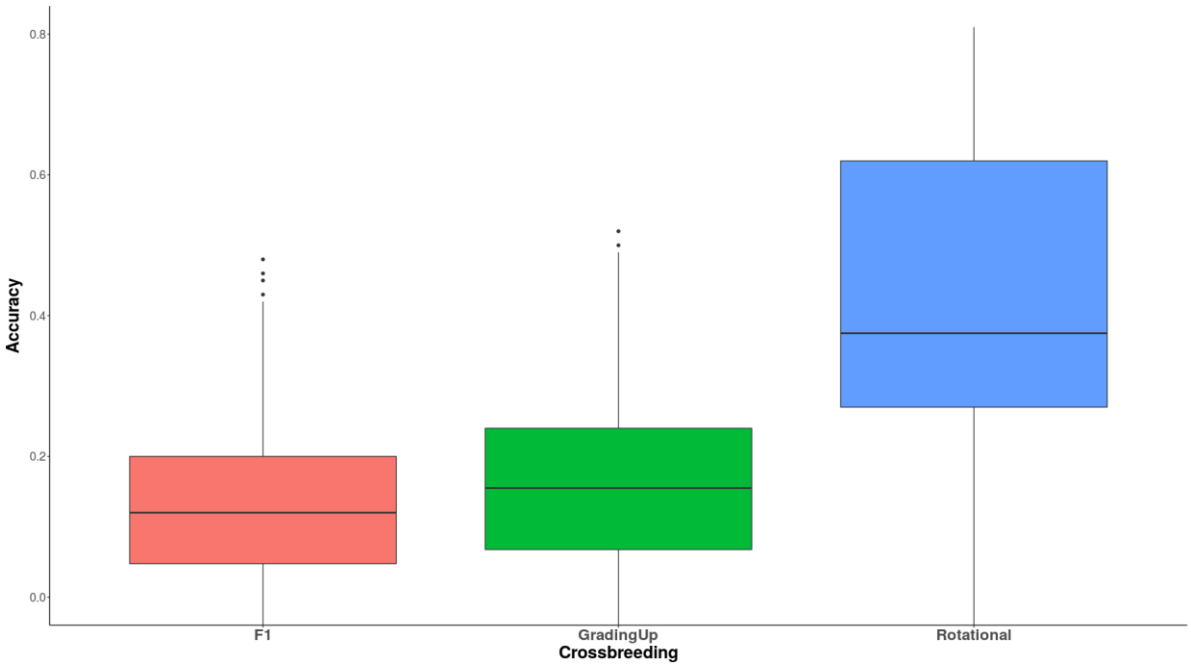

Figure 3. Global effect of the crossbreeding on the phenotypes.

Discussion

Principal component analysis (PCA) of the genotype data (721,136 SNPs) for 220 cattle of 8 breeds allows a clear separation between the breeds based on their *indicine* or *taurine* lineages. These observations are consistent with documented cattle history (The Bovine HapMap Consortium, 2009; Porto-neto *et al.*, 2014; Kasarapu *et al.*, 2017).

Using SNP microarrays could be advantageous by increasing the genomic prediction accuracy of traits with economic interest in beef cattle production systems raised in tropical ecosystems. For this work, subsets of 50k markers were chosen at random, representing a selection from a commercially available high-density genotyping platform. By comparing different SNP selection criteria (see Materials and Methods), Chip 1, corresponding to the selection of SNPs at random from a high-density (700k) chip, was the best option on all proposed traits. Pérez-Enciso *et al.*, (2017), comparing the effectiveness of different genotyping strategies, mentioned that they found a modest advantage in the accuracy of the prediction between the use of sequences over commercial or random SNP matrices using GBLUP.

Crossbreeding, considering either terminal or Rotational crossing, synthetic breed creation or breed replacement, is often promoted as an efficient strategy to increase farmers' income through the improvement of productivity of livestock. Rotational crossing is based on the use of crossbred dams that are alternatively mated to different breeds (usually from two to four), with the genetic composition of crossbred dams varying over generations. Similar to terminal crossing, it requires continuous supply of purebred genetic material, but only on the male side, sparing significant cost for breeders, especially in cattle, when a regular source of semen or low-cost males is available. Terminal and Rotational crossing aim to optimize the heterozygosity of the product, and therefore the heterosis effect (although heterosis is smaller under Rotational crossing). Those two strategies require the management of two (or more) parental lines, with a market chain to provide farmers either purebred reproducers or semen (Leroy *et al.*, 2015). Grading up (or top-crossing) means using the same sire breed each generation, in order to increase the proportion of a certain pure breed within a herd, mainly taurine breeds. Upgrading beyond 75 % temperate blood may, however, lead to problems if the climatic conditions are severe or the level of management does not develop in parallel

with the genetic potential of the stock (Rendel, 1974). Our results suggest that the Rotational scheme results in the best overall accuracy compared to the other two crossing systems (F1 and Grading up). Results from Rotational cross-breeding have shown a marked improvement in animal productivity. This crossing scheme is used or widely advocated in different parts of the tropics as a strategy to maintain high levels of heterozygosity and at the same time achieve specific proportions of the domestic and exotic strains (Galukande *et al.,* 2013).

Conclusion

Genomic selection (GS) has resulted in rapid rates of genetic gains. This success has been supported by well-established conventional genetic evaluation systems. In the case of crossbreeding schemes for tropical climates, our simulations suggest that the Rotational crossing system is optimum in terms of predictive accuracy, and that selecting markers based on allele frequency differences between breeds does not pay off and is even detrimental.

Acknowledgements

MPE is funded by MINECO grant AGL2016-78709-R and from the EU through the BFU2016-77236-P (MINECO/AEI/FEDER, EU) and the "Centro de Excelencia Severo Ochoa 2016-2019" award SEV-2015-0533. LCRA is funded by "Don Carlos Antonio López" Graduate program (BECAL) from Paraguay.

Authors' contributions

LCRA and MPE designed the study, analyzed the data, interpreted the results and wrote the manuscript.

References

Alencar MM, Fraga AB, da Silva AM (2005). Adaptação de genótipos a ambientes tropicais: resistência à mosca-dos-chifres (haematobia irritans, linnaeus) e ao carrapato (boophilus microplus, canestrini) em diferentes genótipos bovinos. *Agrociencia* **9**: 579–585.

Berry DP, Crowley JJ (2013). Cell Biology Symposium: Genetics of feed efficiency in dairy and beef cattle. *Journal of* Animal Science **91**: 1594–1613.

Boonkum W, Duangjinda M (2015). Estimation of genetic parameters for heat stress, including dominance gene effects, on milk yield in Thai Holstein dairy cattle. *Animal Science Journal* **86**: 245–250.

Browning BL, Browning SR (2016). Genotype Imputation with Millions of Reference Samples. *American Journal of Human Genetics* **98**: 116–126.

Bunning H, Wall E, Chagunda MGG, Banos G, Simm G (2019). Heterosis in cattle crossbreeding schemes in tropical regions: Meta-analysis of effects of breed combination, trait type, and climate on level of heterosis. *Journal of Animal Science* **97**: 29–34.

Burrow HM (2001). Variances and covariances between productive and adaptive traits and temperament in a composite breed of tropical beef cattle. *Livestock Production Science* **70**: 213–233.

Burrow HM (2012). Importance of adaptation and genotype × environment interactions in tropical beef breeding systems. *Animal* **6**: 729–740.

Burrow HM, Moore SS, Johnston DJ, Barendse W, Bindon BM (2001). Quantitative and molecular genetic influences on properties of beef: A review. *Australian Journal of Experimental Agriculture* **41**: 893–919.

Caballero A, Tenesa A, Keightley PD (2015). The nature of genetic variation for complex traits revealed by GWAS and regional heritability mapping analyses. *Genetics* **201**: 1601–1613.

Cooke RF, Daigle CL, Moriel P, Smith SB, Tedeschi LO, Vendramini JMB (2020). Cattle adapted to tropical and subtropical environments: social, nutritional, and carcass quality considerations. *Journal of Animal Science* **98**: skaa014.

Cucco DC, Ferraz JBS, Eler JP, Balieiro JCC, Mattos EC, Varona L (2010). Genetic parameters for postweaning traits in Braunvieh cattle. *Genetics and Molecular Research*

9: 545–553.

Elzo MA, Johnson DD, Wasdin JG, Driver JD (2012). Carcass and meat palatability breed differences and heterosis effects in an Angus-Brahman multibreed population. *Meat Science* **90**: 87–92.

Maciel ICF, Barbosa FA, Tomich TR, Pereira Ribeiro LG, Costa Alvarenga R, Lopes LS, *et al.* (2019). Could the breed composition improve performance and change the enteric methane emissions from beef cattle in a tropical intensive production system? *PLoS ONE* **14**: e0220247.

Galukande E, Mulindwa H, Wurzinger M, Roschinsky R, Mwai AO, Sölkner J (2013). Cross-breeding cattle for milk production in the tropics: achievements, challenges and opportunities. *Animal Genetic Resources/Ressources génétiques animales/Recursos genéticos animales* **52**: 111–125.

Henshall JM (2004). A genetic analysis of parasite resistance traits in a tropically adapted line of *Bos taurus*. *Australian Journal of Agricultural Research* **55**: 1109–1116.

Hewetson RW (1972). The inheritance of resistance by cattle to cattle tick. *Australian Veterinary Journal* **48**: 299–303.

Kasarapu P, Porto-Neto LR, Fortes MRS, Lehnert SA, Mudadu MA, Coutinho L, *et al.* (2017). The *Bos taurus*-*Bos indicus* balance in fertility and milk related genes. *PLoS ONE* **12**: e0181930.

Kim JJ, Farnir F, Savell J, Taylor JF (2003). Detection of quantitative trait loci for growth and beef carcass fatness traits in a cross between *Bos taurus* (Angus) and *Bos indicus* (Brahman) cattle. *Journal of Animal Science* **81**: 1933–1942.

Leroy G, Baumung R, Boettcher P, Scherf B, Hoffmann I (2015). Review: Sustainability of crossbreeding in developing countries; definitely not like crossing a meadow... *Animal* **10**: 262–273.

Machado MA, Azevedo AL, Teodoro RL, Pires MA, Peixoto MGCD, de Freitas C, *et al.* (2010). Genome wide scan for quantitative trait loci affecting tick resistance in cattle (*Bos taurus* × *Bos indicus*). *BMC Genomics* **11**: 280.

O'Neill CJ, Swain DL, Kadarmideen HN (2010). Evolutionary process of *Bos taurus* cattle in favourable versus unfavourable environments and its implications for genetic selection. *Evolutionary Applications* **3**: 422–433.

Pérez-Enciso M, Forneris N, de los Campos G, Legarra A (2017). Evaluating sequence-based genomic prediction with an efficient new simulator. *Genetics* **205**: 939–953.

Porto-Neto LR, Reverter A, Prayaga KC, Chan EKF, Johnston DJ, Hawken RJ, *et al.* (2014). The Genetic Architecture of Climatic Adaptation of Tropical Cattle. *PLoS ONE* **9**: e113284.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al.* (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**: 559–575.

Rendel J (1974). The role of breeding and genetics in animal production improvement in the developing countries. *Genetics* **78**: 563–575.

Schutt KM, Burrow HM, Thompson JM, Bindon BM (2009). Brahman and Brahman crossbred cattle grown on pasture and in feedlots in subtropical and temperate Australia. 1. Carcass quality. *Animal Production Science* **49**: 426–438.

Su G, Guldbrandtsen B, Gregersen VR, Lund MS (2010). Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. *Journal of Dairy Science* **93**: 1175–1183.

The Bovine HapMap Consortium (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**: 528–532.

VanRaden PM (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**: 4414–4423.

Watterson GA (1975). On the Number of Segregating Sites in Genetical Models without Recombination. *Theoretical Population Biology* **7**: 256–276.

SUPPLEMENTARY MATERIAL

Figure S1: Population structure**. (A)** Principal component analysis using all samples. Individuals are grouped in *B. taurus*, *B. indicus* and Hybrid. Black: Angus, red: Brangus, blue: Brahman, cyan: Hereford, green: Nelore, orange: Red Angus, magenta: Senepol, purple: Santa Gertrudis; **(B)** Principal component analysis with Angus and *B. indicus* only. Red: Angus (Angus + Red Angus), blue: *B. indicus* (Brahman + Nelore).

Table S1. QTL regions for each of these three sets of phenotypes from the QTL database.

Chapter 4

General Discussion

General Discussion

This thesis aimed to study the application of genomics to beef breeding in the tropics. We have approached this topic from several viewpoints: ranging from the use of sequence data for dissecting the genetic basis of adaptation to investigating genomic selection in *B. indicus* x *B. taurus* crosses. We have also developed appropriate simulation tools.

4.1. The genomic basis of adaptation to tropical climates

Resistance to heat and adaptation to tropical climates varies from breed to breed, clearly showing there is a genetic basis for these complex phenotypes. In cattle, the most dramatic differences are between *B. indicus* and *B. taurus* subspecies. It is also evident that adaptation depends on a multitude of individual traits (e.g., resistance to parasites, thermoregulation, etc.), which makes it also more challenging to detect potential candidate genes. A further complication arises because of the lack of appropriate populations, since the origin of most cattle populations living in the tropics are uncertain.

Here, we analyzed a unique population that has been maintained isolated for the last century and origined from *B. taurus*, Chacuba. Therefore, it is the most suitable candidate to understand adaptation to the tropics in the taurine cattle. We found that a small percentage of intercrossing with zebu has occurred, but its adaptation was not solely explained by this *B. indicus* introgression.

Several genes are involved and play an important role in adaptation. To identify them, there are several methods available that scan the genome, one of which is the Extended haplotype homozygosity (EHH) method that identifies long-range haplotypes (Sabeti *et al.*, 2002) and have been used in chapter 3, paper I. We identified 887 genes, from which we selected those related to adaptation processes, such as thermogenesis, hair development or food efficiency (Table 4.1).

Genomic position	Ensembl Gene ID	Gene name	Condition
8:69500592-69604707	ENSBTAG00000020665	*GFRA2*	Residual Feed Intake
10:61005653-61068184	ENSBTAG00000004013	*FGF7*	Hair Development
10:92624410-92632934	ENSBTAG00000001605	*DIO2*	Thermogenesis
17:17467450-17473822	ENSBTAG00000004647	*UCP1*	Thermogenesis
23:27808496-27812710	ENSBTAG00000021721	*CDSN*	Hair Development
27:14223449-14288333	ENSBTAG00000004344	*ACSL1*	Thermogenesis
28:41817915-41875990	ENSBTAG00000000231	*BMPR1A*	Hair Development

Table 4.1. Genes selected as potential candidates and involved in adaptation processes in cattle.

The *UCP1* gene, belonging to the group of uncoupling proteins (UCPs) is located in the mitochondrial internal membrane and considered crucial regulator of energetic homeostasis. Those proteins can increase the permeability of the internal membrane of the mitochondrion, thus increasing heat dissipation and decreasing the production of ATP (Jin *et al.*, 2018). In cattle, Abd Eldaim *et al.* (2016) argue that the expression of this gene may lead to a decrease in energy (fattening) efficiency, but also that additional studies are needed to determine the Ucp1-mediated thermogenesis. Meanwhile in dairy cows, Zhou *et al.* (2017) reported genetic variation in four regions of bovine *UCP1* gene in crosses of Holstein-Friesian × Jersey dairy cows. This variation mainly affected the daily milk yield and also had a minor effect on the fat and protein percentage. Another interesting gene is *DIO2*, related to thyriod hormone regulation, which play a critical role in thermogenesis and metabolism (Silvestri *et al.*, 2005). This was reported by Naidu (2016), who studied gene expression in two cattle breeds, Tharparkar and Karan Fries (Tharparkar × Holstein Friesian). They reported that lower mRNA expression of the *DIO2* gene, together with lower levels of thyroid hormones in Tharparkar cattle than in Karan Friesian cattle, is associated to better adaptability of Tharparkar cattle than the Karan Friesian cattle in tropical climatic conditions while maintaining their productivity. The same gene was reported in an independent study by Howard *et al.* (2014) who conducted a genome association study in beef cattle to better understand the genetic basis of body temperature regulation.

Another gene of interest is PRLR. In the Senepol breed, a taurine adapted to tropical conditions, the slick hair locus has been found, a gene with a dominant pattern of inheritance that controls the length of hair in cattle. Olson *et al.* (2003) reported that animals carrying this gene maintain lower rectal temperature, with differences of 0.18 to 0.4°C with animals carrying 'normal' hair. This same pattern was observed by the same authors in the composite Venezuelan breed Carora, a cross between the Brown Swiss taurine breed and a Venezuelan Creole breed, which indicates the presence of the same gene in Creole breeds of Spanish origin in Latin America. Dikmen *et al.* (2014) studying Holstein cows with and without the slick gene, indicate that the former have a superior thermoregulatory capacity and experience a less drastic depression in production during the summer. Mariasegaram *et al.* (2007) have mapped this gene on chromosome 20 and Littlejohn *et al.* (2014) propose mutations in the prolactin receptor (*PRLR*) on chromosome 20 and the prolactin gene (*PRL*) on chromosome 23, as being responsible for the condition of straight hair.

We were also interested in the *ACSL1* gene, one of the five isoforms and member of the long-chain Acyl-CoA synthetase (LCA) family. Widmann *et al.* (2011) reported its role as a candidate gene for fatty acid composition in bovine skeletal muscle. The same gene was identified by Zhao *et al.* (2020) when studying the genome-wide autozygosity pattern and inbreeding level in a commercial cattle populations. The authors found that *ACSL1* was involved in promoting growth and development of the animal's body.

In relation to feed efficiency, we found the *GFRA2* gene. This gene was reported to influence basal metabolic rates, suggesting a mechanism by which genetic variation may contribute to residual feed intake. In this context, Higgins *et al.* (2018) carried a genomics associations study in Irish cattle and identified a new eQTL for residual feed intake in the *GFRA2* gene. Increased *GRFA2* expression is correlated with increased feed efficiency, suggesting that this gene is involved in a mechanism or pathway involved in the reduction of metabolic rates.

4.2. SeqBreed as a general tool to investigate genomic experiments

Computer simulation has a long history in genetics, both in the areas of population genetics and in quantitative genetics. For decades, it has been a fundamental tool to investigate the effects of selection of complex traits and to efficiently design breeding strategies, in plants as

well as in animals. The appearance of genomic selection has exacerbated the need of having appropriate and specific simulation tools, because genomic selection is an expensive strategy and there are numerous alternatives to optimize genotyping strategies.

Here we developed SeqBreed (https://github.com/miguelperezenciso/SeqBreed) to fill out some of the gaps on extant software. Although several other programs have been developed for similar purposes as SeqBreed, including our own pSBVB (Zingaretti *et al.*, 2019), AlphaSim (Faux *et al.*, 2016) and its successor AlphaSimR (https://alphagenes.roslin.ed.ac.uk/wp/software-2/alphasimr/), PedigreeSim (Voorrips and Maliepaard, 2012), simuPOP (Peng & Kimmel, 2005), and QMSim(Sargolzaei & Schenkel, 2009), SeqBreed offers a unique combination of features for simulation of genomic prediction of complex traits. It includes built-in implementation of several genomic prediction methods, the possibility of simulating polyploid genomes, and several options to specify QTN and SNP arrays. Given that is developed in python, it is easier to use than software developed in compiled languages, although at the cost of efficiency. It is especially suited for teaching purposes.

Although SeqBreed has many advantages in its use, additional features have yet to be developed to generalize available genetic architectures (e.g. imprinting, epistasis), integration with automatic learning tools (scikit, keras) for genetic evaluation, development of an educational tool with an html-based interface, and improving output and plotting features.

4.3. Design of genomic selection experiments in the tropics

A final contribution of this thesis was to investigate, via simulation, optimum strategies for genomic selection in tropical climates. We focused on the different breeding schemes that are being currently used in Paraguay so that our results could be implemented with minimum distortion in Paraguayan beef breeding schemes. We considered different genotyping strategies appropriate for multiple breed evaluation, but none outperformed the simply random sampling. Overall, we found that the rotational crossing system is optimum in terms of predictive accuracy.

There is abundant literature describing the different crossbreeding schemes that were tested in this thesis and that provide an important basis for an optimized livestock production in the tropics. The terminal cross or F1 aims to generate offspring that have the above-average performance of the two parent breeds together.

F1 crosses are widely used. For instance, crosses between typical taurine (Angus and Hereford) and zebu (Brahman) breeds were evaluated for maternal and reproductive traits in the southern United States but using other African breeds as well, such as Tuli (*B. taurus*) and Boran (*B. indicus*) breeds adapted to the harsh tropical environment. Muntean *et al.* (2018) reported greater production efficiency in F1 resulting from the British-Boran cross compared to the other types of crosses (British-Brahman and British-Tuli). More recently, Mendonça *et al.* (2019) reported similar results with a group of purebred (Angus, Hereford and Nelore) and crossbred (Angus-Hereford, Hereford-Angus, Angus-Nelore, Nelore-Angus, Caracu-Angus) beef cows raised in Southern Brazil. These studies confirm the effect of heterosis in the cross wich can result for example, in higher average of mature weight or higher rate of growth. However, the Hereford breed is an exception with similar maturity weight as its crosses. F1 cows produced a greater amount of milk and heavier calves, although the authors mention that they were less dependent on the distance between the breeds involved in the cross. Unfortunately, hybrid vigour is primarily a one-generation phenomenon. That extra vigor is not passed on to the next generation when the initial cross-bred animals are re-bred due to recombination.

Backcross has also been reported as one of the crossing schemes in tropical climates. This mechanism consists of a first cross between two purebred breeds, all male calves are sold for slaughter, while the female offspring are crossed with males of one of the parent breeds, and the offspring of these are sold for slaughter. This breeding system makes maximum use of heterosis for fertility and half of the possible heterosis for growth. When adaptation to a specific environment of a particular mother breed is required, the crossed F1 female is also successfully adapted.

Amen *et al.* (2007a) studied the effects of the backcross system in backcrossed *B. taurus* x *B. indicus* calves, first evaluating birth and weaning traits and then post-weaning, carcass, and

meat traits (Amen *et al.*, 2007b). In the first study, calves with a higher percentage of Brahman in the sire compared to the proportion in the dam were classified as heavier for weaning weight than for birth weight, although these differences were not significant. In the following study, while the authors identified an increase in body weight up to 18 months of age, when theoffpring possessed more Brahman influence on the sire than the dam, body composition and meat quality traits in the Brahman × *B. taurus* cross produce calves that do not seem to explain the performance differences, as it does for weight traits.

The term backcrossing is associated with grading-up. The variation in genetic composition with this type of crossing was described by Hill (1993) and many breeds were "created" as a result of this methodology. Koufariotis *et al.* (2018) identified that the Brahman breed has a mosaic genome, as a result of the process of backcrossing *B. taurus* females with *B. indicus* bulls. Swan & Graser (1988) reported that animals of the Simmental breed were first introduced to Australia in the 1970s, with the importation of semen from Europe and live animals from New Zealand. Subsequently, grading-up programs were initiated until calves were at least 0.9375 Simmental.

The rotational crossing system occurs when males of two or more breeds mate with crossed females. Over time, each breed will have contributed its strengths and weaknesses equally. The term 'criss-cross' is normally used when two breeds are mated (i.e. zebu x *taurine*) and the resulting female offspring are kept as replacements and mated again with one of the breeds. In subsequent generations, the females are raised with the opposite breed to their sire. Traits of reproductive interest were analyzed by Williams *et al.* (1990) in a study of two, three, and four breeds in rotating crosses, spanning four generations. The authors concluded that rotational breeding systems are more productive than the parental breeds. The three- and four-breed rotational systems had a slight advantage over the two-breed rotational system for weaning rate. The two-breed Hereford-Brahman rotation system and the three-breed Angus-Brahman-Hereford rotation system tended to be superior to the other two- and three-breed rotation systems for weaning rate. DeRouen *et al.* (1992) used the same breeds and rotational crossing systems and evaluated carcass-related traits, concluding that three- and four-breed rotational mating systems were superior to the two-breed rotation for hot carcass weight, retail yield, and muscle area longissimus, but were similar for fat thickness, marbling score, and

Warner-Bratzler for shear force. Rotational combinations exceeded purebred breeds in all carcass traits except marbling score.

Synthetic breeds were mostly developed because individuals of the pure breeds *B. taurus* and *B. indicus* do not perform satisfactorily in all traits of economic importance for tropical climates. Thrift *et al.* (2010) conducted a major review covering research results over approximately 22 years, including comparisons of pre-weaning and post-weaning traits and carcass traits of sired progeny of several breeds widely used in meat production (purebred *B. taurus*, *B. indicus*, and synthetic breeds). They highlighted that the performance before and after weaning is lower in the progeny sired by the Brahman breed compared to the progeny sired by subtropically adapted non-*B. indicus* breeds (Tuli, Romosinuano, Bonsmara, and Senepol). These breeds help to reduce dystocia, along with other *B. indicus* bull breeds, such as Gir and Sahiwal, but not Indu-Brazil, and seem to improve the merit of the carcasses, especially their sensitivity, compared to the Brahman breed. As for the *B. taurus* sire breeds, especially Angus and Hereford, they present superior carcass merit in terms of marbling score, quality grades and tenderness. In general, progeny derived from *B. indicus* (Brangus, Beefmaster and Santa Gertrudis) are consistently heavier at birth and weaning than progeny sired from *B. taurus* (Angus, Hereford and Red Poll).

Genomic selection aims to improve production by exploiting molecular genetic markers to design new breeding programmes. New marker-based models can be developed for genetic evaluation, as well as to improve selection routines, especially in species with long breeding cycles, late or sex-limited traits, or complex traits. Genomic selection requires certain initial conditions that are very well met, for example, in dairy cattle (large amount of accumulated information and reliable genetic testing of many bulls thanks to international genetic evaluation, high fertility of artificial insemination and adequate ratio between the price of a stallion and the cost of the SNP chip, among others). Above all it should be stressed that genomic selection does not eliminate any of the activities that are currently carried out in "classical" selection programmes. It is therefore important to continue to maintain BLUP genetic evaluation of animals, as the reference population must be redefined regularly. Similarly, genealogy recording should be maintained for similar reasons. The application of genomic selection became feasible with the availability of panels of thousands of SNPs that

could be genotyped at a reasonable cost and has revolutionized all livestock breeding programs. The real power of genomic selection is to estimate the value of breeding more accurately than could be done with pedigree data alone (Goddard *et al.*, 2010).

The application of genomic estimation leaded to significant changes in dairy farming, but has not been widely accepted in beef breeding. Although beef cattle also have a long generation interval, several obstacles arise for the application of genomic selection in beef cattle such as the variety of breeds and in general the small number of genotyped and phenotyped individuals per breed. On the other hand, the disadvantages of storing phenotype data of satisfactory value and size hinder traditional method of conventional evaluation and makes genomic selection effective by significantly improving genetic gain when increases the reliability of selection at an early age (Jonas & de Koning, 2015). In livestock breeding, index selection is usually market-specific, but its adoption is slower due to the phenotype characteristics of interest such as growth rate, carcass, reproduction and health that contribute to profitability (Montaldo *et al.*, 2012). In fact, the efficiency of genomic selection remains lower in beef cattle than in dairy cattle, possibly due to the heterogeneity of breeds, less advanced breeding structures and schedules, predominance of natural services, crosses in commercial herds, as well as effective population size (Johnston *et al.*, 2012; Van Eenennaam *et al.*, 2014). The lower reliability is due to the fact that the quality and quantity of the beef cattle population data is lower than that of dairy cattle. In addition, the target population and validation animals may be less closely related to the reference population for meat cattle than for dairy cattle. Combining data across countries and/or breeds would possibly increase the accuracy of the prediction by solving the problem of small reference populations (De Roos *et al.*, 2009).

The application of genomic selection, particularly on cattle, has grown in recent years in tropical environments. In this context, several works have been published in Brazil, a country that has a cattle population of over 230 million, making it the largest exporter of meat in the world (Vale *et al.*, 2019; Grigoletto *et al.*, 2020). Cardoso *et al.* (2015) investigated the usefulness of genomic prediction as a tool to select animals resistant to one of the main health problems facing livestock production in tropical climates, such as bovine ticks, which causes a decrease in yield, devaluation of the skin, increased production costs and transmission of

infectious diseases. They used two breeds that are well established in this type of environment, the Hereford taurine breed and the synthetic Braford. The results showed that the accuracy values were moderately high for genomic predictions of tick load in Braford cattle, indicating that genomic predictions could be used as a practical tool to improve genetic tick resistance and in the development of tick-resistant lines in this breed. While in the Hereford breed, those accuracies were found from low to moderate (from 0.29 to 0.36). According to the same authors, although the accuracies of genomic predictions can be reliably estimated, it is needed to expand the training population with more individuals with phenotypic and genotypic information, before using genomic selection for tick resistance in practice. Neves *et al.* (2014) presented the first results of the implementation of genomic selection in the Nelore breed for traits related to productivity and fertility. In their results they concluded that further improvements are needed to reduce the deflation of predictions, which would require regular updates of the training population to allow accurate prediction of the genetic merit of young animals. However, they stress the technical feasibility of applying genomic prediction. Similar results were found by Magalhães *et al.* (2019) studying traits related to meat quality in the same population. It is well known that the Nelore breed is not selected precisely for the quality of its meat, since this trait is mostly exploited in breeds from temperate climates. Improving this phenotype would be of great advantage for a breed that is well adapted to the harsh characteristics of the tropical climate. Therefore, the authors conclude that meat quality traits should be incorporated into the genetic evaluation scheme of the breeding program, and support the feasibility of applying genomic selection in Nelore cattle.

Another country with numerous studies on genomic selection in beef cattle is Australia, which, due to its large size, is home to around 24 million head of cattle, making it the third largest exporter of beef in the world (Suybeng *et al.*, 2019). In particular, research carried out on many commercial breeds, such as the study of Boerner *et al.* (2014), which presents the accuracies of the genomic breeding estimates (GEBV) that were calculated from genomic predictions. These accuracies were generally low (from 0.1 to 0.4) in comparison with published breed-specific estimated accuracies (Angus, Murray Grey, Shorthorn, Hereford, Brahman, Belmont Red, Santa Gertrudis, and Tropical Composite), but they were also consistent with those derived from other multi-breed populations. Bolormaa *et al.* (2013)

sought to evaluate the accuracy of genomic predictions for several traits, including feed efficiency, growth, carcass and meat quality traits in cattle. They concluded that, although the accuracies are low compared to those observed in dairy cattle, they show that genomic selection would still be beneficial for traits that are difficult to improve through conventional selection, such as tenderness and residual feed intake. Zhang *et al.* (2014) focused their research on studying the utility of genomic selection on traits related to breeding of females of breeds adapted to the tropics (Brahman and Tropical Composite). Although the accuracy of the predictions was generally low (from 0.2 to 0.4) due to the limited number of data from the experimental populations, they found a similar level of accuracy for some of the early measures of female reproduction in commercial cows, indicating that the potential for genomic selection may be limited by the number of animals with phenotypes.

4.4. Future research

Adaptation remains a complex issue and future studies should allow us to increase our understanding of this phenomenon and how it affects production, especially for cattle breeds that are used in tropical climates.

The main limitation we faced in our first study was that very few Chacuba animals were sequenced. A larger number of sequences from this population, as well as that of other widely exploited breeds in tropical conditions, could confirm our findings and extend our knowledge of how this genetic group can, to this day, survive and produce in that harsh tropical environment.

On the other hand, since reference populations are much larger in dairy cows and many progenies tested are used for genomic selection is more appropriate in dairy than in beef cattle. In addition, the target population and validation animals may be less closely related to the reference population in beef cattle than in dairy cattle. Despite these difficulties, genomic selection has been applied in beef cattle (e.g., Meuwissen *et al.*, 2016). As DNA information from different commercial breeds of cattle increases, predictions can improve and therefore optimize breeding programs. Finally, tools such as SeqBreed can help to optimize allocation of resources in tropical countries.

Chapter 5

Conclusions

Conclusions

1) Using whole-genome sequence data, we validate that the Chacuba genetic group is closely genetically related to the French and Canadian Charolais breeds, although with a limited zebu introgression.

2) We found signs of recent adaptation to tropical conditions between Chacuba and its founding Charolais French population, as well as genes and variants that could explain the phenotypic differences with the Charolais breed as well as its adaptation.

3) SeqBreed is a tool that allows the optimization of genomic selection as well as genome-wide association studies, incorporating some of the most widely used genomic selection methods in an easy and flexible way.

4) Our simulations of the application of genomic selection in beef breeds using real SNP data from *B. indicus* and *B. taurus* in tropical environments show that the rotational crossing system is optimal in terms of predictive accuracy, and that marker selection based on allele frequency differences between breeds does not lead to improve upon SNP random sampling.

Chapter 6

References

Asociación Rural del Paraguay (2010). *Manual Compendio de la Ganaderia Paraguaya*. Asunción, Paraguay.

Abd Eldaim MA, Hashimoto O, Ohtsuki H, Yamada T, Murakami M, Onda K, *et al.* (2016). Expression of uncoupling protein 1 in bovine muscle cells. *Journal of Animal Science* **94**: 5097–5104.

Acosta AC, Uffo O, Sanz A, Ronda R, Osta R, Rodellar C, *et al.* (2013). Genetic diversity and differentiation of five Cuban cattle breeds using 30 microsatellite loci. *Journal of Animal Breeding and Genetics* **130**: 79–86.

Aggarwal A, Upadhyay R (2013). Heat stress and animal productivity. Springer, New Delhi, India.

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* **93**: 743–752.

Albuquerque LG, Fernandes Júnior GA, Carvalheiro R (2017). Beef Cattle Genomic Selection in Tropical Environments. In: *Proceedings Association for the Advancement of Animal Breeding and Genetics Conference* **22**: 255-263.

Amen TS, Herring AD, Sanders JO, Gill CA (2007a). Evaluation of reciprocal differences in *Bos indicus* × *Bos taurus* backcross calves produced through embryo transfer: I. Birth and weaning traits. *Journal of Animal Science* **85**: 365–372.

Amen TS, Herring AD, Sanders JO, Gill CA (2007b). Evaluation of reciprocal differences in *Bos indicus* × *Bos taurus* backcross calves produced through embryo transfer: II. Postweaning, carcass, and meat traits. *Journal of Animal Science* **85**: 373–379.

Ames D (1980). Thermal Environment Affects Production Efficiency of Livestock. *BioScience* **30**: 457–460.

Berihulay H, Abied A, He X, Jiang L, Ma Y (2019). Adaptation mechanisms of small ruminants to environmental heat stress. *Animals (Basel)* **9**: 75.

Bertram J, Carric M, Holroyd D, Lake M, Lehman W, Taylor K, *et al.* (2002). *Breeding for Profit*. Department of Primary Industries. 2ª Ed. Queensland.

Boerner V, Johnston DJ, Tier B (2014). Accuracies of genomically estimated breeding values from pure-breed and across-breed predictions in Australian beef cattle. *Genetics Selection Evolution* **46**: 61.

Bolormaa S, Pryce JE, Kemper K, Savin K, Hayes BJ, Barendse W, *et al.* (2013). Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. *Journal of Animal Science* **91**: 3088–3104.

Bunning H, Wall E, Chagunda MGG, Banos G, Simm G (2019). Heterosis in cattle crossbreeding schemes in tropical regions: Meta-analysis of effects of breed combination, trait type, and climate on level of heterosis. *Journal of Animal Science* **97**: 29–34.

Burrow HM (2001). Variances and covariances between productive and adaptive traits and temperament in a composite breed of tropical beef cattle. *Livestock Production Science* **70**: 213–233.

Burrow HM, Henshall JM (2014). Relationships between adaptive and productive traits in cattle, goats and sheep in tropical environments. In: *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production*, Vancouver, Canada, 17–22 August 2014.

Cardoso FF, Gomes CCG, Sollero BP, Oliveira MM, Roso VM, Piccoli ML, *et al.* (2015). Genomic prediction for tick resistance in Braford and Hereford cattle. *Journal of Animal Science* **93**: 2693–2705.

Carvalheiro R, Costilla R, Neves HHR, Albuquerque LG, Moore S, Hayes BJ (2019). Unraveling genetic sensitivity of beef cattle to environmental variation under tropical conditions. *Genetics Selection Evolution* **51**: 29.

Christensen OF, Lund MS (2010). Genomic relationship matrix when some animals are not genotyped. *Genetics Selection Evolution* **42**: 2.

Coop RL, Kyriazakis I (1999). Nutrition and Parasite Interaction. *Veterinary Parasitology* **84**: 187–204.

Daetwyler HD, Calus MPL, Pong-Wong R, de los Campos G, Hickey JM (2013). Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. *Genetics* **193**: 347–365.

DeRouen SM, Franke DE, Bidner TD, Blouin DC (1992). Two-, three-, and four-breed rotational crossbreeding of beef cattle: carcass traits. *Journal of Animal Science* **70**: 3665–3676.

Dikmen S, Hansen PJ (2009). Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment? *Journal of Dairy Science* **92**: 109–116.

Dikmen S, Khan FA, Huson HJ, Sonstegard TS, Moss JI, Dahl GE, *et al.* (2014). The SLICK hair locus derived from Senepol cattle confers thermotolerance to intensively managed lactating Holstein cows. *Journal of Dairy Science* **97**: 5508–5520.

Faux A-M, Gorjanc G, Gaynor RC, Battagin M, Edwards SM, Wilson DL, *et al.* (2016). AlphaSim: Software for Breeding Program Simulation. *The Plant Genome* **9**: 10.3835/plantgenome2016.02.0013..

Fernández M (2003). *Informe de país sobre la situación nacional de los recursos zoogenéticos en animales de granja*. La Habana, Cuba.

Forbes AB, Huckle CA, Gibb MJ, Rook AJ, Nuthall R (2000). Evaluation of the effects of nematode parasitism on grazing behaviour, herbage intake and growth in young grazing cattle. *Veterinary Parasitology* **90**: 111–118.

Frisch JE (1999). Towards a permanent solution for controlling cattle ticks. *International Journal for Parasitology* **29**: 57–71.

Galukande E, Mulindwa H, Wurzinger M, Roschinsky R, Mwai AO, Sölkner J (2013). Cross-breeding cattle for milk production in the tropics: achievements, challenges and opportunities. *Animal Genetic Resources/Ressources génétiques animales/Recursos genéticos animales* **52**: 111–125.

Georges M, Charlier C, Hayes B (2019). Harnessing genomic information for livestock improvement. *Nature Reviews Genetics* **20**: 135–156.

Ghebrewold R (2018). Genome-Wide Association Study for the Relationship Between Temperature and Feed Intake in Beef Cattle (*Master's Thesis, University of Nebraska, Lincoln, Nebraska, USA*): 95.

Goddard ME, Hayes BJ (2007). Genomic selection. *Animal Breeding and Genetics* **124**: 323–330.

Goddard ME, Hayes BJ, Meuwissen THE (2010). Genomic selection in livestock populations. *Genetics Research (Camb)* **92**: 413–421.

Gosey J (2005). Crossbreeding the forgotten tool. *In: Proceedings, The Range Beef Cow Symposium XIX*. USA: University of Nebraska – Lincoln, 6–8 December. Rapid City, South Dakota, pp. 39–48.

Grigoletto L, Ferraz JBS, Oliveira HR, Eler JP, Bussiman FO, Abreu Silva BC, *et al.* (2020). Genetic Architecture of Carcass and Meat Quality Traits in Montana Tropical® Composite Beef Cattle. *Frontiers in Genetics* **11**: 123.

Habeeb AA, Gad AE, Atta MA (2018). Temperature-Humidity Indices as Indicators to Heat Stress of Climatic Conditions with Relation to Production and Reproduction of Farm Animals. *International Journal of Biotechnology and Recent Advances* **1**: 35–50.

Haskell MJ, Simm G, Turner SP (2014). Genetic selection for temperament traits in dairy and beef cattle. *Frontiers in Genetics* **5**: 368.

Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution* **41**: 51.

Hayes BJ, Lewin HA, Goddard ME (2013). The future of livestock breeding: Genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Genetics* **29**: 206–214.

Henderson CR (1984). *Applications of linear models in animal breeding*. Guelph: University of Guelph.

Hernández A, Ponce De León R, García SM (2015). Genetic evaluation of multibreed populations of milking bovines. *Cuban Journal of Agricultural Science* **49**: 459–463.

Herring AD (2014). Beef Cattle Production Systems. *CABI Publishers*, Wallingford, UK.

Hickey JM, Gorjanc G (2012). Simulated data for genomic selection and genome-wide association studies using a combination of coalescent and gene drop methods. *Genes, Genomes, Genetics* **2**: 425–427.

Higgins MG, Fitzsimons C, McClure MC, McKenna C, Conroy S, Kenny DA, *et al.* (2018). GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and *GFRA2* expression in beef cattle. *Scientific Reports* **8**: 14301.

Hill WG (1993). Brief Communications: Variation in genetic composition in backcrossing program. *The Journal of Heredity* **84**: 212–213.

Howard JT, Kachman SD, Snelling WM, Pollak EJ, Ciobanu DC, Kuehn LA, *et al.* (2014). Beef cattle body temperature during climatic stress: A genome-wide association study. *International Journal of Biometeorology* **58**: 1665–1672.

Jin S, Yang L, He T, Fan X, Wang Y, Ge K, *et al.* (2018). Polymorphisms in the uncoupling protein 3 gene and their associations with feed efficiency in chickens. *Asian-Australasian Journal of Animal Sciences* **31**: 1401–1406.

Johnston DJ, Tier B, Graser HU (2012). Beef cattle breeding in Australia with genomics: Opportunities and needs. *Animal Production Science* **52**: 100–106.

Jonas E, de Koning DJ (2015). Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. *Frontiers in Genetics* **6**: 49.

Kelleher J, Etheridge AM, McVean G (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS Computational Biology* **12**: e1004842.

Köbrich C, Bravo-Peña F, Rivas T, Maino M, Agüero D, Barcellos J, *et al.* (2018). *Estudio para la caracterización de la faena de animales y de los canales de comercialización de la carne bovina en los países del MERCOSUR*. Santiago, Chile.

de Koning DJ (2016). Meuwissen et al. On genomic selection. *Genetics* **203**: 5–7.

Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, Kingan SB, *et al.* (2018). De novo assembly of haplotype-resolved genomes with trio binning. *Nature Biotechnology* **36**: 1174–1182.

Koufariotis L, Hayes BJ, Kelly M, Burns BM, Lyons R, Stothard P, *et al.* (2018). Sequencing the mosaic genome of Brahman cattle identifies historic and recent introgression including polled. *Scientific Reports* **8**: 17761.

Kuehn LA, Keele JW, Bennett GL, McDaneld TG, Smith TPL, Snelling WM, *et al.* (2011). Predicting breed composition using breed frequencies of 50,000 markers from the US Meat Animal Research Center 2,000 bull project. *Journal of Animal Science* **89**: 1742–1750.

Lees AM, Sejian V, Wallage AL, Steel CC, Mader TL, Lees JC, *et al.* (2019). The impact of heat load on cattle. *Animals (Basel)* **9**: 322.

Legarra A, Aguilar I, Misztal I (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* **92**: 4656–4663.

Littlejohn MD, Henty KM, Tiplady K, Johnson T, Harland C, Lopdell T, *et al.* (2014). Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. *Nature Communications* **5**: 5861.

Liu H, Tessema BB, Jensen J, Cericola F, Andersen JR, Sørensen AC (2019). ADAM-Plant: A software for stochastic simulations of plant breeding from molecular to phenotypic level and from simple selection to complex speed breeding programs. *Frontiers in Plant Science* **9**: 1926.

Mader TL, Davis MS, Brown-Brandl T (2006). Environmental factors influencing heat stress in feedlot cattle. *Journal of Animal Science* **84**: 712–719.

Magalhães AFB, Schenkel FS, Garcia DA, Gordo DGM, Tonussi RL, Espigolan R, *et al.* (2019). Genomic selection for meat quality traits in Nelore cattle. *Meat Science* **148**: 32–37.

Mariasegaram M, Chase CC, Chaparro JX, Olson TA, Brenneman RA, Niedz RP (2007). The slick hair coat locus maps to chromosome 20 in Senepol-derived cattle. *Animal Genetics* **38**: 54–59.

Mendonça FS, MacNeil MD, Leal WS, Azambuja RCC, Rodrigues PF, Cardoso FF (2019). Crossbreeding effects on growth and efficiency in beef cow-calf systems: Evaluation of Angus, Caracu, Hereford and Nelore breed direct, maternal and heterosis effects. *Translational Animal Science* **3**: 1286–1295.

Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**: 1819–1829.

Meuwissen T, Hayes B, Goddard M (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers* **6**: 6–14.

Miller S (2010). Melhoramento genético de bovinos de corte utilizando informações da genômica. *Revista Brasileira de Zootecnia* **39**: 247–255.

Ministerio de Agricultura y Ganadería (2004). *Informe pais sobre la situación de los recursos zoogenéticos del Paraguay*. Asunción, Paraguay.

Mishra SP, Taraphder S, Roy M, Sarkar U, Datta S, Saikhom R, *et al.* (2017). Breeding techniques to exploit non-additive gene action for improvement of Livestock. *Bulletin of Environment, Pharmacology and Life Science* **6**: 126–134.

Montaldo HH, Casas E, Ferraz JBS, Vega-Murillo VE, Román-Ponce SI (2012). Opportunities and challenges from the use of genomic selection for beef cattle breeding in Latin America. *Animal Frontiers* **2**: 23–29.

Morenikeji OB, Ajayi OO, Peters SO, Mujibi FD, De Donato M, Thomas BN, *et al.* (2020). RNA-seq profiling of skin in temperate and tropical cattle. *Journal of Animal Science and Technology* **62**: 141–158.

Mrode R, Ojango JMK, Okeyo AM, Mwacharo JM (2019). Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects. *Frontiers in Genetics* **9**: 694.

Muntean CT, Herring AD, Riley DG, Gill CA, Sawyer JE, Sanders JO (2018). Evaluation of F1 cows sired by Brahman, Boran, and Tuli bulls for reproductive, maternal, and cow longevity traits. *Journal of Animal Science* **96**: 2545–2552.

Naidu C (2016). Metabolic profile and expression pattern of some genes in Tharparkar and Karan Fries (Tharparkar × Holstein Friesian) heifers during different seasons. *M.V.Sc Thesis submitted to ICAR-NDRI*, Karnal, Haryana: 56.

Neufeld-Arce RO (2006). Evaluation of heterosis and heterosis retention in *Bos taurus-Bos indicus* crossbred cattle for reproductive and maternal traits in cows (*Master's Thesis*, Texas A&M University, *Texas, USA*): 118.

Neves HHR, Carvalheiro R, O'Brien AM, Utsunomiya YT, do Carmo AS, Schenkel FS, *et al.* (2014). Accuracy of genomic predictions in *Bos indicus* (Nellore) cattle. *Genetics, selection, evolution* **46**: 17.

Nin-Pratt A, Freiría H, Muñoz G (2019). *Productivity and efficiency in grassland-based livestock production in Latin America: The cases of Uruguay and Paraguay*. Inter-American Development Environment, Rural Development and Risk Management Division IV. Title V. Series IDB-WP-1024. Washington DC, USA: 69.

Olson TA, Lucena C, Chase CC, Hammond AC (2003). Evidence of a major gene influencing hair length and heat tolerance in *Bos taurus* cattle. *Journal of Animal Science* **81**: 80–90.

Paim T do P, Hay EHA, Wilson C, Thomas MG, Kuehn LA, Paiva SR, *et al.* (2020). Dynamics of genomic architecture during composite breed development in cattle. *Animal Genetics* **51**: 224–234.

Peng B, Kimmel M (2005). simuPOP: A forward-time population genetics simulation environment. *Bioinformatics* **21**: 3686–3687.

Pérez-Enciso M, Forneris N, de los Campos G, Legarra A (2017). Evaluating sequence-based genomic prediction with an efficient new simulator. *Genetics* **205**: 939–953.

Pitt D, Sevane N, Nicolazzi EL, MacHugh DE, Park SDE, Colli L, *et al.* (2019). Domestication of cattle: Two or three events? *Evolutionary Applications* **12**: 123–136.

Polsky L, von Keyserlingk MAG (2017). Invited review: Effects of heat stress on dairy cattle welfare. *Journal of Dairy Science* **100**: 8645–8657.

Porto-Neto LR, Bickhart DM, Landaeta-Hernandez AJ, Utsunomiya YT, Pagan M, Jimenez E, *et al.* (2018). Convergent evolution of slick coat in cattle through truncation mutations in the prolactin receptor. *Frontiers in Genetics* **9**: 57.

De Roos APW, Hayes BJ, Goddard ME (2009). Reliability of genomic predictions across multiple populations. *Genetics* **183**: 1545–1553.

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, *et al.* (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832–837.

Sargolzaei M, Schenkel FS (2009). QMSim: A large-scale genome simulator for livestock. *Bioinformatics* **25**: 680–681.

Scasta JD (2015). Fire and Parasites: An Under-Recognized Form of Anthropogenic Land Use Change and Mechanism of Disease Exposure. *EcoHealth* **12**: 398–403.

Setshwaelo LL, Cundiff L V., Dickerson GE (1990). Breed effects on crossbred cow-calf performance. *Journal of animal science* **68**: 1577–1587.

Silvestri E, Schiavo L, Lombardi A, Goglia F (2005). Thyroid hormones as molecular determinants of thermogenesis. *Acta Physiologica Scandinavica* **184**: 265–283.

Sollero B., Gomes CCG, Roso VM, Higa RH, Yokoo MJ, Cardoso LL, *et al.* (2014). Goodness of Fit Comparisons among Five Bayesian Models in Genome-Wide Association of Tick Resistance in Brazilian Hereford and Braford Beef Cattle. *In: Proceedings, 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada*: 7–9.

Sutarno, Setyawan AD (2016). The diversity of local cattle in Indonesia and the efforts to develop superior indigenous cattle breeds. *Biodiversitas* **17**: 275–295.

Suybeng B, Charmley E, Gardiner CP, Malau-Aduli BS, Malau-Aduli AEO (2019). Methane emissions and the use of desmanthus in beef cattle production in Northern Australia. *Animals (Basel)* **9**: 542.

Swan AA, Graser HU (1988). Genetic correlations between gestation length, birth weight, 200-, 365- and 550-day weights in beef cattle. *In: Proceedings Association for the Advancement of Animal Breeding and Genetics* **7**: 493.

Tesema Z, Taye M, Ayichew D (2019). The role of phenotypic and genetic basis of livestock selection for climate change adaptation and mitigation: A review. *Journal of Applied and Advanced Research* **4**: 66.

Thrift FA, Sanders JO, Brown MA, Brown AH, Herring AD, Riley DG, *et al.* (2010). Preweaning, Postweaning, and Carcass Trait Comparisons for Progeny Sired by Subtropically Adapted Beef Sire Breeds at Various US Locations. *The Professional Animal Scientist* **26**: 451–473.

Vale P, Gibbs H, Vale R, Christie M, Florence E, Munger J, *et al.* (2019). The Expansion of Intensive Beef Farming to the Brazilian Amazon. *Global Environmental Change* **57**: 101922.

Van Eenennaam AL, Weigel KA, Young AE, Cleveland MA, Dekkers JCM (2014). Applied Animal Genomics: Results from the Field. *Annual Review of Animal Biosciences* **2**: 105–139.

VanRaden PM (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**: 4414–4423.

Voorrips RE, Maliepaard CA (2012). The simulation of meiosis in diploid and tetraploid organisms using various genetic models. *BMC Bioinformatics* **13**: 248.

Wakchaure R, Ganguly S, Kumar P, Sharma S, Kumar A, Tanvi M, *et al.* (2015). Importance of Heterosis in Animals : A Review. *International Journal of Advanced Engineering Techonology And Innovate Science* **1**: 1–5.

Wallage AL, Gaughan JB, Lisle AT, Beard L, Collins CW, Johnston SD (2017). Measurement of bovine body and scrotal temperature using implanted temperature sensitive radio transmitters, data loggers and infrared thermography. *International Journal of Biometeorology* **61**: 1309–1321.

Wallage AL, Johnston SD, Lisle AT, Beard L, Lees AM, Collins CW, *et al.* (2017). Thermoregulation of the bovine scrotum 1: measurements of free-range animals in a paddock and pen. *International Journal of Biometeorology* **61**: 1381–1387.

Weller JI, Kashi Y, Soller M (1990). Power of Daughter and Granddaughter Designs for Determining Linkage Between Marker Loci and Quantitative Trait Loci in Dairy Cattle. *Journal of Dairy Science* **73**: 2525–2537.

Weller JI, VanRaden PM, Wiggans GR (2013). Application of a posteriori granddaughter and modified granddaughter designs to determine Holstein haplotype effects. *Journal of Dairy Science* **96**: 5376–5387.

Widmann P, Nuernberg K, Kuehn C, Weikard R (2011). Association of an *ACSL1* gene variant with polyunsaturated fatty acids in bovine skeletal muscle. *BMC Genetics* **12**: 96.

Williams AR, Frankez DE, Saxton AM, Turner JW (1990). Two-, Three- And Four-Breed Rotational Crossbreeding of Beef Cattle: Reproductive Traits. *Journal of Animal Science* **68**: 1536–1546.

Wolfenson D, Roth Z (2019). Impact of heat stress on cow reproduction and fertility. *Animal Frontiers* **9**: 32–38.

Wright FA, Huang H, Guan X, Gamiel K, Jeffries C, Barry WT, *et al.* (2007). Simulating association studies: A data-based resampling method for candidate regions or whole genome scans. *Bioinformatics* **23**: 2581–2588.

Yuan X, Miller DJ, Zhang J, Herrington D, Wang Y (2012). An overview of population genetic data simulation. *Journal of Computational Biology* **19**: 42–54.

Zhang YD, Johnston DJ, Bolormaa S, Hawken RJ, Tier B (2014). Genomic selection for female reproduction in Australian tropically adapted beef cattle. *Animal Production Science* **54**: 16–24.

Zhao G, Zhang T, Liu Y, Wang Z, Xu L, Zhu B, *et al.* (2020). Genome-wide scan for runs of homozygosity identifies potential candidate regions associated with growth traits in beef cattle. *BMC Genomics* [***Preprint***].

Zhou H, Cheng L, Byun SO, Aizimu W, Hodge MC, Edwards GR, *et al.* (2017). Haplotypic variation in the *UCP1* gene is associated with milk traits in dairy cows. *Journal of Dairy Research* **84**: 68–75.

Zingaretti ML, Monfort A, Pérez-Enciso M (2019). pSBVB: A versatile simulation tool to evaluate genomic selection in polyploid species. *Genes, Genomes, Genetics* **9**: 327–334.

Chapter 7

Annexes

7.1. Supplementary material Paper I: 'Whole-genome sequencing reveals insights into adaptation of French Charolais cattle to Cuban tropical conditions'

Figure S1. Proportion of *B. indicus* per CHCU samples.

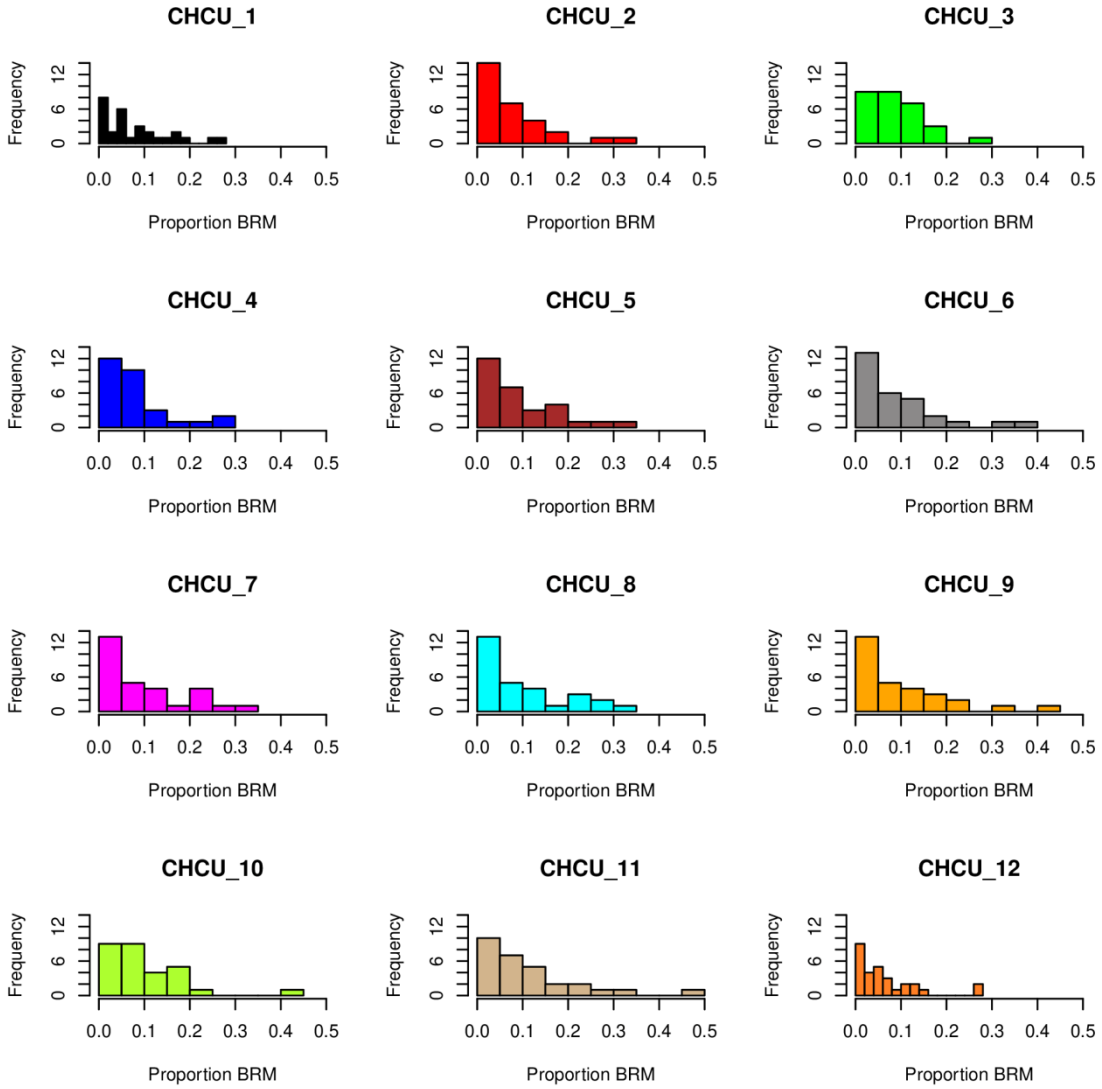

Figure S2. Relationship between proportion of *B. indicus* in CHCU and Fst CHCU-CHFR, each dot corresponds to a 30kb window

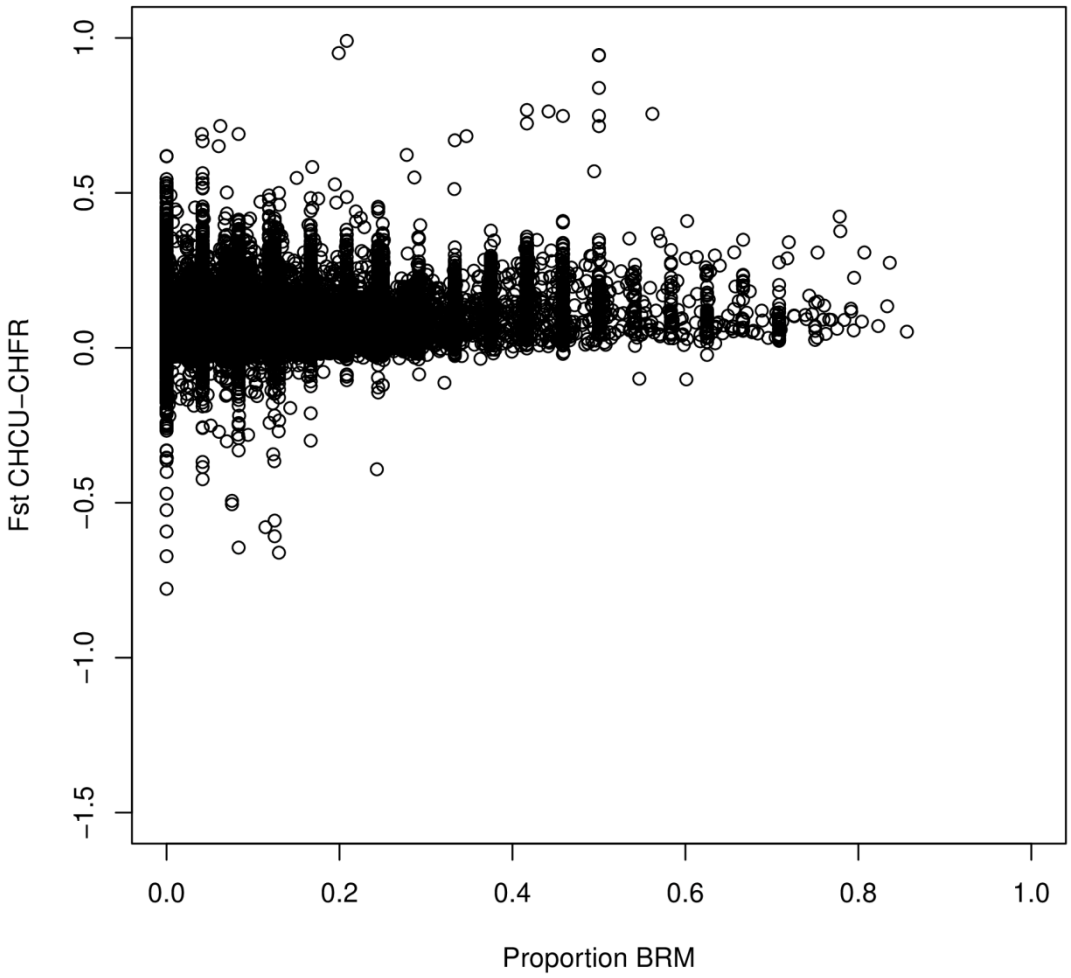

Table S1. Description of the samples employed in the study.

SampleID	Breed	Country	Accession	Depth
BRG1	Brangus	US	SRS1603295	10,8
BRG2	Brangus	US	SRS1603296	11,4
BRG3	Brangus	US	SRS1603297	13,3
BRG4	Brangus	US	SRS1603298	11,1
BRG5	Brangus	US	SRS1603299	11,3
BRM1	Brahman	US	SRS1603286	9,3
BRM2	Brahman	US	SRS1603288	11,1
BRM3	Brahman	US	SRS1603289	11,3
BRM4	Brahman	AU	SRS2894458	12,4
BRM5	Brahman	AU	SRS2894456	9,8
BRM6	Brahman	AU	SRS2894453	10,3
BRM7	Brahman	AU	SRS2894448	10,1
BRM8	Brahman	AU	SRS2928488	13,1
BRM9	Brahman	AU	SRS1603315	10,2
BRM10	Brahman	US	SRS1603285	10,7
CHCA1	Charolais	CA	SRS628369	14,3
CHCA2	Charolais	CA	SRS628370	11,4
CHCA3	Charolais	CA	SRS628372	11,5
CHCA4	Charolais	CA	SRS629321	14,0
CHCA5	Charolais	CA	SRS629334	10,0
CHCA6	Charolais	CA	SRS629687	9,9
CHCA7	Charolais	CA	SRS629691	15,8
CHCA8	Charolais	CA	SRS631230	15,1
CHCA9	Charolais	CA	SRS666710	6,9
CHCA10	Charolais	CA	SRS666715	6,8
CHCA11	Charolais	CA	SRS42866	8,7
CHCA12	Charolais	CA	SRS428702	7,2
CHCA13	Charolais	CA	SRS428703	7,8
CHCA14	Charolais	CA	SRS428698	8,1
CHCA15	Charolais	CA	SRS428678	8,7
CHCU1	Charolais_cu	CU	CU12_GATCAG_L006.bam	16,3
CHCU2	Charolais_cu	CU	CU13_TAGCTT_L006.bam	11,4
CHCU3	Charolais_cu	CU	CU16_GGCTAC_L006.bam	12,6
CHCU4	Charolais_cu	CU	CU20_CTTGTA_L006.bam	11,5
CHCU5	Charolais_cu	CU	CU2_ATCACG_L006.bam	10,1
CHCU6	Charolais_cu	CU	CU3_CGATGT_L006.bam	11,6

CHCU7	**Charolais_cu**	**CU**	**CU4_TTAGGC_L006.bam**	**17,5**
CHCU8	Charolais_cu	CU	CU5_TGACCA_L006.bam	11,3
CHCU9	Charolais_cu	CU	CU6_ACAGTG_L006.bam	10,5
CHCU10	Charolais_cu	CU	CU7_GCCAAT_L006.bam	10,1
CHCU11	Charolais_cu	CU	CU8_CAGATC_L006.bam	14,5
CHCU12	Charolais_cu	CU	CU9_ACTTGA_L006.bam	13,4
CHFR1	Charolais	FR	1000 Bull Genomes Project	9,4
CHFR2	Charolais	FR	FR0395121015_NoIndex_L001.bam	8,1
CHFR3	Charolais	FR	FR4286100325_NoIndex_L004.bam	12,1
CHFR4	Charolais	FR	1000 Bull Genomes Project	11,6
CHFR5	Charolais	FR	1000 Bull Genomes Project	12,9
CHFR6	Charolais	FR	FR7121520725_NoIndex_L002.bam	10,8
CHFR7	Charolais	FR	FR7185119662_NoIndex_L007.bam	8,1
CHFR8	Charolais	FR	FR7187110011_NoIndex_L002.bam	12,7
CHFR9	Charolais	FR	1000 Bull Genomes Project	11,6
CHFR10	Charolais	FR	1000 Bull Genomes Project	12,0
CHFR11	Charolais	FR	1000 Bull Genomes Project	12,4
CHFR12	Charolais	FR	1000 Bull Genomes Project	12,8
CHFR13	Charolais	FR	FR8587103951_NoIndex_L007.bam	12,5
CHFR14	Charolais	FR	FR8589103317_NoIndex_L005.bam	11,6
CHFR15	Charolais	FR	PAPRIKA_NoIndex_L006.bam	13,5
LIMS1	Limousin	FR	ERS212674	8,1
LIMS2	Limousin	FR	ERS212672	7,0
LIMS3	Limousin	FR	1468LM	10,7
LIMS4	Limousin	FR	ERS2582780	7,0
LIMS5	Limousin	FR	ERS477550	8,3
LIMS6	Limousin	FR	ERS212673	9,0
LIMS7	Limousin	CA	SRS428692	7,4
LIMS8	Limousin	CA	SRS428696	7,6
LIMS9	Limousin	CA	SRS428697	7,6
LIMS10	Limousin	CA	SRS631219	12,0
LIMS11	Limousin	CA	SRS631220	12,0
TXL1	Texas Longhorn	US	SRS1602493	10,8
TXL2	Texas Longhorn	US	SRS1602494	10,7
TXL3	Texas Longhorn	US	SRS1602495	9,2
TXL4	Texas Longhorn	US	SRS1602496	10,6

Table S2. Description of the intervals identified as selective sweeps between Cuban and French Charolais (DOI: 10.6084/m9.figshare.12465845).

Table S3. Putative regulatory SNPs (rSNPs) located upstream of *UCP1* gene (DOI: 10.6084/m9.figshare.12465845).

7.2. Supplementary material Paper III: 'Evaluating genomic prediction strategies in crossbred *indicine* x *taurine* cattle programs'

Figure S1. Structure of the breeds analysed**. (A)** Principal component analysis using all samples. Individuals are grouped in *B. taurus*, *B. indicus* and Hybrid. Black: Angus, red: Brangus, blue: Brahman, cyan: Hereford, green: Nelore, orange: Red Angus, magenta: Senepol, purple: Santa Gertrudis; **(B)** Principal component analysis with Angus and *B. indicus* only. Red: Angus (Angus + Red Angus), blue: *B. indicus* (Brahman + Nelore).

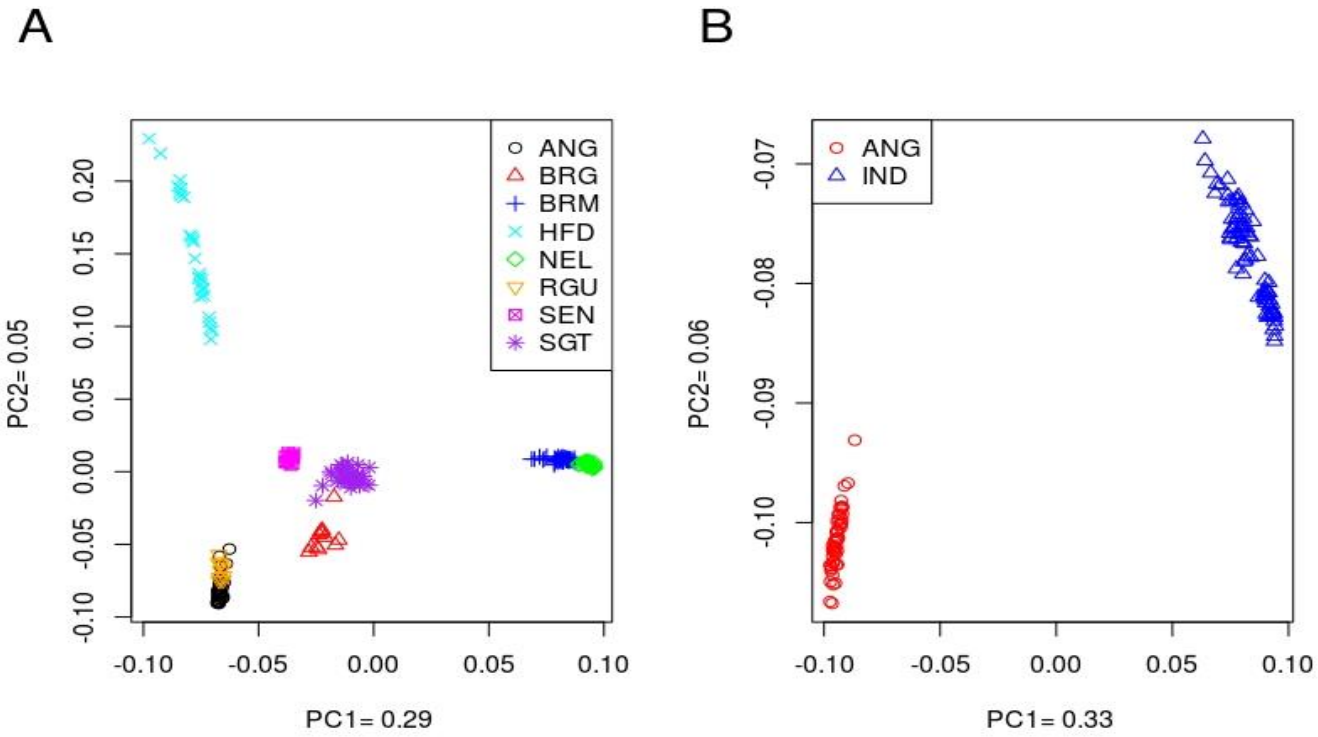

Table S1. QTL regions for each of these three sets of phenotypes from the QTL database (DOI: 10.6084/m9.figshare.12465845).

Acknowledgments

La vida es una eterna escuela me dijo una vez mi madre… Y empezar a escribir estas líneas con el temor de olvidarme de alguien que estuvo presente en mi escuela en particular, me resulta un poco difícil.

Antes que nada, doy gracias a Dios y a la Virgen por darme la posibilidad de completar una etapa más en mi vida.

Fueron muchas personas las que estuvieron involucradas en esta travesía. Primero mi familia. Aunque ya no esté más con nosotros físicamente, mi padre Aldo Luis fue la persona que construyó los cimientos para forjarnos un futuro mejor. Mi eterno agradecimiento por los años que nos brindó todo lo que estaba a su alcance y dejar encaminado nuestro futuro antes de emprender el viaje de su vida. Después, la persona que comenzó a construir sobre esos cimientos, mi madre Zulmina Edit, que siempre quiso que salgamos del país para ver y escuchar otro tipo de "problemas" como dice ella, basándose en su propia experiencia. Mi más sincera admiración, respeto y amor a la persona que me dio la vida y que nunca escatimo recursos hacia mí. Ahora que soy papá puedo entender, capaz, una pequeña parte de todo lo que se hace por un hijo. Solamente decirte "gracias" se queda en nada a lado de todo lo que me diste Mamá! Espero que la vida me dé la oportunidad de retribuirte en algo porque no me va a alcanzar la misma para devolverte todo lo que hiciste por mí. Detrás vienen mis herman@s, Mirta Edith, Liz Teresita, Víctor Aldo y Lourdes Angelina, que me malcriaron toda la vida. Gracias por aguantarme durante tantos años y aun así estar siempre para mí, incluso en mis momentos de malas decisiones. Al igual que a Mamá, me faltan las palabras para expresarles lo que siento al escribir estas palabras. Gracias! Y en el sentido más amplio de la palabra!

A mi cuñado, Luis Fernando, que nos dio a nuestr@s primer@s amores, José Daniel, Juan Miguel, Jazmín de María y Gema Teresita. Por tu paciencia y por soportar tantas cosas, que ni nosotros mismos podemos dimensionar.

A esos hermanos que te regala la vida y que no son de sangre, Blas, Gaby, Ron y Navarro. Fueron partícipes de todo el proceso y saben todo lo que tuvimos que remar para llegar a este día. Gracias por la amistad incondicional, por estar para nosotros a cualquier hora del día, por las palabras de aliento y la buena onda que siempre nos brindaron. Son muchos años de amistad y espero que sea siempre así.

Acknowledgments

A mi tutor de Master, Roberto (alias Tulo o Profe) que fue la persona que me dijo “andate”, textualmente. Gracias Profe por todas las ayudas! Vos me diste la oportunidad de crecer como persona y como profesional y eso te voy a agradecer toda mi vida.

A mis amigos de la Alianza Francesa, que juntos emprendimos este desafío con el Profe Carli a la cabeza, una persona que, con mucha paciencia, nos enseñó y supo transmitir las mejores directrices para que nos sepamos desenvolver en la vida. A todos y cada uno de ellos, gracias por los excelentes momentos que pasamos juntos, en especial a Luis (manex) y Heidy (mami), que siempre estuvieron dispuestos a ayudarme en todo sentido, por darme un lugar en París cuando aún mi destino era incierto y por seguirme en cada macanada.

A mis amigos paraguayos en Barcelona, Hugo (Hugay), Noemí (amichi), Germán (Papex o vamos a la playa con pandi), Nathi (Helena), Fátima (Fatuchi), Leticia (Letti), Vicky (whisky), Lourdes (prima), Giorgio (italo-paraguayo), Alma (italo-paraguaya-catalana) y Rolando (Bebeto), que me abrió las puertas de su casa ni bien llegué acá. Gracias por el aguante y las ayudas desinteresadas.

A los amigos que me regaló el idioma catalán, Sergio, Jele, Graciela, Ana, Miguel, Astrid, Erika, Gleison, Abiu, Jimmy, Alejandro, Selene, Katy, Elena y en especial a Andrés (pana) y Diana (Di) o en general, l@s ti@s bichit@s en honor a Aby, por esa amistad verdadera y sincera, por las ayudas incondicionales, por cada detalle que tienen con nosotros, por esas actividades que organizan con tal de verle feliz a mi pequeña y sobre todo, por el gran amor que le tienen.

A mis directores de tesis. Primero a Miguel (alias Doctor o big boss), por darme la oportunidad de estar en este mundo paralelo y por la paciencia extraordinaria que desarrollaste solo para mí. Por tu ayuda y preocupación en todos los ámbitos y el tiempo que me dedicaste para finalizar este doctorado. A Yuli (alias Maestro Jedi, Power Ranger Rojo, Maestro de los 7 picos o Jefecito), que llegó en el momento exacto para ayudarnos a terminar el doctorado. Por esos consejos sobre como sobrellevar el doctorado, las cosas que me enseñaste en tu estado “normal” (moto GP) y esas charlas alentándome siempre a no decaer. Perdón por enviarte más trabajo los findes y gracias por también la paciencia extraordinaria, Jefecito.

A Sebas, por darme de comer y alentar así mi nula vida saludable. Por esas charlas que empezábamos hablando del paper y terminábamos hablando de nuestras hijas. Gracias por ese sentido del humor y hacer tonterías para alegrarnos el día.

A los demás IPs, JM, Alex, Marcel, con quienes compartíamos charlas de pasillo sobre temas varios, y con Marcel, evaluaciones de los seguimientos, siempre con esa mirada de "a ver que le pregunto….". Espero que siempre compartamos la afición hacia la Voll-Damm.

A todas las personas con quienes tenía charlas de pasillo de sobre temas random, sobre todo a Maite (piruleta), que aparte de las charlas, siempre se ofrecía a acercarme a mi casa después de una larga jornada, aunque a veces nos quedábamos a comer y tomar (nuestra especialidad) por nuestra querida Cerdanyola.

A mi primera compañera de despacho, Nati Forneris (amea), quien me dio la bienvenida al despacho 301 y me enseñó mis primeros scripts. Por esas charlas sobre la vida, que hasta hoy las recuerdo y me han ayudo un montón.

A mis compañeros de despacho vegetalógolos (inventado), Miguel (Micky), Nuria (Nuri) y Christian (Chris), por aguantarme todos estos años, sobre todo cuando entraba gritando y asustándoles (cada día durante más de 3 años). A Alice (Alicia), por adaptarse tan rápido a nosotros, los animalógolos (más inventado), por esos días que poníamos música para bailar en el despacho y aguantar también mis sustos.

A mi compañero de despacho y amigo, Elies (my friend). Por las ayudas en todo sentido! Por esas charlas de catarsis intensas (en persona y por hangouts), por tu amistad desinteresada y por sobre todo, por mostrarme otro punto de vista sobre la vida. Sos un tipazo y sabes lo mucho que te aprecio. Ojalá todas las personas se pudieran tomar el tiempo de conocer a alguien y no prejuzgarla antes. Una de las tantas cosas que aprendí contigo. Por esas sesiones de metal ópera para canalizar nuestro Ki (en términos de Dragon Ball).

Aunque fue por muy poco tiempo compañero de despacho, siempre estaba presente en cada actividad que hacíamos todos juntos, Joan. Un tipazo y quizás la persona más loca que conocí programando. Gracias por estar siempre en todas nuestras reuniones (sobre todo nuestros días sagrados de cine + la cueva) y aportar un poco de cordura al grupo.

Por orden de antigüedad (se refiere a la finalización del doctorado, no a la edad, para no herir sensibilidades), a mi amigo Rayner (papichulo o tiburón), por esa forma de ser que solo vos podes tener. Compartimos muy poco en el Crag, pero lo suficiente crear una amistad que se mantiene hasta hoy. Gracias en particular por el ISAG, que me diste un lugar donde quedarme y por la alegría que siempre transmitís. Por esas salidas de baile que ayudaban un montón a reducir el estrés. Y acá añadir a mi querida Antonia (Anto), que junto a la Petretto (próxima

incorporación al Crag) irradiaban alegría y ganas de pasarla bien. Por esas invitaciones para salir solo a divertirnos, fiel a sus estilos. Se les extraña!

A Manuel (Manu), que también compartimos muy poco en el Crag, pero que a pesar de la distancia, mantenemos el contacto. La verdad, nunca había conocido a alguien tan diplomático como vos Manu! Y gracias por aquella maratón de madrugada para alcanzar el bus… nunca lo voy a olvidar! (lastimosamente no se pueden poner emojis acá).

A mi querida brasilera Tainã (Tai), por las ayudas en todos los sentidos! La brasilera más rara que conocí en mi vida (no le gusta sambar). Por esas salidas en donde a las 10 p.m. cada uno ya se tenía que ir a su casa porque vos tenías sueño, una noche loca digamos. Gracias Tai por tus palabras de aliento, por tu amistad, tus consejos y por aguantarme, sobre todo con el tema de Leoncito de Peluche. Sabes que soy pesado (físicamente y de carácter) pero es mi manera de demostrar cariño.

A mi querido amigo Dani, que nunca le puse un apodo pero me gusta decirle Piazuelo. Ufff Dani… en vos encontré un verdadero e incondicional amigo. Me ayudaste muchísimo y en todos los sentidos! Sos una excelente persona y me aguantaste muchísimo. Gracias por esas charlas a altas horas de la noche en el Crag, en mi casa o donde fuese, siempre dispuesto a escuchar y ofrecer su ayuda. Por el cariño que junto a Celia le ofrecieron a mi hija desde el primer día. Sabes el cariño que le tenemos a los dos y te agradezco ese don de gente que tenes. Y por sobre todo, por aportar datos que nadie nunca sabía acerca de las películas en nuestros días sagrados de cine + la cueva.

A Marta (Martuki), por las incontables plantas de menta para mi tereré, por hacer que las horas del almuerzo sean las mejores, invitándome de tus comidas (incluye acelgas), por las largas charlas a altas horas de la noche en el Crag. Por tu ayuda y preocupación hasta el final de mi doctorado, por aceptarme tal cual soy!

A Jordi (George, boludo, feo, SENSEI), que si pongo por las ayudas en todos sentidos me quedo corto! Con ese peculiar sentido del humor (mi yo español) supimos tener una amistad que nos llevó a compartir muchas cosas, entre ellas el amor por nuestra hijas (ah y por nuestras respectivas esposas también…). Pero también por otras cosas que solo vos y yo sabemos (copiado textualmente de tus agradecimientos, #MdV, entre otras cosas). Gracias George por ese humor pesado, que creo es lo que más nos une (y nos distancia de los demás). Por esas salidas al cine y después al lugar quizás más "cutre" que puede existir, pero donde fuimos, somos y seremos siempre felices (claro, si seguís teniendo permiso) Y aprovechar

agradecerle a Vane por las veces arriesgadas que te dejó salir solo conmigo. Por esos miles de scripts a las 4:45 p.m (hasta el último día) y por ser realmente una persona con la que siempre me la paso superrrrrr.

A Lourdes (mi pequeña Pony), a quien le debo más que unas simples gracias también. Sabes lo mucho que te quiero y la verdad necesitaría muchas líneas para expresarte todo mi agradecimiento. Fuiste la primera persona que se acercó a mí cuando llegué y fuiste la última con la que hablé al momento de escribir estas líneas. Fuiste Santa Claus, Reyes Magos, agencia de viajes y guía turística de vacaciones, entre otras muchas cosas. Siempre dispuesta a ayudarme, en todos los sentidos. Nunca me dijiste que no, siempre buscabas la manera y lo conseguías. Le diste a mi hija sus primeras vacaciones a la montaña, su primer paseo en caballo y eso lo voy a llevar en mi corazón toda la vida. Sos de las que me manda de paseo (para no poner una mala palabra) cuando no te pido ayuda, siempre fiel a tu estilo. Gracias mi pequeña Pony! Y por ende, a tus padres (Doña Lourdes y Pepe) por todos los regalos para mi hija y porque sé que Pepe siempre quiso tener un hijo como yo (…). Y obviamente no podía faltar Manel (cariño), porque se supo amoldar a nuestra situación (fiestas controladas por presencia de menores de edad) y por enseñarme la frase que lo llevo muy presente cada día, "si tenes la lengua dura, tenes que tener los oídos duros también". Simplemente, fantástico.

A Emilio (dear Sir Emilic), uno de los crack siendo estudiante. Admiro muchísimo esa capacidad que tenes de saber un poco de todo Emilic, sos un capo! Me ayudaste a entender muchísimas cosas y hasta días antes de leer la tesis, porque hablando contigo, la ciencia parece super fácil. Gracias por el tiempo y consejos que siempre te lo tomas muy en serio, que según vos, no estabas tan seguro, pero siempre terminas acertando. No tengo ninguna duda de que vas a conquistar todos tus sueños.

Siguiendo con otra crack, Laura (Lauchis, baby). Hace unos días volvimos a hablar y escribo textualmente todo lo que te dije y que siempre pensé de vos. Desde el día 1 que nos conocimos, me demostraste una amistad incondicional. También contigo me quedo corto al decirte solo "gracias". Nunca voy a olvidar todo lo que me ayudaste, el tiempo que me dedicaste y todo lo que me enseñaste. Gracias por todo eso y en especial, por aquella noche en el hospital, que no dudaste ni un minuto para quedarte conmigo. Lo llevo como un tatuaje en el corazón ese detalle que tuviste conmigo. Amo tus salidas zarpadas y ojalá nunca cambies eso, porque te hace ser única. Tod@s coincidimos que, junto con Emilic, están por encima del promedio de cualquier estudiante normal de doctorado.

Acknowledgments

A María Gracia (Mary Poppins o cualquier otro apodo menos el que ella quiso siempre). Ni bien llegaste, ya te uniste de lleno a nosotros, por esa extraordinaria capacidad que tenes de amoldarte al ambiente (sobre todo en cinéfilos). Siempre te digo que en vos encontré esa amiga "camionera", con la que podes quedar para tomar en cualquier parte, a cualquier hora. Mil gracias por la ayuda en todo sentido Poppins! Por esas tardes en la que fuiste una mala influencia y me llevabas a la cueva bar para tomar y hablar de todo y a la vez, de nada. Cuando de repente, la charla ya no tenía sentido, igual no nos callábamos. Voy a extrañar esas salidas contigo! Porque no sé si vas a seguir teniendo permiso, Doña Poppins.

To Ioanna (Juanita), for the caveman talks in English and for helping me improve on that topic. Thanks for the times you approached my house and let me come to Nikos' side while you were in the back seat (I always understood that it was not going to fit there).

To Dailu (uno), who although he didn't understand me at all, was always ready to have a conversation. Thanks to you, they will never again invite us to a happy hour at the Crag, I think that was the best contribution you could make in the group.

A l@s nuev@s, con cariño, Yron, Jesús, Magí, Lian y Cristina, porque aportan diversidad al grupo y nos ayudaron a renovar nuestras bromas pesadas. Les deseo el mayor de los éxitos en este camino que ya están transitando. Nunca dejen de reunirse para intercambiar papers como lo hacíamos nosotros (…).

Y lo mejor viene al final. Mis amores, Moni y Aby. Necesitaría más que una tesis para decirles todo lo que significan en mi vida. Aunque muchas veces no lo demuestre, son los pilares sobre la que mi vida transcurre, los motores que hacen que cada día valga la pena luchar, las fuerzas para remar, muchas veces contra la corriente, pero a sabiendas que juntos, los tres, conseguiremos llegar a buen puerto. Gracias por malcriarme, por las comidas (de verdad a Moni y de juego a Aby), por respetar el tiempo que tenía que trabajar, por no escuchar, ni ver, reclamo alguno del tiempo que no pasé con ustedes. Gracias Moni por amarme incondicionalmente desde hace más de 10 años, por las palabras de aliento, por ser una super mamá y esposa. Perdón porque sé que sufriste muchas veces en silencio para que no me preocupe y porque muchas veces estuve ausente, centrado en mi mundo paralelo. Y por sobre todas las cosas, gracias a mi chiquitita, por ser la persona que más me enseño en tan poco tiempo. Cambiaste mi vida radicalmente para bien desde la primera vez que te tuve en mis brazos, porque un solo abrazo o sonrisa tuya es suficiente para hacerme el día. Le das

color a mi vida, cada día, todos los días. Espero ser el papá que siempre quisiste tener. Te amo más que a nada en esta vida.

Espero no haberme olvidado de nadie y de ser así, si tienen la oportunidad de leer estas últimas líneas decirle a todos y cada uno, ***Aguyjevete!!!***

Y ahora si oficialmente,

Ala! Me'n vaig!!!