Head Office: Università degli Studi di Padova

Department of Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE)

Ph.D. COURSE IN: Animal Food Science

CURRICULUM: Animal breeding and Genetic

SERIES XXXIV

Introduction of new genetic and genomic tools in local cattlebreeds

Thesis written with the financial contribution of Università degli studi di Padova

Coordinatore: Ch.mo Prof. Angela Trocino

Supervisore: Ch.mo Prof. Roberto Mantovani

Co-Supervisore: Ch.mo Prof. Cristina Sartori

Dottorando : Enrico Mancin

To Mendelian sampling otherwise

I would be unemployed

Thesis contents

1. ABSTRACT

Local cattle breeds are characterized by strict historical and environmental connection with the areas of diffusion. They presented different phenotypic and genetics aspect among them, due to the adaptation to different environment and partially to genetic drift due to the environment. Preserving this diversity is essential to ensure future food production in an environment that is constantly changing due to climate change, i.e., temperature increment and reduced distribution of rainfall. Furthermore, they play an important role in producing services unrelated to food production such as ecosystem services, but also cultural and historical services. In addition, they have a primary role to sustain the local economy especially in rural areas, being linked often to specific traditional products. Despite this a deterioration of local breeds consistency has taken place in last decades. This was due by the progressive substitution with more specialized/productive breeds that ensured a generally higher profitability. However, higher profitability to these breeds is not entirely impossible, and it can be done by farming them in low-input environments, and/or through a further valorization of breed specific product. The competitiveness of these breeds can also be achieved by adequate breeding plans that guarantees a progressive increase of the productive traits with the maintenance of the typical traits and low increase of inbreeding levels. The DUALBREEDING is a ministerial project aimed at promoting the competitiveness of local dual-purpose breeds in Italy through the good breeding practices described above. It is based on some milestones as the monitoring of inbreeding, selection for functional traits and longevity. Three breeds involved in the DUALBREEDING project were considered in the present thesis: the Alpine Grey (Grigio Alpina), Reggiana, and Rendena. In each breed three different approaches are developed to promote the local breed within the territory of origin. For example, the Alpine Grey breed is the quintessential alpine breed, where there is a strong link between farming environments and breeding. The Reggiana, on the other hand, represented the symbol of the valorization of autochthonous breeds through the animal-breed-food association. Lastly, the Rendena breed is particularly appreciated by breeders for its "rusticity" and its ability to apt in different environment combined with good milk and meat production. In this study, thanks to the data and directives provided by the DUALBREEDING project, some strategies have been developed to improve and enhance the genetic value of the three breeds. The first approach was based on the development variance components estimation and response to selection for each breed. A technical note was then produced to demonstrate how to

derive selections indexes by attributing specific economic weights considering even constrains on some traits.

New selection indices and response to selection were then calculated for Alpine Grey breeds. In that study we demonstrate that the current selection index leads in the medium-long term to a detriment of genetic progress for beef, functional characteristics of the cattle. For these reasons we have presented various selection indices more oriented to the dual aptitude of these cattle without worsening some morphological characteristics appreciated by breeders, maintaining a modest selective response for milk production. While for Reggiana the variance of milk components and fertility traits was estimated. Furthermore, for these phenotypes the quote of Genotyped from the environmental interaction (GxE) was calculated. In that study we identify a significant quote of GxE expressed by those traits, however the models that consider GxE do not differ in terms of EBV accuracy and bull re-ranking. The second study was focused on the introduction of genomic selectionin Rendena breed. At first, data on performance test were analyzed with “classical” BLUP. Three models were then compared: (i) Pedigree-BLUP (PBLUP); (ii) single-step GBLUP (ssGBLUP), and (iii) weighted single-step GBLUP (WssGBLUP). We identify that the models including genomic information presented higher accuracies than PBLUP, especially WssGBLUP. However, the model with the best overall properties was the ssGBLUP, showing higher accuracy than PBLUP with optimal values of bias and dispersion parameters. This study demonstrated that integrating phenotypes for beef traits with genomic data can be helpful to estimate performance test EBVs, even in a small local breed. The subsequent study consisted in a further implementation of genomic selection in the Rendena cattle. This time we investigated several alternative methods to improve the accuracy of genomic selection in the population. Particularly, the impact of using only a subset of informative markers regarding accuracy of prediction, bias, and dispersion, was investigated. We tested different machine learning variable selection algorithms to select the SNPs, i.e., LASSO, recursive recursive feature elimination and Extreme Gradient Boost. At first, in a simulated dataset we benchmark the performance of ssGBLUP with variable selection models with the models mentioned above. Simulation differs in terms of number of QTLs and effective population size. Then, these approaches were implemented on the Rendena performance test dataset. Our results showed that the accuracy of GBLUP in small sized populations increase when performed with SNPs selected via variable selection methods both in simulated and actual datasets. In addition, the use of variable selection

models – especially those using XGboost – in the actual dataset did not impact on bias and the dispersion of estimated breeding values.

In the last part of this thesis, a study on the genetic makeup of local breeds, by conducting a GWAS analysis on Rendena cattle breed, was carried out. However, since there were several sources of information for Rendena (animals with and without genotyping or phenotype), an efficient methodcombining each information was needed. For this purpose, single-stepGWAS (ssGWAS) promises a good strategy. However, its ability to account for population structure has not been explored. We investigated the equivalence among ssGWAS, efficient mixed-model association expedited(EMMAX), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS), and how they differ from the single-SNP analysis without correction for population structure (SSA-NoCor). Simulated datasets were then constructed and structured populations that mimicked fish, beef cattle, and dairy cattle populations with 1,040, 5,525, and 1,400 genotyped individuals, respectively, were produced. A larger population that had up to 10-fold more genotyped animals were also simulated. In dairy cattle phenotypes of daughters were projected into genotyped sires (i.e., de-regressed proofs) before applying EMMAX and SSA-NoCor. Although SSA-NoCor had the largest number of true positive SNPs among the four methods, the number of false negatives was two–fivefold that of true positives. Interesting we found that GBLUP-GWAS and EMMAX had a similar number of true positives, which was slightly smaller than in ssGWAS, although the difference was not significant. After the validation of single-step GWAS performance, the equation was used in the Rendena actual dataset. GWAS a post-GWAS analysis for body weight (BW), average daily gain (ADG), carcass fleshiness (CF) and dressing percentage (DP) in 1,690 individuals were then carried out. Moreover, we considered two of the target phenotypes (BW and ADG) at different times in the individuals' life, a potentially important aspect in the study of the traits' genetic architecture. We identified 8 significant and 47 suggestively associated SNPs, located in 14 autosomal chromosomes (BTA). Among the strongest signals, 3 significant and 16 suggestive SNPs were associated with ADG and were located on BTA10 (50–60 Mb), while the hotspot associated with CF and DP was on BTA18 (55–62 MB). Among the significant SNPs some were mapped within genes, such as SLC12A1, CGNL1, PRTG (ADG), LOC513941 (CF), NLRP2 (CF and DP), CDC155 (DP). Concluding, although the improvement of local breeds plays a secondary role, the results produced in this thesis seem suitable for the breeding

systems even in small local populations, in terms of selection / accuracy plans but also in the enhancement of their genomic heritage

2. ABSTRACT (ITA)

Le razze bovine autoctone sono caratterizzate da uno stretto legame storico e ambientale con i territori di diffusione. Esse presentano una varietà aspetti fenotipici e genetici, dovuti all'adattamento al diverso ambiente e anche in parte alla deriva genetica dovuta all'ambiente. La conservazione di questa diversità è essenziale per garantire la futura produzione alimentare in un ambiente in costante mutamento a causa dei cambiamenti climatici, come l'aumento della temperatura e la ridotta distribuzione delle precipitazioni. Inoltre, svolgono un ruolo importante nella produzione di servizi esterni alla produzione alimentare come i servizi ecosistemici, ma anche i servizi culturali e storici. Inoltre, hanno un ruolo primario nel sostenere l'economia locale soprattutto nelle zone rurali poiché esse sono spesso legate a specifici prodotti tradizionali. Nonostante ciò, negli ultimi decenni si è verificato un deterioramento della consistenza delle razze locali. Ciò era dovuto alla progressiva sostituzione con razze più specializzate/produttive che garantivano una redditività generalmente maggiore. Tuttavia, una maggiore redditività per queste razze non è del tutto impossibile e può essere ottenuta allevandole in ambienti a basso input e/o attraverso un'ulteriore valorizzazione del prodotto specifico della razza. La competitività di queste razze può essere raggiunta anche da adeguati piani di allevamento che garantiscano un progressivo incremento dei tratti produttivi con il mantenimento dei tratti tipici e un basso incremento dei livelli di consanguineità. Il DUALBREEDING è un progetto ministeriale volto a promuovere la competitività delle razze locali a duplice attitudine in Italia attraverso le buone pratiche di allevamento sopra descritte. Si basa su alcune pietre miliari come il monitoraggio della consanguineità, la selezione dei tratti funzionali e la longevità. Nella presente tesi sono state considerate tre razze coinvolte nel progetto DUALBREEDING: la Grigia Alpina (Grigio Alpina), la Reggiana e la Rendena. In ogni razza vengono sviluppati tre diversi approcci per promuovere la razza locale all'interno del territorio di origine. Ad esempio, la razza Alpine Grey è la razza alpina per eccellenza, dove esiste un forte legame tra ambienti di allevamento e allevamento. La Reggiana, invece, rappresentava il simbolo della valorizzazione delle razze autoctone attraverso l'associazione animale-razza-alimentazione. La razza Rendena, infine, è particolarmente apprezzata dagli allevatori per la sua "rusticità" e per la sua capacità di adattamento in ambienti diversi abbinata ad una buona produzione di latte e carne. In questo studio, grazie ai dati e alle direttive fornite dal progetto DUALBREEDING, sono state sviluppate alcune strategie per migliorare e valorizzare il valore genetico delle tre razze. Il primo approccio era basato sulla stima delle

componenti della varianza dello sviluppo e sulla risposta alla selezione per ciascuna razza. Come prima cosa è stata sviluppata una nota tecnica per dimostrare come ricavare indici di selezione attribuendo pesi economici specifici considerando anche i vincolisu alcuni tratti.

Sono stati quindi calcolati nuovi indici di selezione e risposta alla selezione per le razza Grigio Alpina. In questo dimostriamo che l'attuale indice di selezione porta nel medio-lungo periodo a scapito del progresso genetico della carne bovina, caratteristiche funzionali del bovino. Per questi motivi abbiamo presentato vari indici di selezione più orientati alla duplice attitudine di questi bovinisenza peggiorare alcune caratteristiche morfologiche apprezzate dagli allevatori, mantenendo una modesta risposta selettiva per la produzione di latte. Mentre per la Reggiana è stata stimata la varianza delle componenti del latte e dei tratti di fertilità. Inoltre, per questi fenotipi è stata calcolata la quota di interazione genotipo per ambiente (GxE). In questo studio identifichiamo una quota significativa di GxE espressa da quei tratti; tuttavia, i modelli che considerano GxE non differiscono in termini di accuratezza EBV e riclassificazione rialzista. Il secondo gruppo di ricerche si è concentrato sull'introduzione della selezione genomica nella razza Rendena. Inizialmente, i dati sul test delle prestazioni sono stati analizzati con BLUP "classico". Sono stati quindi confrontati tre modelli: (i) Pedigree-BLUP (PBLUP); (ii) GBLUP a fase singola (ssGBLUP) e (iii) GBLUP a fase singola ponderata (WssGBLUP). Identifichiamo che i modelli che includono informazioni genomiche presentavano precisioni maggiori rispetto a PBLUP, in particolare WssGBLUP. Tuttavia, il modello con le migliori proprietà complessive è stato ssGBLUP, che mostra una maggiore precisione rispetto a PBLUP con valori ottimali di bias e parametri di dispersione. Questo studio ha dimostrato che l'integrazione dei fenotipi per i tratti della carne bovina con i dati genomici può essere utile per stimare gli EBV dei test di prestazione, anche in una piccola razza locale. Lo studio successivo è consistito in un'ulteriore implementazione della selezione genomica nei bovini Rendena. Questa volta abbiamo studiato diversi metodi alternativi per migliorare l'accuratezza della selezione genomica nella popolazione. In particolare, è stato studiato l'impatto dell'utilizzo di solo un sottoinsieme di marcatori molecolari informativi. Abbiamo testato diverse varianti di machine learning. algoritmi di selezione per selezionare questi marcatori, ovvero LASSO, recuersive feature eliminations e Extreme Gradient Boost. Inizialmente, in un set di dati simulato, confrontiamo le prestazioni di ssGBLUP con modelli di selezione variabile con i modelli sopra menzionati. Le simulazioni differiscono in termini di numero di QTL e dimensione effettiva della popolazione.

Quindi, questi approcci sono stati implementati sul set di dati del test delle prestazioni Rendena. Inostri risultati hanno mostrato che l'accuratezza di GBLUP in popolazioni di piccole dimensioni aumenta se eseguita con SNP selezionati tramite metodi di selezione variabile sia in set di dati simulati che effettivi. Inoltre, l'uso di modelli di selezione variabile, in particolare quelli che utilizzano l'algoritmo XGboost, nel set di dati effettivo non ha avuto alcun impatto sulla distorsionee sulla dispersione dei valori riproduttivi stimati.

Nell'ultima parte di questa tesi è stato condotto uno studio sul corredo genetico delle razze locali, effettuando un'analisi GWAS sulla razza bovina Rendena. Tuttavia, poiché esistevano diverse fonti di informazione per Rendena (animali con e senza genotipizzazione o fenotipo), era necessario un metodo efficiente che combinasse ciascuna informazione. A questo scopo, single-stepGWAS (ssGWAS) si promette una buona strategia. Tuttavia, la sua capacità di tenere conto della struttura della popolazione non è stata esplorata. Abbiamo studiato l'equivalenza tra ssGWAS, efficiente associazione di modelli misti accelerati (EMMAX) e miglior previsione lineare imparziale genomica GWAS (GBLUP-GWAS) e come differiscono dall'analisi a singolo SNP senza correzione per la struttura della popolazione (SSA-NoCor) . Sono stati quindi costruiti set di dati simulati e sono state prodotte popolazioni strutturate che imitavano le popolazioni di pesci, bovini da carne e bovini da latte con rispettivamente 1.040, 5.525 e 1.400 individui genotipizzati. Sono state anche simulate popolazioni più grandi che avevano fino a 10 volte più animali genotipizzati. Nei bovini da latte i fenotipi delle figlie sono stati proiettati in tori genotipizzati (cioè prove de-regredite) prima di applicare EMMAX e SSA-NoCor. Sebbene SSA-NoCor avesse il maggior numero di SNP veri positivi tra i quattro metodi, il numero di falsi negativi era due-cinque volte quello dei veri positivi. Èinteressante notare che GBLUP-GWAS ed EMMAX avevano un numero simile di veri positivi, leggermente inferiore rispetto a ssGWAS, sebbene la differenza non fosse significativa. Dopo la convalida delle prestazioni GWAS a fase singola, l'equazione è stata utilizzata nel set di dati effettivo di Rendena. GWAS è stata quindi eseguita un'analisi post-GWAS per peso corporeo (BW), guadagno medio giornaliero (ADG), carcassa della carcassa (CF) e percentuale di medicazione (DP) in 1.690 individui. Inoltre, abbiamo considerato due dei fenotipi target (BW e ADG) in momenti diversi della vita degli individui, un aspetto potenzialmente importante nello studio dell'architettura genetica dei tratti. Abbiamo identificato 8 SNP significativi e 47 associati in modo suggestivo, situati in 14 cromosomi autosomici (BTA). Tra i segnali più forti, 3 SNP significativi e 16 suggestivi erano associati all'ADG e si trovavano su BTA10 (50–60 Mb),

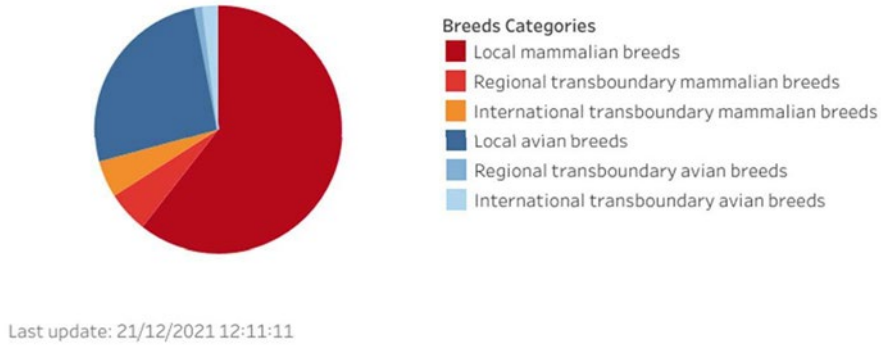
mentre l'hotspot associato a CF e DP era su BTA18 (55–62 MB). Tra gli SNP significativi alcuni sono stati mappati all'interno dei geni, come SLC12A1, CGNL1, PRTG (ADG), LOC513941 (CF), NLRP2 (CF e DP), CDC155 (DP). In conclusione, sebbene il miglioramento delle razze locali svolga un ruolo secondario, i risultati prodotti in questa tesi sembrano adatti ai sistemi di allevamento anche in piccole popolazioni locali, in termini di piani di selezione/accuratezza ma anche nella valorizzazione del loro patrimonio genomico.

3. INTRODUCTION

IMPORTANCE OF LOCAL BREEDS

Livestock farming, while being one of the drivers that marginally contribute to climate change (14.5% of GHG emission source: https://www.fao.org/), is at the same time one of the activities that is and will be most impacted by environmental changes (Mastrangelo et al., 2014). The negative effect on food production and distribution, the increase in animal diseases or other effects on feed supply are some of the consequences caused by climate change (Hoffmann, 2010). For example, a critical situation is identified in Northern Italy, an area historically devoted to diary production, where the reduction in rainfall and the increase in temperatures have negative impacts on the ability of farms to produce feed and milk, with expected future economic losses (Vitali et al., 2019). For that this reason, it is evident that the current livestock production system based on few and highly specialized species/breeds, will not be able to guarantee food security in the future (Boudalia et al., 2020). This becomes even more critical, if in addition to climate change, an increased future human demand for food is considered (FAO, 2003). For this perspective, preserve livestock biodiversity, especially across breed genetic diversity, is becoming essential (Hoffmann, 2010). Local breeds are the main contributor to a crossbreed animal genetic biodiversity, as reported on Figure 1 (https://www.fao.org/dad-is).

Figure 1. Number of livestock breeds obtained from Domestic Animal Diversity Information System(DAD-IS).



Although native breeds were once a key to ensure food security in specific local/rural areas andsupporting their economies, in the near future, these breeds will become essential to ensure the global food security (Hoffmann, 2013). Indeed, such breeds often present unique characteristics that allow them to adapt to different conditions (Krupová et al., 2016; Sutera et al., 2021), which in turn implies a better response to environmental changes or new challenges due to these changes (Biscarini et al., 2015, Mancin et al 2021). Thus, these non-specialized breeds represent an unexploited source of diversity for the animal breeding sector since they have not been under excessive specialization. Recently, thanks to the no longer prohibitive costs of genotyping with single nucleotide polymorphisms (SNP; Blasco and Toro, 2014), many studies focused on the characterization of genomic assets even in these breeds (Sechi et al., 2007; Bertolini et al., 2018; Yin and König, 2019). Results identified in these studies further confirmed what expected, i.e., i) the wide genetic variability carried by local breeds; ii) and the different genomic asset in terms of genes and biological pathways involved. For example, (Ben-Jemaa et al., 2021), showed that natural selection (i.e., selection for adaptability in a specific environments), shaped several immunity genes, not present in most cosmopolite breeds, involved in both innate and adaptive response to the pathogens. Furthermore, in some cases even the candidate genes associated with the main productive traits seem to be targets of natural or semi-artificial selection, which shows a differential number of genes and a greater complexity of the pathways involved for these traits (Mancin et al., 2022). Regarding the large genetic variability carried by these breeds, a striking example can be identified in Senczuk et al.(2020), where authors indicates that although many of the local Italian breeds analyzed are endangered, they still retain a significant amount of genetic variation compared to some specialized breeds (i.e. Holstein and Brow Swiss).

Furthermore, preserving the genetic biodiversity of livestock can be fundamental for contrasting new undue phenomena caused by climate change such as the increase in the circulation of potentially pandemic viruses (Fan et al., 2020), SARS-CoV-2, among all. Indeed, there is a clear link between intensification and specialization of the worldwide livestock sector and other recent zoonotic pandemics, e.g., influenza viruses (Gandini and Hiemstra, 2021). This is why livestock has an important component aimed at a "One Health" perspective to control viral infections affecting both animals and humans. For instance, in Bovo et al. (2021), it has been demonstrated that pig sector can be a "*One Healt*" approach against coronavirus through the inclusion of less specialized breeds.

In addition to the new role assigned to native breeds as a “gene pool” to ensure future food production in an ever-changing environment, these breeds have always played a key role in supporting the local ecosystem and economy (Hiemstra et al., 2010). They can generate immaterial socio-cultural or ecosystem benefits. Indeed, local breeds are generally raised in an "open" livestock system in which the production is based on a resource-driven activity linked to the circumscribed ecosystems, while production for the cosmopolitan breed is a question-driven activity, disconnected from the surrounding environment (FAO, 2003; Marsoner et al., 2018). For these reasons native breeds provide more local ecosystem services and cultural values, commonly called “externalities”, compared to specialized breeds (Leroy et al., 2018). The externalities provided by local cattle breeds can be mainly grouped in i) cultural services, and ii) ecosystem services. The ecosystem services can in turn be divided in supply services and regulatory services. Supply services include food-feed, fiber, genes, while the regulatory ones include nutrient cyclin, as disease control (Ovaska and Soini, 2017). Cultural services, on the other hand, consist in preservation of heritage and uniqueness through local gastronomy and landscapes, but also through rural services, such as agritourism and rehabilitation. (Ovaska and Soini, 2017).

Despite this, in the last century (from 1950 to 1980) unprecedented deterioration of livestock genetic diversity has taken place (FAO, 2007). The main driver was the interest in increasing farm profitability through the augmentation of output production (milk and beef). This geared towards the use of specialized/high-yielding breeds with a consequently a decline in the use of multi- purpose/local cattle breed (Gandini et al., 2010). Other secondary drivers are: the unbalanced assessments, the genetic introgression of other breeds, the lack of market and public incentives, the excessive use of the same sires, or for developing countries natural disasters or political instability (Bett et al., 2013). Fashion-driven factors also play an important role in the abandonment of local breeds, since the breeders themselves felt the pressure to switch from native breeds to "modern" and "more efficient" specialized breeds (Hiemstra and Gandini, 2010). However, this decline did not follow the same patterns in all countries: for example a milder decline was seen in the countries of the Mediterranean belt (Spain, Grece and Italy), while in the Nord Europe countries the local breeds were almost replaced with cosmopolitan breeds (Hiemstra and Gandini, 2010). For example, in Spain, the percentage of local breeds decreased form 74% in 1995 to 26% in 1986 (Gómez et al., 1997). In Finland a more intensive decline was observed: the Finncattle breed rapidly declined from approximately

500,000 animals 1950 to only few thousand today (https://faba.fi/en/history/). The most effective defense to counter the progressive abandonment of local breeds consists in increasing the market value of local breed farms (Gandini et al., 2007; Hiemstra et al., 2010; Biscarini et al., 2015). This could be done through public subsidies or, more effectively, by raising the awareness of farmers. Indeed, increasing the competitiveness of these breeds can be done i) by enhancing these breeds in low input-output systems (i.e., alpine pastures) ii) by considering livestock a multifunctional activity, capable of producing both food and other services that are not easily tangible (Gandini et al., 2010). An example of the second point is the enhancement of local and unique products derived from these breeds (FAO, 2007), where in addition to the intrinsic products values, an added value is guaranteed by the uniqueness of the products given by the simultaneous connection between food, the breed, the environment and the historical heritage of the place. A striking example is what happened to the Reggiana cattle, a native breed of northern Italy. Thanks to the creation of high-quality single-breed cheeses (*Parmigiano Reggiano delle Vacche Rosse*, founded in 1990), the population progressively grew from 800 heads in 1980 to 3,000 heads nowadays. This was guarantee by high profitability of Parmigiano *Reggiano delle Vacche Rosse*, that ensured over time a sustained price for milk that compensated the lower productivity of the breed (Gandini et al., 2007). Although the increase in the market values of these breeds is the key factor in the perspective of conservation, albeit secondary, it is given by adequate breeding programs, which are sometime poorly implemented in the local breeds (Biscarini et al., 2015). Although, most technologies and equation have been implemented for high-input and larger breeds are therefore applied more effectively in that context, their application in native breeds is not prevented and can bring considerable advantages. The uses of this technologies can be more successfully if they are utilized in close cooperation with farmer and breeding associations (Biscarini et al., 2015). In addition to an increased production level, these plans are needed to preserve genetic variability within each livestock population. This is essential for the long-term success of the animal husbandry industry to ensure productive and reproductive efficiency, health, survival, and overall resilience in future unforeseen environmental pressures (Mastrangelo et al., 2014). For this reason, the objectives of this thesis consisted in applying and developing *ad hoc* breeding strategies and equations, for i) broadening the knowledge of the genetic architecture of the local breed but above all ii) aiming to increase genetic progress net of maintenance of the functional and peculiar characteristics of these breeds.

THE DUALBREEDING PROJECT

Although the competitiveness of native breeds must be mainly ensured by the commercial ability of the breeder, public incentives can be supportive. More than "direct" economic aid, the "indirect" ones are useful. Public incentives that stimulate the technological progress or that aim at a reorganization of the local breed sector are therefore welcome. The DUALBREDING project is an example of this indirect public intervention to support local dual-purpose breeds. It is financed by the European Agricultural Fund for Rural Development (PSRN 2014-2020) through the National Rural Development Program, specifically by sub-measure 10.2. The National Rural Development Program (PSRN) is an instrument developed by the Italian Ministry of Agricultural, Food and Forestry Policies (MIPAAF) and co-financed by the European Agricultural Fund for Rural Development (EAFRD; Reg. (EU) n.1305 / 2013). The 2014-2020 PSRN to finance these sub-measures has a total public funding of 2.14 billion euros approved by the European Commission with decision (C2015) 8312 of 20/11/2015. The PSRN 2014-2020 pays attention to sustainability and supports rural areas with various objectives:

1) Protect and safeguard the environment and animal biodiversity
2) Investing in irrigation resources by promoting investments to facilitate water saving
3) Promote the use of risk management tools such as mutuality funds and income stabilization

As aforementioned, DUALBREEDING was financed by sub-measure 10.2: "Support for the sustainable conservation, use and development of genetic resources in agriculture". This sub-measure is aimed at the conservation of the genetic heritage and the maintenance of animal genetic variability and the safeguard of biodiversity. Sub-measure 10.2 also indirectly includes the sustainable use of farm animal biodiversity, as well as the preservation, restoration, and enhancement of connected ecosystems. The most significant application of sub-measure 10.2 was the creation of 9 livestock sectors: dairy cattle, beef cattle, dual-purpose cattle, buffaloes, sheep/goats, pigs, burrows, equine, and poultry. The DUALBREEDING is the specific measure for dual purpose, mostly local breeds.

The project involves 5 main national breeders associations, i.e., the Italian Simmental breed (Pezzata Rossa, ANAPRI), the Alpine Grey (Grigio Alpina, ANAGA), the Rendena (ANARE), the Reggiana (ANABORARE) and the Valdostana breeds (ANABORAVA).. The

DUALBREEDING also involves other 11 autochthonous small and less diffused dual-purpose breeds distributed throughout the national territory, i.e., Pinzgauer, Modicana, Cinisara, Pezzata Rossa D'Oropa, Pustertaler Sprinzen/Barà, Modenese/Bianca Val Padana, Burlina, Agerolese, Cabannina, Varzese-Ottonese Tortonese, and Garfagnina breeds, as described in Table 1. As respect the main 5 breeders associations, these breeds do not present a selection, but a conservation program. A total contribution of € 7,920,298.79 have been granted by a Ministerial decreet (Decreto Minesteriale; no. 7388) released on 23rd of February 2018, equivalent to 90% of the Admitted Expenditure of € 8,800,331.99 The DUALBREDING object is based on milestones. The first is the biodiversity or genetic variability, the specifically involves reduction and constant monitoring of inbreeding in populations. The second milestone is the progress of cattle husbandry toward the environmental sustainability, that can be achieved directly by increasing feed efficiency to reduce emission, but also indirectly through the increase of cows' longevity or placing new selection criteria that emphasize the double attitude, since dual-purpose animals have a lower environmental impact due to the co-production of milk and meat in the same process (Kaptijn, 2016). The increase in disease resistance is another milestone of the project. In fact, reducing the occurrence of diseases it may increase production efficiency and at same time reduce the costs for farmers, and to improve the conditions of animal welfare. One last point, transversal to all the others, is the so-called "Open data", that is the usability of the information collected with the project actions by the users.

Table 1 Table represented the list of breeds involved in the DUALBREEDING project and their characteristics**.**

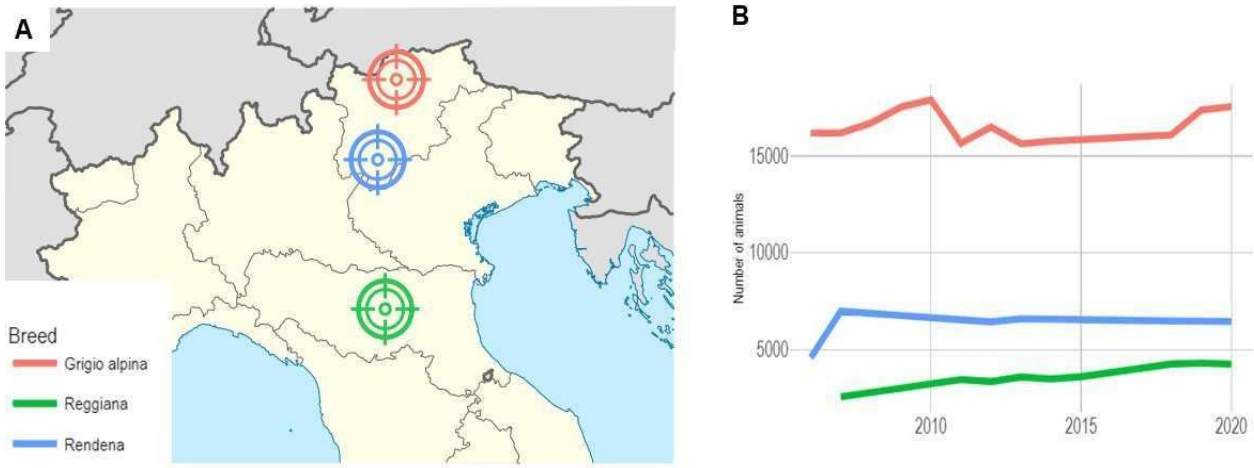
Breed	Type[1]	Competent body	Heads	Herds	Inbreeding
Pezzata Rossa Italiana	LG	ANAPRI	64,544	5,163	1.3
Valdostana PR,PN,Castana*	LG	ANABORAVA	19,500	1,322	2.7-1.5-2.2
Grigio Alpina	LG	ANAGA	7,930	1,258	2.2
Rendena	LG	ANARE	3,985	199	5.5
Reggiana	LG	ANABORARE	2,408	145	3.7
Pinzgauer	LG	AIA	1,308	222	2.9
Modicana	RA	AIA	1,825	147	2.2
Cinisara	RA	AIA	1,638	134	3.0
Pezzata-Rossa-D'Oropa	RA	AIA	2,039	120	3.9
Pustertaler Sprinzen/Barà	RA	AIA	286	43	3.7
Modenese/Bianca Val Padana	RA	AIA	451	40	2.3
Burlina	RA	AIA	426	23	3.6
Agerolese	RA	AIA	166	34	1.0
Cabannina	RA	AIA	100	17	2.9
Varzese-Ottonese-Tortonese	RA	AIA	34	5	2.6
Garfagnina	RA	AIA	0	0	1.0

*Inbreeding was distinguish for the 3 different Valodostana cattle populations, (PR: pezzata rossa, PN: pezzata nera and Castana)
[1]Type of selection, LG means that genetic improvement has carried on alongside conservation; while RA means only conservation propose (LG: Libro Genealogico; GA registro anagrafico)

LOCAL BREED CONSIDERED IN THE PRESENT STUDY

Three different local breeds were considered in current study: the Rendena, the Alpine Grey (Grigio Alpina), and the Reggiana breeds. The origin and consistency of these breeds are shown in Figure 2. Although these breeds have many similarities, they show three different approaches to the valorization process. For example, the Alpine Grey breed is the quintessential alpine breed, where a strong link between environments and breeding is considered. The Reggiana, however, represented the symbol of the valorization of autochthonous breeds through the food product-breed association. On the other hand, the Rendena breed is particularly appreciated by breeders for its "rusticity" and itsability to adapt in different environment combined with good milk and meat production.

Figure 2. Origin (**A**) and total number of animals distributed per year (from 2005 to 2020) (**B**) for the three breeds considered int that study Data extracted form (www.fao.org/dad-is/; update: 26 December 2021) and plotted in R (R Core Team, 2017).



Rendena

History

The Rendena is a native breed originating from the Alps, precisely from the homonymous valley located in the Trento province. The origin of this breed dates back to about 1700, where due to the rinderpest that struck the Rendena valley, breeders began to import cows from Switzerland and to cross them with local Brown Swiss (Bonsembiante et al., 1988). The imported animals were chosen by Rendena breeders for a certain affinity with the characteristics of native Trentino cattle populations to integrate them harmoniously. Then in the 1800 the new populations have been rapidly expanded form the original areas to South Tyrol, then in Veneto and in Lombardy. The Rendena population reach its peak in 1900, where more than 200,000 cows were present. However, after the first word-war, like in many other local breeds, the number of animals dramatically declined to few thousand head (Bonsembinate et al., 1988). This was determined by the new policies of the fascist regime (Serpieri and Mortara, 1934), which encouraged the abandonment of local breeds in favour of the more specialized ones such as the Brown Swiss (Bruna Alpina). However, after the Second World War the agricultural policies changed again and in 1947 a breeders association (*Associazione Nazionale Allevatori di Razza Rendena*) was created, although an effective herd book was created in 1976. Only in 1981 the data collection of productive, reproductive, morphological, and genealogical information was started, and consequently the first genetic indices were implemented in the middle of eighties. (Mantovani et al., 1997; Del Bo et al., 2001; Mazza and Mantovani, 2012).

Current diffusion and type of farming

According to the FAO (www.fao.org/dad-is/; update: 26 December 2021) the population of Rendena is considered at risk. However, in the past 10 years, the number of animals has remained stable, with a total population ranging between 5,000-7,000 per year (www.fao.org/dad-is/; update: 26 December 2021). Populations in 2020 was about 6.512 animals register at the herd-book, of which 4.543 cows. Animals are reared in 199 farms mainly distributed in Trentino Alto Adige and Veneto, specifically in the provinces of Trento, Padova, Vicenza, and Verona in the North-east of Italy (Bittante et al., 1993). The farms are of small-medium size, (about 29 animals/farm) locatedboth in mountain and in plain.

Rendena is reared in two main different way characterized by different management and feeding strategies. The first husbandry method is more connected with the traditional farming systems, in which animals are tied in stalls and fed with hay and concentrates, but with a large diffusion of the grazing practice on alpine pasture of the entire flock during the summer season; this form is obviously largely diffused in the farms located in the mountain area of origin. An opposite situation occurs in the plains, (Pò Valley, Veneto Region) in which a more intensive farming system is applied (i.e., a feeding system based on corn silage and with summer pasture being common mainly for replacement heifers. Rendena cattle put on grazing in alpine pasture usually spend about four months on mountains (from early June to late September) of both Val Rendena and in the Altopiano di Asiago (ANARE, 2017). Alpine pasture involves almost all the cows reared in Trentino and more than 50% of those reared in the Veneto region. As effect of the pasture, this breed maintains the seasonality of calving, with a maximum number of calving between October and December.

Characteristics, Appearance and Production:

Rendena cattle presented brown coat with different shades of dark brown with hairs of tuft and the dorsal line characterized by lighter dorsal stripe. The coat is almost black in males, and a white ring around the black muzzle is always present in both sexes Figure 3.

Figure 3. Rendena breed cow specimen

As other Alpine breeds, Rendena is a small size cattle (132 cm of height at withers in cows) with a good beef conformation (Forabosco et al., 2011). Rendena cattle show also

excellent fertility and longevity performance, together with a good milk production higher than the other local breeds and possess a fairly good beef conformation (Mazza et al., 2014). Specifically, primiparous cows that spend 100 days or more on high alpine pasture average 4,733 kg per lactation; the milk has3.50% fat and 3.36% protein; However, cows breed in more intensively product system shows an average milk production nearly 6,000 kg per lactation. Regarding beef production, Rendena calves may be slaughtered as milk veal, or as beef cattle at the age of 16–18 months, when they weigh 450–550 kg and yield 58–60% of good quality meat (Forabosco & Mantovani, 2011). The current selection index account the dairy and beef attitudes in the ratio 70:30% (Sartori et al., 2018) However the main characteristic of the breed is the rusticity, i.e., the ability to cope in different environment especially in harsh environments with low quality forages, as well as its suitability to graze during the summer season in the alpine high pastures (Mantovani et al. 1997). The Rendena breed also has a single breed product called "*Spressa delle Giudicarie DOP cheese*", produced and directly sold in many mountain farms in the territory of origin.

Alpine Grey - Grigio Alpina

History

The Alpine Grey-Grigio Alpina belongs to the group of Gray breeds of the Alpine arc and is probably one of the oldest breeds of the Alps. The Alpine Grey has originated nearly in 1800, where different "breeds" or strains presented in some Alpine valley have been merged into the current Alpine Grey-Grigio Alpina breed, in particular animals present in Passiria, Senales, Sarentino, Fassa and Fiemme valleys were joined to form the current breed . The characteristics of these strains remained almost unchanged over time due to the isolation of these valleys and the difficult genetic exchanges (Senczuk et al., 2020) and still they are present in animals belonging to the Alpine Grey breed. The first attempt of selection in Alpine grey dates to the beginning of the last century (1905), when the first breeding companies were founded in Trentino Alto Adige. Their role was to manage the breeding of bulls locally. This was also moved by the fear of losing the identity of the breed due to the increasingly crosses with the Brown Swiss, suggested as for the Rendena, by the autarchy policy imposed during the Fascist government era in Italy. The war events with the consequent economic crisis also had devastating repercussions on the breeding of the Alpine Gray. After the Second World War a first local associations of breeders were established; this association led in 1949 to the

foundation of the Federation of Breeders of the Alpine Grey Breed in Bolzano. In 1956 the "*Società Allevatori Grigia Alpina*" was refunded with the headquarters in Predazzo (TN). The latter joined the Breeders Provincial Federation of Trento. The National Association of the Alpine Grey Cattle Breeders was established on 19th of June 1980. After receiving legal recognition in 1985 and consequently being recognized the management of the National Herd Book, the National Association of Alpine Grey, began its own activities, setting up the Central Office in Bolzano in 1986 and developing its own selection program.

Current diffusion and type of farming

The Alpine Grey-Grigio Alpina breed is not considered at risk according to FAO (https://www.fao.org/dad-is/browse-by-country-and-species/en). The breed is traditionally diffused in Provinces of Bolzano and Trento, with small presence even in Belluno. However, in recent years some herds are emerging out of this area both in Norther and in Southern regions of Italy, such as in the provinces of Udine, Como, Torino, and Campobasso (ww.anaga.it). The estimated population is composed of about 25,000 head, and about **17,583** are registered in the Herd Book, (www.fao.org/dad-is/; update: 26 December 2021). In recent decades, the number of this population has remained almost constant, albeit with fluctuation over years. These animals are widespread in 1,788 farms located mostly in the provinces of Bolzano and Trento. The Alpine Grey-Grigio Alpina play an irreplaceable role for a multifunctional sustainable development of mountain environment (Marsoner et al., 2018). In fact, breeding is the main activity of that mountain area, and therefore represents the basic livelihood of mountain inhabitants.

Characteristics, Appearance and Production:

The Alpine Grey cattle is small-sized (120-125 cm of height at withers), it is robust breed and demonstrate excellent adaptability to unfavorable grazing environmental conditions, thanks to their agility with hard and resistant claws and an innate instinct to search for the best forage, that permit to venture to remote pastures. The breed standard foresees the following characteristics: the coat is light silver in color with darker ones on the head, neck, shoulder, hips, thighs, and limbs; the skin is fine and soft; the eyes are large and bright; the horns are fine, white at the base, black at the tip, directed forward upwards and with diverging tips; the tail is very thin, long and with abundant bow (Mantovani & Forabosco 2011), Figure 4.

Figure 4: Grigio Alpina cows breed specimen

As many locals' breeds, Alpine Grey has good characteristics in terms of longevity and fertility. Despite this, the breed has good production parameters, especially considering the environment in which it is raised, with on average milk production of 5,000 kg per lactation with also a good milk quality (3.75 % of fat and 3.44% of protein and low SCS). Alpine Gray is a double proposed breed, in fact it has good attitude to beef production, with an average daily gain of about 1.2 kg/d, and a good carcass conformation, and about 58% of dressing percentage (www.ANAGA.it; Mancin et al., 2021)

Reggiana

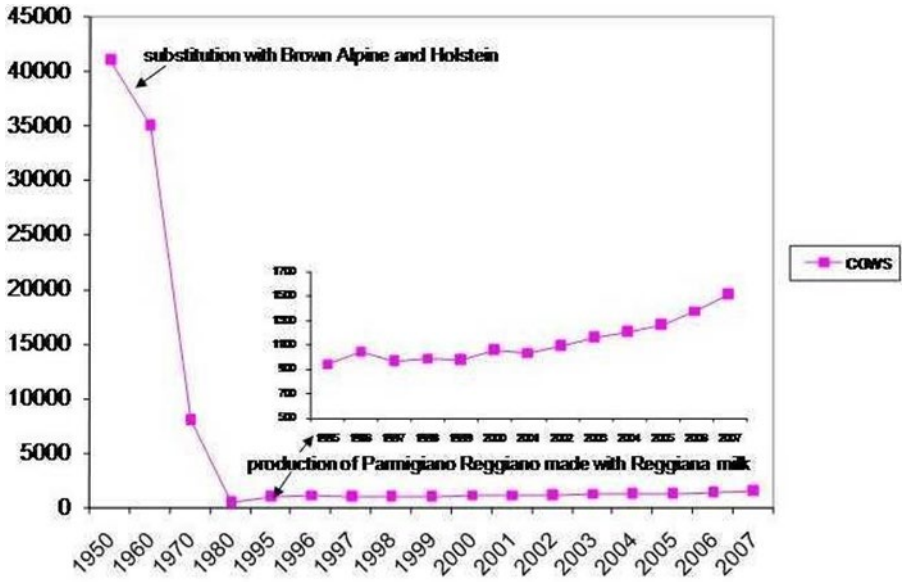
History

The origins of the Reggiana breed date back to the barbarian invasions in 568,where the invaders brought with them red cattle originating from southern Russia and the Pannonia regions, that efficiently adapted to new plain environment of the Pò Valley Even today the red color “fromentino” - like wheat grains - is still a characteristic of the cattle of Ukraine and central Russia (www.razzareggiana.it). The ancient Reggiana was a rustic and triple aptitude breed with good milk production. The cheese produced by Reggiana was the precursor of the current Parmigiano Reggiano. Around the 9th century the monks reported the presence of Reggiana cattle in Parma and Reggio Emilia. The breed at that time was a main player in the agricultural and livestock context of the area. Indeed, Reggiana was so widespread in central Italy that Renaissance painters constantly illustrate the red ox in the frescoes of the Nativity, Figure 5.

Figure 5. Natività - Andrea de Litio, 1460 - Cathedral of Atri (TE). Detail of the cycle of frescoes, the largest in Abruzzo and one of the largest in central-southern Italy by the Abruzzese painter Andrea de Litio (Lecce nei Marsi, 1420 - Atri, 1490)

Reggiana was also presented at the Vienna Expo in 1873, as most representative cattle breeds of the Pò Valley **(**https://www.regionalcattlebreeds.eu/**).** The breed reached its peak in 1954 with 139,695 heads. However, post-war Italian agricultural policy, aimed at increasing national agricultural production, led to replace/cross local Reggiana cows with more specialized cattle breeds (Serpieri and Mortara, 1934). Like many other local breeds, there has been a decline in animal numbers since the sixties, reaching less than 1000 cows in the 1980s. Fortunately, a slight but gradual increase in the number of animals has been observed since the 1990s, because of the creation of the "*Parmigiano Reggiano delle Vacche Rosse*" consortium which encouraged the breeding of the Reggiana due to the higher than average price of the milk(Figure 6).

Figure 6 Trends of Reggiana breeds population**,** the slight increase after the creation of the consortium *Parmigiano Reggiano delle Vacche Rosse* in 1995 is highlighted, figure was extracted form **(**https://www.regionalcattlebreeds.eu/**).**



The National Association of Reggiana Breeders (ANABoRaRe) was founded in 1962, but only in 1996 the herd book recording was started. Furthermore, the association has the task of enhancing the products deriving from the Reggiana breed, in fact ANABoRaRe retains the brand "*Parmigiano Reggiano delle Vacche Rosse*" (http://dualbreeding.com/it/associazioni/anaborare).

Current diffusion and type of farming

FAO reported the Reggiana populations at risk, in endangered situations. However, as mentioned above, the size of populations has been steadily increasing over the past two decades, although today the population is composed by 3,896 head (registered in the Herd Book), of which 2,409 lactating cows (www.fao.org/dad-is/; update: 26 December 2021). The animals are reared in about 50 herds. The average number of animals per herd (48) is much greater than that found in the other local breeds belonging to the DUALBREEIDNG project. The largest number of farms (95%) are in the provinces of Reggio Emilia, Parma, and Modena. The Reggiana breeding system is more like that of cosmopolitan breeds where animals are kept in free stalls with little or no presence of pasture

Characteristics, Appearance and Production

Reggiana presents a “fromentino” red coat colour (like the colour of wheat kernels) varying between light and darker Formentino. It may present lighter colour e in the internal and inferior areas of the limbs, around the eyes, and around the pink snout (https://www.razzareggiana.it), Figure 7.

Figure 7. Reggiana cows breed specimen

Compared to the other previous two breeds, the Reggiana is structurally slender (140-145 cm of height at withers in cows), with a structure closer to the "dairy type" as respect to the dual-purpose one. However, Reggiana has been officially recognize as dual cattle breeds (https://www.regionalcattlebreeds.eu/). For these reasons, future selection plans are necessary to improve, or at least to preserve, the dual-purpose attitude of Reggiana (Mantovani & Fontanesi, personal communication). The average milk production in 305 days of lactation is 5,557 kg (3.45% protein; 3.54% fat) (https://www.consorziovaccherosse.it)

EVOLUTIONS OF ANIMAL BREEDING

Animal breeding is the discipline that aims to maximize the genetic merit over time by selecting/mating the" best "animals; for doing that it is necessary: i) define what the merit is and ii) how to estimate that merit (Céron-Rojas and Crossa, 2018). The modern animal breeding is based on theory of random effects and correlations between relatives developed in (Fisher, 1919); and thanks to the earlier discoveries as segregation of traits and Mendelian inheritance. From here, two fundamental theories arose: i) the aggregate selection index (*how merit is defined*) ii) Mixed Models Equations (MME) and Best linear unbiased predictor (BLUP) (*how merit is estimated*).

The aggregate selection index (Hazel, 1943), has been used both for estimating aggregate genetic merit and predicting response to selection. However, nowadays the aggregate selection index is only used to design optimum breeding program as it presented some disadvantages in genetic evaluations (Satoh et al., 2000). The selection index is common procedure to select many traits at once. It is based on the selection of an unobservable variable through an observable variable. The unobservable variable is the genetic merit ($H = w'g$), where *w* is the vector of economic weight and *g* is the vector of true breeding values. Therefore, the observable variable is ($I = b'y$) where *y* is the vector of phenotype, and *b* is the vector that maximized the correlation between I and H. Once $cor(H, I)$ is maximized we obtained $b = P^{-1}G'a$. After, that b is integrated in function I to rank the animals.

Henderson 1948 and then formalized in (Henderson, 1975) has developed a straightforward approach to predict the genetic called mixed model equations (MME). MME allow to estimate at once environmental effect ($\hat{b}$, systematic effect) and genetic values of the animals ($\hat{u}$, random effect). MME were derived by maximizing the joint density for the two effects y and u, after some derivation and matrix handling, we obtained (proof 1):

$$\begin{bmatrix} X'R^{-1}\,\mathrm{X} & X'R^{-1}\mathrm{Z} \\ \mathrm{Z}'R^{-1}\,\mathrm{X} & \mathrm{Z}'R^{-1}\mathrm{Z}+G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ \mathrm{Z}'R^{-1}y \end{bmatrix} \quad (1)$$

where **y** is a vector of observations, **X** is a known matrix of incidence connecting phenotype with fixed effect, **Z** is a known matrix of incidence connecting phenotype with random effect, **b** is the vector of unknown systematic effect, environmental values, **a** is the vector of unknown random effect, animal breeding values, **e** the vector of residuals. The

momentum of models give to **u** and **e** are equal to $\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \begin{matrix} G & 0 \\ 0 & R \end{matrix}$. Since residuals are uniform distributed variances $V = ZGZ' + R$.

Henderson demonstrated that the second part of the equations is the equivalent to the BLUP of genetic effect: $\hat{u} = GZ'V^{-1}(Y - X\hat{\beta})$. Then, when $\hat{u}$ is substituted in the first part of equations, after some matrix handling $\hat{b}$ is equivalent the Best Linear Unbiased Estimator (BLUE) $\hat{b} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$. Note that MME in (Henderson, 1975)was an efficient method (respect to BLUP and BLUE equation shown above) since does not require V^{-1}, that at the time was computationally costly. The animal model proposed in (Henderson, 1975), was an arrangement of (1), where $\mathbf{G} = \boldsymbol{A} \otimes \sigma_a^2$. **A** is the covariance structure, that were built form pedigree relationship and σ_a^2 is the additive genetic variance. Motivation behind **A** is based on the study of Fisher 1919 in which the similarity between relative has been hypothesized, while $\hat{\mu}$ is considered "random" due to Inheritance or Mendelian error studies.

When $R = I \otimes \sigma_e^2$ (homogenous residual) models (1) can be arranged in:

$$\begin{bmatrix} X'\,\mathrm{X} & X'\mathrm{Z} \\ \mathrm{Z}'\,\mathrm{X} & \mathrm{Z}'\mathrm{Z} + A^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ \mathrm{Z}'y \end{bmatrix} \qquad (2)$$

However, MMEs were not immediately routinely implemented for two reasons: i) difficulties to estimate the variance of the random effect ii) the inversion of **A** was still too computationally demanding at the time.

Regarding variance components, (Henderson, 1982) developed three different methos. These methods have the same procedure:(i) calculation of the average squares of some kind, (ii) getting their expectations and (iii) solving linear equations in the components of the unknown variance, derived from by equating the calculated mean squares to their expectations. However, this approach was too far to be straightforward, especially in unbalance situations. Maximum Likelihood (ML) was also adopted, but it led to a is biased downwards, i.e. it underestimates the true variance. The problem was resolved by Patterson & Thompson (1971), using the Restricted Maximum Likelihood estimator (REML). In simplistic terms the REML considers systematic effect and the mean (not the variance as for ML) as disturbance parameter, that must be removed from the equation. Regarding the second problem, the computational demand of inverting A, it was brilliantly solved by (Henderson, 1976). In this

study Henderson demonstrated that A^{-1} can be directly construct form pedigree, without the need to construct A and then inverted it (Figure 8).

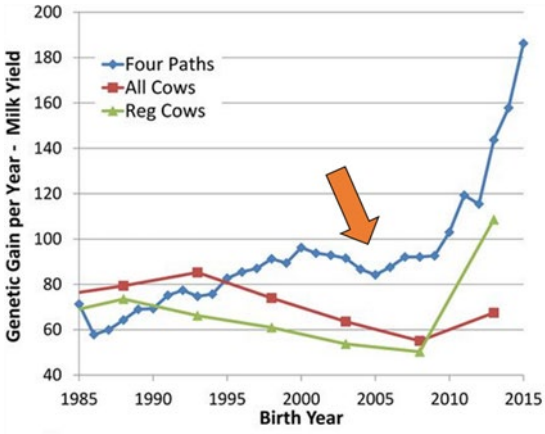
Figure 8. Simple Fortran code representing the algorithm to compute A inverse directly

```
print *,  'calculating A inverse ...'
do i=1,n
!call cpu_time(start)
  add=0.00
  if((dams_list(i) .eq. 0 ) .and.  (sire_list(i) .eq. 0)) then

 add=1.0
 A(i,i)=add
  else if ((dams_list(i) .eq. 0 ).and.  (sire_list(i) > 0)) then

 add=4.0/3.0
 A(i,i)=add
 A(i,sire_list(i))=A(i,sire_list(i)) - (add/2)
 A(sire_list(i),sire_list(i))= A(sire_list(i),sire_list(i)) + (add/4)
    A(sire_list(i),i)=A(sire_list(i),i) - (add/2)

  else if((dams_list(i) > 0) .and. (sire_list(i) .eq. 0)) then

 add=4.0/3.0
 A(i,i)=add
 A(i,dams_list(i))=A(i,dams_list(i)) - (add/2)
 A(dams_list(i),dams_list(i))=A(dams_list(i),dams_list(i))  +  (add/4)
 A(dams_list(i),i)=A(dams_list(i),i) - (add/2)


  elseif((dams_list(i) > 0) .and.  (sire_list(i) > 0)) then

add=2.0
A(i,i)=2.00
A(i,dams_list(i))=A(i,dams_list(i)) - (add/2)
 A(i,sire_list(i))=A(i,sire_list(i)) - (add/2)
 A(sire_list(i),sire_list(i))= A(sire_list(i),sire_list(i)) +  (add/4)
 A(dams_list(i),dams_list(i))=A(dams_list(i),dams_list(i))  +  (add/4)
 A(dams_list(i),sire_list(i))=A(dams_list(i),sire_list(i))  +  (add/4)
 ! ovviamente anche il rovescio
 A(dams_list(i),i)=A(dams_list(i),i) - (add/2)
 A(sire_list(i),i)=A(sire_list(i),i) - (add/2)
 A(sire_list(i),dams_list(i))=A(sire_list(i),dams_list(i))  +  (add/4)
  endif
  enddo
print *, 'done ..'
```

The development and implementation of MME was a game changer in animal breeding, with an increasing number of Henderson's MME adaptations implemented in the following years. Examples are maternal effect, social interaction models, non-additive genetic models, random regression, or reaction norm models.

The other turning point is what we now call "genomic selection" (GS), occurred around the first years of the 21st century. Indeed, GS has permitted unprecedented advances in animals breeding, involving a doubling of dairy cattle genetic progress compared with traditional BLUP(de Koning, 2016), Figure 9.

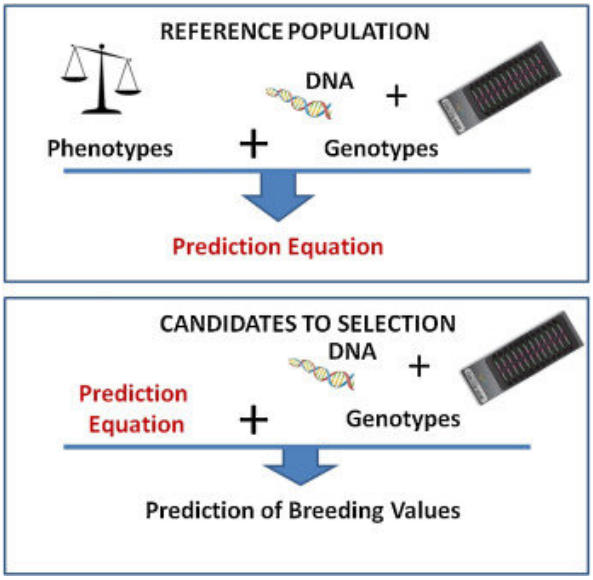
Figure 9. Changes in genetic selection differentials in US Holstein dairy cattle (orange arrow indicated when GS has been started), figure modify form (García-ruiz et al., 2016).



However, the first studies attempting to estimate genetic value from molecular data had already been conducted a decade earlier. These are based on Marked Assisted Selection (MAS) proposed by Lande and Thompson 1990. The MAS is a two-step procedure where at first the markers associated with the phenotype are identify and then selection occur increasing the frequencies of favorable allele. However, MAS found little interest since the few markers associated marginally contributed of the total genetic variance express by the phenotype (Blasco and Toro, 2014).

What we now call genomic selection (GS) was first implemented in (Meuwissen et al., 2001). The main assumption behind GS is that Single Nucleotide Polymorphisms panels (SNP) must be dense enough to be in LD with all quantitative traits loci (QTL). GS is a two-step procedure where the allelic substitution effects are estimated for all markers and then the genomic values are calculated as the summatory of these effects. The marker effect is estimated in a validation dataset (animals with both genotype and phenotype) while genomic prediction was calculated for candidate to selection (animal with genotype and not necessarily phenotype, generally young bulls), represented in Figure 10

Figure 10. Schematic representation of two-step procedure of GS (Boichard et al., 2016)



(Meuwissen et al., 2001) estimated the SNPs effect as a regression SNPs on the phenotype, in that study three models were used: BLUE, BLUP, and Bayesian models (Bayes A and B). Later, the Bayesian approaches have been extensively optimized and revised by other researchers, with many improvements related to prior's distributions, ucalled Bayesian Alphabet (Gianola, 2013). Despite the great accuracy of Bayesian model (i.e. variable selection models as Bayes B), especially in presence of small training populations these models have been scarcely implemented in real selection scheme (Habier et al., 2013). However, the most common method is the SNP-BLUP models (or SNPs ridge regression) are represented as follow:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+I\frac{\sigma_e^2}{\sigma_{a0}^2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \qquad (3)$$

Where **Z**, in this case, is the matrix representing the gene content, and $\sigma_{a0}^2$is the SNPs variances. Note since is a SNP-BLUP model $\sigma_{a0}^2$is equal for all SNPs. However, GS has two main drawback: i) for an accurate estimation of SNPs effects are required an large enough training population ii) the SNPs effect and the training population need to be update every 3-4 generation, due the change of the LD between SNPs and QTLs (Ibáñez-Escriche et al., 2014). In addition, the use of multi-trait models is far from easy to implement.

(VanRaden, 2008)reinvented again the concept GS. This study proposes to replace **A** matrix present in animal model (equation 2) with a genomic relationship matrix **G**, such as:

$$\begin{bmatrix} X'\mathrm{X} & X'\mathrm{Z} \\ \mathrm{Z}'\mathrm{X} & \mathrm{Z}'\mathrm{Z}+\mathrm{G}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ \mathrm{Z}'y \end{bmatrix} \qquad (4)$$

G is and Identical By State matrix (IBS), constructed as:

$$\mathrm{G} = \frac{\mathrm{MM}'}{2\sum p_i\,(1-p_i)} \qquad (5)$$

Where **M** is matrix conteined the gene content in (3) and p is the alleic frequencies of i^{th} locus. For that reason, these models are usually called Genomic BLUP (GBLUP).

This method is matematical equivalent of BLUP of SNP regression (equation 3). GBLUP made GS more flexible and rutinable form some reason:

i) It reduced MME's dimesion ,at the time, since SNP-BLUP is *(systematic effect + number of SNP)*^{2} while GBLUP is *(systematic effect + number of animals)*^{2}*.*

ii) GBLUP has less convergence problem respect of SNP-BLUP

iii) Relex the concept of training and test population

iv) Above all it has made the transition from BLUP to GS much easier and understable, since it allowed conceptual comparisons between pedigree-based and genome-based predictions (Misztal et al., 2020).

However, both methods (GBLUP and SNP-BLUP) are comonly called multi-step methods. Multi-steps means that some pre and post genomic analyzes are needed. The first step a genetic evalution based on pedigree information has carried on, and then pseudo-phenotype are calculated. Second step consist to proceed with genomic estimation. Finally, the genomic and genetic selection index are combined by removing the parent average. Multisteps are necessary to combine the different sources of information (animals with and without genotype and animals with and without phenotype) (Misztal et al., 2020). However, multi-step methods are not easy to implement as many operations are required and especially in some cases the pseudo-phenotypes are trivial and accuracy is by definition approximate(Masuda et al., 2018).

On this point (Legarra et al., 2009) proposed an new apprach of GS called single step GBLUP (ssGBLUP). The MME presented in the ssGBLUP are equal to (1 or 4), but the covariances structure is rappresented by **H** matrix. The idea behind **H** matrix was based on previus study of (Gengler et al., 2007) in which pedigree and genomic relashionship are jointly distributed. Infact, (Legarra et al., 2009) considered that that pedigree (u_1) and genomic (u_2) information are multinormal distributed, as:

$$H = \begin{bmatrix} A_{11} - A_{12}{A_{22}}^{-1}A_{21} + A_{12}A_{22}^{-1}GA_{22}^{-1}A_{21} & A_{21}A_{22}^{-1}G \\ G{A_{22}}^{-1}A_{21} & G \end{bmatrix} \qquad (6)$$

Derivation of (8) was intuitively described in (Lourenco et al., 2020).

However,ssGBLUP was not easy to implement becouse **H** inverse was too compuational demands. For this reason, as append to BLUP in (Henderson, 1976) ssGBLUP become felasible when inverse of relashionship matrix was directly compute. Indeed (Aguilar et al., 2010) directly compute **H** inverse as:

$$\mathrm{H}^{-1} = \mathrm{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathrm{G}^{-1} - \mathrm{A}_{22}^{-1} \end{bmatrix} \qquad (7)$$

However, over the years the number of genotyped animals has grown more and more, which has made it difficult to make ssGBLUP computationally flexible (since cost was *animals*^{3}). (Misztal, 2016) proposed the APY algorithm to simplify the construction of G, based on the main eigenvalues express by **G** matrix.

An alternative approach commonly called single step SNP-BLUP (ssSNP-BLUP) was proposed by (Fernando et al., 2014), SNP-BLUP is equivalent to ssGBLUP.The idea behind ssSNP-BLUP is based on avoid the computational cost in ssGBLUP of “imputing” genotypes for non-genotyped animals. Indeed, ssGBLUP based on less demanding SNP-based prediction or “imputation”. The other advantages of ssSNP-BLUP respect to ssGBLUP consist of frequencies of the SNP alleles in the founders are not required and that different priorities for the distribution of the SNP can be easily integrated. Additionally, SNPs effect are easier to use in interim prediction, since it doesn’t required a second step as ssGBLUP (Taskinen et al., 2017).

However, ssSNP-BLUP exhibited less flexibility in indirect genetic effect models (such as maternal models) or in multi-trait models. Other limitation of ssSNP-BLUP, it is the poor

convergence of when preconditioned conjugate gradient (PCG), but it was partially solved by in (Vandenplas et al., 2021), using a second levels preconditioner PCG (PCG is the most common method for solving MME of genetic evaluations).

Nowadays there are many promising new perspectives in animal breeding. Omics technologies are into this category. Despite their scientific interest, the main advantage of this consists of predict a phenotype at low-cost. Nevertheless, their effectiveness seems to be limited only to certain phenotype. Machine learning and deep learning algorithms are also seen as a new perspective. However so far they do not seem to have caught on in animal genetics due to the small increase in accuracy over "traditional methods"(Abdollahi-Arpanahi et al., 2020). Lastly the genome editing apperead as promising strategy. However, despite the initial hype, it could be an auxiliary tool for increasing genetic progress for disease-related traits, but many technical limitations are still present (Tait-Burkard et al., 2018). However, few models and technologies have been developed ad-hoc for small populations/locals,due the marginal economic interest of thosse breeds. For those reasons, in the present work we aimed to applied new and existed equation in a local breeding framework, and evaluated the impact of them.

Bibliography:

Henderson CR. (1949). Estimation of changes in herd environment. J Dairy Sci. (Abstract) 32: 706.

Bonsembiante M., G. Bittante, M. Ramanzin E C. Neri, 1988. Caratteristiche, evoluzione e miglioramento della Razza Rendena. Ed. Pragmark.

Mantovani, R., Gallo, L., Carnier, P., Cassandro, M., and Bittante, G. (1997).Vienna, Austria. The Use of a Juvenile Selection Scheme for Genetic Improvement of Small Populations: the Example of Rendena Breed, Proc. 48th EAAP Annu. Meet. 25–28 .

Abdollahi-Arpanahi, R., Gianola, D., and Peñagaricano, F. (2020). Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genet. Sel. Evol.* 52, 1–15. doi:10.1186/s12711-020-00531-z.

Lande R, Thompson R. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics. 1990 Mar;124(3):743-56. doi: 10.1093/genetics/124.3.743. PMID: 1968875; PMCID: PMC1203965.

Patterson, H. D.; Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika. 58 (3): 545. doi:10.1093/biomet/58.3.545.

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93, 743–752. doi:10.3168/jds.2009-2730.

Ben-Jemaa, S., Senczuk, G., Ciani, E., Ciampolini, R., Catillo, G., Boussaha, M., et al. (2021). Genome-Wide Analysis Reveals Selection Signatures Involved in Meat Traits and Local Adaptation in Semi-Feral Maremmana Cattle . *Front. Genet.* 12, 669. Available at: https://www.frontiersin.org/article/10.3389/fgene.2021.675569.

Bertolini, F., Schiavo, G., Galimberti, G., Bovo, S., D'Andrea, M., Gallo, M., et al. (2018). Genome-wide association studies for seven production traits highlight genomic regions useful

to dissect dry-cured ham quality and production traits in Duroc heavy pigs. *Animal* 12, 1777–1784. doi:10.1017/s1751731118000757.

Bett, R. C., Okeyo, M. A., Malmfors, B., Johansson, K., Agaba, M., Kugonza, D. R., et al. (2013). Cattle breeds: Extinction or quasi-extant? *Resources* 2, 335–357. doi:10.3390/resources2030335.

Biscarini, F., Nicolazzi, E., Alessandra, S., Boettcher, P., and Gandini, G. (2015a). Challenges and opportunities in genetic improvement of local livestock breeds. *Front. Genet.* 5, 1–16. doi:10.3389/fgene.2015.00033.

Biscarini, F., Nicolazzi, E. L., Stella, A., Boettcher, P. J., and Gandini, G. (2015b). Challenges and opportunities in genetic improvement of local livestock breeds. *Front. Genet.* 6, 33. doi:10.3389/fgene.2015.00033.

Blasco, A., and Toro, M. A. (2014). A short critical history of the application of genomics to animal breeding. *Livest. Sci.* 166, 4–9. doi:10.1016/j.livsci.2014.03.015.

Boichard, D., Ducrocq, V., Croiseau, P., and Fritz, S. (2016). Genomic selection in domestic animals: Principles, applications and perspectives. *Comptes Rendus - Biol.* 339, 274–277. doi:10.1016/j.crvi.2016.04.007.

Boudalia, S., Said, S. Ben, Tsiokos, D., Bousbia, A., Gueroui, Y., Mohamed-Brahmi, A., et al. (2020). Bovisol project: Breeding and management practices of indigenous bovine breeds: Solutions towards a sustainable future. *Sustain.* 12, 1–9. doi:10.3390/su12239891.

Bovo, S., Schiavo, G., Ribani, A., Utzeri, V. J., Taurisano, V., Ballan, M., et al. (2021). Describing variability in pig genes involved in coronavirus infections for a One Health perspective in conservation of animal genetic resources. *Sci. Rep.* 11, 1–14. doi:10.1038/s41598-021-82956-0.

Céron-Rojas, J. J., and Crossa, J. (2018). *Linear Selection Indices in Modern Plant Breeding*. Available at: https://link.springer.com/content/pdf/10.1007%2F978-3-319-91223-3.pdf.

de Koning, D. J. (2016). Meuwissen et al. On genomic selection. *Genetics* 203, 5–7. doi:10.1534/genetics.116.189795.

Fan, J. L., Da, Y., Zeng, B., Zhang, H., Liu, Z., Jia, N., et al. (2020). How do weather and climate change impact the COVID-19 pandemic? Evidence from the Chinese mainland. *Environ. Res. Lett.* 16. doi:10.1088/1748-9326/abcf76.

FAO (2003). Food Insecurity in the World. *Food Agric. Organ. United Nations*. Available at: http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E%5Cnhttp://faostat3.fao.org/.

Fernando, R. L., Dekkers, J. C. M., and Garrick, D. J. (2014). A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genet. Sel. Evol.* 46, 1–13. doi:10.1186/1297-9686-46-50.

Fisher, R. A. (1919). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinburgh* 52, 399–433. doi:10.1017/S0080456800012163.

Food and Agriculture Organization of the United Nations (FAO) (2007). *THE STATE OF FOOD AND AGRICULTURE*. ROME doi:10.18356/36f9d44d-en.

Forabosco, F., Mantovani, R., and Meneghini, B. (2011). *European and Indigenous Cattle Breeds in Italy*. Schiel & Denver Publishing Limited Available at: https://books.google.it/books?id=BiA0YAAACAAJ.

Gandini, G., Avon, L., Bohte-Wilhelmus, D., Bay, E., Colinet, F. G., Choroszy, Z., et al. (2010). Motives and values in farming local cattle breeds in Europe: a survey on 15 breeds. *Anim. Genet. Resour. génétiques Anim. genéticos Anim.* 47, 45–58. doi:10.1017/s2078633610000901.

Gandini, G., and Hiemstra, S. J. (2021). Farm animal genetic resources and the COVID-19 pandemic. *Anim. Front.* 11, 54–56. doi:10.1093/af/vfaa049.

Gandini, G., Maltecca, C., Pizzi, F., Bagnato, A., and Rizzi, R. (2007). Comparing local and commercial breeds on functional traits and profitability: The case of reggiana dairy cattle. *J. Dairy Sci.* 90, 2004–2011. doi:10.3168/jds.2006-204.

García-ruiz, A., Cole, J. B., Paul, M., Wiggans, G. R., Ruiz-lópez, F. J., and Curtis, P. (2016). Erratum: Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection (Proceedings of the National Academy of Sciences of the United States of America (2016) 113 (E3995-E4004) DOI:10.1073. *Proc. Natl. Acad. Sci. U. S. A.* 113, E4928. doi:10.1073/pnas.1611570113.

Gengler, N., Mayeres, P., and Szydlowski, M. (2007). A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal* 1, 21–28. doi:https://doi.org/10.1017/S1751731107392628.

Gianola, D. (2013). Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* 194, 573–596. doi:10.1534/genetics.113.151753.

Gómez, M., Plazaola, J. M., and Seiliez, J. P. (1997). The Betizu Cattle of the Basque country Resumen Origin , Population Numbers and Distribution Characteristics of the Breed. 1–5.

Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genomic BLUP decoded: A look into the black box of genomic prediction. *Genetics* 194, 597–607. doi:10.1534/genetics.113.152207.

Hazel, L. N. (1943). The Genetic Basis for Constructing Selection Indexes. *Genetics* 28, 476–490.

Henderson, C. R. (1975a). Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics* 31, 423–447. doi:10.2307/2529430.

Henderson, C. R. (1975b). Best Linear Unbiased Estimation and Prediction under a Selection Model Author (s): C . R . Henderson Published by : International Biometric Society Stable URL : https://www.jstor.org/stable/2529430 International Biometric Society is collaborating with J. *Biometrics* 31, 423–447.

Henderson, C. R. (1976). A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics* 32, 69. doi:10.2307/2529339.

Henderson, C. R. (1982). Analysis of Covariance in the Mixed Model: Higher-Level, Nonhomogeneous, and Random Regressions. *Biometrics* 38, 623–640. doi:10.2307/2530044.

Hiemstra, S. J., De Haas, Y., Mäki-Tanila, A., and Gandini, G. (2010). *Local cattle breeds in Europe: Development of policies and strategies for self-sustaining breeds*. doi:10.3921/978-90-8686-697-7.

Hiemstra, S. J., and Gandini, G. (2010). *Local cattle breeds in Europe*. doi:10.3920/978-90-8686-697-7.

Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Anim. Genet.* 41, 32–46. doi:10.1111/j.1365-2052.2010.02043.x.

Hoffmann, I. (2013). Adaptation to climate change--exploring the potential of locally adapted breeds. *Animal* 7 Suppl 2, 346–362. doi:10.1017/S1751731113000815.

Ibáñez-Escriche, N., Forni, S., Noguera, J. L., and Varona, L. (2014). Genomic information in pig breeding: Science meets industry needs. *Livest. Sci.* 166, 94–100. doi:10.1016/j.livsci.2014.05.020.

Kaptijn, G. (2016). Evaluation of the performance of dual-purpose cows in European pasture-based systems. *Univ. Wageningen*, 1–45.

Krupová, Z., Krupa, E., Michaličková, M., Wolfová, M., and Kasarda, R. (2016). Economic values for health and feed efficiency traits of dual-purpose cattle in marginal areas. *J. Dairy Sci.* 99, 644–656. doi:10.3168/jds.2015-9951.

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. doi:10.3168/jds.2009-2061.

Leroy, G., Hoffmann, I., From, T., Hiemstra, S. J., and Gandini, G. (2018). Perception of livestock ecosystem services in grazing areas. *Animal* 12, 2627–2638. doi:10.1017/S1751731118001027.

Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., and Misztal, I. (2020). Single-step genomic evaluations from theory to practice: using snp chips and sequence data in blupf90. *Genes (Basel).* 11, 1–32. doi:10.3390/genes11070790.

Mancin, E., Sartori, C., Guzzo, N., Tuliozi, B., and Mantovani, R. (2021). Selection Response Due to Different Combination of Antagonistic Milk, Beef, and Morphological Traits in the Alpine Grey Cattle Breed. *Animals* 11. doi:10.3390/ani11051340.

Mancin, E., Tuliozi, B., Pegolo, S., Sartori, C., and Mantovani, R. (2022). Genome Wide Association Study of Beef Traits in Local Alpine Breed Reveals the Diversity of the Pathways Involved and the Role of Time Stratification. 12, 1–22. doi:10.3389/fgene.2021.746665.

Marsoner, T., Egarter Vigl, L., Manck, F., Jaritz, G., Tappeiner, U., and Tasser, E. (2018). Indigenous livestock breeds as indicators for cultural ecosystem services: A spatial analysis within the Alpine Space. *Ecol. Indic.* 94, 55–63. doi:10.1016/j.ecolind.2017.06.046.

Mastrangelo, S., Saura, M., Tolone, M., Salces-Ortiz, J., Di Gerlando, R., Bertolini, F., et al. (2014). The genome-wide structure of two economically important indigenous sicilian cattle breeds. *J. Anim. Sci.* 92, 4833–4842. doi:10.2527/jas.2014-7898.

Masuda, Y., VanRaden, P. M., Misztal, I., and Lawlor, T. J. (2018). Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection. *J. Dairy Sci.* 101, 5194–5206. doi:10.3168/jds.2017-13310.

Mazza, S., Guzzo, N., Sartori, C., Berry, D. P., and Mantovani, R. (2014). Genetic parameters for linear type traits in the Rendena dual-purpose breed. 131, 27–35. doi:10.1111/jbg.12049.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome wide dense marker map. *Genetics* 157, 1819–1829.

Misztal, I. (2016). Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. *Genetics* 202, 401–409. doi:10.1534/genetics.115.182089.

Misztal, I., Lourenco, D., and Legarra, A. (2020). Current status of genomic evaluation. *J. Anim. Sci.* 98. doi:10.1093/jas/skaa101.

Ovaska, U., and Soini, K. (2017). Local Breeds – Rural Heritage or New Market Opportunities ? Colliding Views on the Conservation and Sustainable Use of Landraces. *Sociol. Ruralis* 57, 709–729. doi:10.1111/soru.12140.

R Core Team (2017). R: A Language and Environment for Statistical Computing.

Sartori, C., Guzzo, N., Mazza, S., and Mantovani, R. (2018). Genetic correlations among milk yield, morphology, performance test traits and somatic cells in dual-purpose Rendena breed. *Animal* 12, 906–914. doi:10.1017/S1751731117002543.

Satoh, M., Hicksh, C., Ishii, K., and Furukawa, T. (2000). Prediction of Response Values Index to Selection by Expected Selection Based on BLUP to of Breeding Family Response Supporting Pig Selection. 71, 17–25.

Sechi, T., Usai, M. G., Miari, S., Mura, L., Casu, S., and Carta, A. (2007). Identifying native animals in crossbred populations: The case of the Sardinian goat population. *Anim. Genet.* 38, 614–620. doi:10.1111/j.1365-2052.2007.01655.x.

Senczuk, G., Mastrangelo, S., Ciani, E., Battaglini, L., Cendron, F., Ciampolini, R., et al. (2020). The genetic heritage of Alpine local cattle breeds using genomic SNP data. *Genet. Sel. Evol.* 52, 1–12. doi:10.1186/s12711-020-00559-1.

Serpieri, A., and Mortara, G. (1934). POLITICA AGRARIA FASCISTA. *Ann. di Econ.* 9, 209–303. Available at: http://www.jstor.org/stable/23229929.

Sutera, A. M., Moscarelli, A., Mastrangelo, S., Sardina, M. T., Di Gerlando, R., Portolano, B., et al. (2021). Genome-Wide Association Study Identifies New Candidate Markers for Somatic Cells Score in a Local Dairy Sheep. *Front. Genet.* 12, 409. doi:10.3389/fgene.2021.643531.

Tait-Burkard, C., Doeschl-Wilson, A., McGrew, M. J., Archibald, A. L., Sang, H. M., Houston, R. D., et al. (2018). Livestock 2.0 - Genome editing for fitter, healthier, and more productive farmed animals. *Genome Biol.* 19, 1–11. doi:10.1186/s13059-018-1583-1.

Taskinen, M., Mäntysaari, E. A., and Strandén, I. (2017). Single-step SNP-BLUP with on-the-fly imputed genotypes and residual polygenic effects. *Genet. Sel. Evol.* 49, 1–15. doi:10.1186/s12711-017-0310-9.

Vandenplas, J., Calus, M. P. L., Eding, H., van Pelt, M., Bergsma, R., and Vuik, C. (2021). Convergence behavior of single-step GBLUP and SNPBLUP for different termination criteria. *Genet. Sel. Evol.* 53, 1–15. doi:10.1186/s12711-021-00626-1.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980.

Vitali, A., Segnalini, M., Esposito, S., Lacetera, N., Nardone, A., and Bernabucci, U. (2019). The changes of climate may threat the production of Grana Padano cheese: past, recent and future scenarios. *Ital. J. Anim. Sci.* 18, 922–933. doi:10.1080/1828051X.2019.1604087.

Yin, T., and König, S. (2019). Genome-wide associations and detection of potential candidate genes for direct genetic and maternal genetic effects influencing dairy cattle body weight at different ages. *Genet. Sel. Evol.* 51, 1–14. doi:10.1186/s12711-018-0444-4.

4. GENERAL AIMS

The present thesis has carried out with the intent of provided suitable methos of valorization of Italian locals breeds through a careful selection scheme and through the valorization of their genetic heritage. This project arises with collaboration between university of Padova (Department of Agronomy, Food, Natural resources, Animals and Environment) and the breeding association who took part to the DUALBREEDING project. This thesis can be divided into three main chapters. Each chapter contains a theoretical part in which the new formulas / equations are applied to simulated data while the second part consists in the application of these methods to real data. The three main research line followed were in the direction of: i) genetic selection plans ii) genomic selection iii) genome wide association study.

Genetic and plans:

The objective of this part of the study was to produce an algorithm able to process genetic and phenotypic correlations to derive economic weights to be used in the selection indices for small dual-purpose cattle, finalized to select for many antagonistic traits (milk yield and quality, beef characteristics, morphological, and functional traits) to guarantee even zero genetic progress for some traits of interest (constrains of maintenance of the genetic level in the population).

A second objective was then to develop a new selection index for Alpine Grey cattle breed in which milk, meat and functional characteristics were considered, aiming at improving the dual-purpose attitude and maintain the breed's peculiar characteristics by attributing specific economic weights at each trait. A preliminary analysis of genetic correlations among different traits was the preliminary objective of this study.

A third part was aimed at estimating genetic and phenotypic correlations among milk yield and fertility traits considering a genotype by environment (GxE) approach in the Reggiana breed, looking accuracy and bulls re-ranking respect classical single traits models were compared.

Genomic selection

The aim of studies on genomic selection were addressed at first to the analysis of performance test data belonging to the Rendena breed to evaluate the impact of genomic selection as EBVs accuracy due to use of genomic information.

A second objective was the identification of *ad-hoc* models to be used for small populations including different genomic relationships matrix constructed from selected SNPs using different variable selection algorithms. As preliminary part, this study also aimed at developing a variable selection algorithm by considering different genomic architecture available through simulated dataset

Genome wide association study

The objective of studies carried out on this research line were at first to identify strategies that would allow GWAS to be conducted on small populations like the Rendena cattle breed, with a limited number of genotyped animals. Particularly, the objective of the study was to investigate the equivalence between ssGWAS, efficient association of accelerated mixed models (EMMAX) and best impartial linear genomic prediction GWAS (GBLUP-GWAS) and to analyze how they differ from single-SNP analysis without correction for population structure (SSA-NoCor).

The final aim of this research line was the GWAS analysis on performance test traits obtained from the Rendena breed A secondary objective of the study was the analysis of two of the target phenotypes at different times in the lives of individuals, to deeply evaluate the genetic architecture of the traits.

5. TECHNICAL NOTE: ECONOMIC WEIGHTS FOR RESTRICTION OF SELECTION INDEX:

STATUS: ON SUBMISSION TO JOURNAL OF DAIRY SCIENCE

Economic weights for restriction of selection index

Enrico Mancin,Roberto Mantovani and Cristina Sartori*

INTRODUCTION

The aggregate economic selection index is commonly used to classify animals based on a combination of the estimated breeding values (EBVs) for traits of interest and weighted by the economic values of traits (Hazel, 1943). Weights of combined traits are given considering market, breeding system or functional need of each breed. However, the main drawback of including many traits in the selection index is to ensure a positive genetic progress for all these traits, as many of them are negatively correlated. For example in dual-purpose breeds beef conformation is also considered a productive trait alongside milk production (Cunningham and McClintock, 1974), but an antagonistic genetic relationship among the two aptitudes exist (Fuerst-Waltl et al., 2016; Mazza et al., 2016; Sartori et al., 2018). Additionally, not all traits require positive genetic progress, and for some traits is necessary at least to guarantee a non-worsening genetic progress over time (Miglior et al., 2005). For example, dairy cattle have some traits of interest that are not strictly related to production but to individual welfare, such as somatic cells score (SCS), fertility, longevity, or some morphological traits that not necessarly need improvement, but just their stability over time (Mancin et al., 2021). In literature, several methods to restrict to zero the genetic gain of target traits have been proposed (e.g., Kempthorne and Nordskog, 1959; Xie and Xu, 1997). The restriction to zero of the genetic progress for some traits can be advantageous in the aforementioned situations, allowing to maximize the productive characteristics at the net of conservation of functional and/or morphological characteristics. Furthermore, the restriction to zero can be particularly useful for local breeds, where the addition in the selection index of some non-economical traits is

sometime required. These are mainly phenotypes linked to some peculiar aspects of the breed, e.g., head tipicality in Alpine Grey cattle (Mancin et al., 2021). This technical note aims to analytically demonstrate how to derive economic weights to be used in selection indices to ensure a zero genetic progress of certain traits of interest. With this in mind, we analyzed (i) the classical formula of univarited genetic progress; (ii) the genetic progress in a multiple traits framework where the economic weigth for multiple traits is obtained, and (iii) the procedure to restrict some traits (i.e., ensure gentic progress equal to zero). Last, the procedure to obtain the resultant economic weights with restriction on some trais is also reported. A small example was provided using the genetic and phenotypic correlation of a dual-purpose cattle population. Where restriction on SCS and muscularity while maximizing the genetic gain for other milk and meat production traits is presented including a R-code to run all the requested steps.

MATERIAL AND METHODS

Genetic progress for one trait

The response to selection (R) represents the difference between the mean phenotypic value of the progeny of selected parents and that of the whole parental generation before selection (Falconer and Mackay, 1996). The selection response for one trait of interest has been represented in many formulas. For consistency to what will be reported later, the formula presented in Harris (1964), is:

$$R = \beta_{gy}\Delta_I \qquad (1),$$

Where β_{gy} is the linear regression coefficient of the offspring genetic value g on selected parents' phenotype (y), and Δ_I stands for the selection differential, representing the deviation of selected parents' mean from the population mean. Replacing β_{gy} with its mathematical

expression $\frac{\mathrm{Cov(g,y)}}{\sigma_y 2}$ and then grouping Δ_I/σ_y into *i*, i.e., the intensity of selection in standardized units, we obtain (Falconer and Mackay, 1996):

$$R = \frac{\mathrm{Cov(g,y)}}{\sigma_y} i \quad (2),$$

Where, under the assumption of a normal distribution of the phenotype *y*, the integration of *i* or Δ_I/σ_y provides the proportion of selected individuals (the parents).

Economic weight and genetic progress for multiple traits

In a multiple trait framework, the parents' phenotype (***y***) are represented in matrix of form, such as the phenotypic (co)variances **P** = $(\mathbf{y}^T\mathbf{y})$, and thengenetic (co)variances, i.e., **G** = $(\mathbf{g}^T\mathbf{g})$, for traits (*T*) of interest. The maximization of the genetic progress for multiple traits is based on the indirect selection for an unobservable variable (**H**), by the truncated selection of an observable variable, I (Harris, 1964). The observable variable **I**, is a linear combination of ($\mathbf{I} = \mathbf{b}'\mathbf{y}$), where $\mathbf{b}'$ is the vector of selection index coefficients for the traits of interest. Therefore, the variance of selection index **I** (σ_i) can be expressed as var(b'y) or, in matrix form, as $\mathbf{b'Pb}$. The role of vector **b** is to maximize the genetic progress through a maximization of the covariance between **H** and **I**, $\mathrm{Cov}(\mathbf{H},\mathbf{I})$ (Harris, 1964). Regression of **H** on **I** is linear for any set of **b** values. Since no economic weight are applied, **H** is equal to the vector of the true breeding values $\mathbf{H} = \mathbf{a}'$. Given that $\mathrm{Cov}(\mathbf{G},\mathbf{P}) = \mathbf{G}$. It can be demonstrated by put equal the two different selection response formula in Falconer and Mackaay (1996), that the covariance between **H** and **I,** $(\mathrm{Cov}(\mathrm{y'b},\mathrm{g'}))$, becomes $\mathbf{Gb}'$.

Considering $\mathrm{i},\mathbf{P},\mathbf{G}$ as constants (they are known values), maximizing genetic progress is equal to maximize $\mathrm{Cov}(\mathbf{H},\mathbf{I})$, that means minimizing the squared differences between **H** and **I**,

that is $(\mathrm{E}[(\mathbf{H}-\mathbf{I})^2])$ (Céron-Rojas and Crossa, 2018). Setting the partial derivative of $\mathbf{b}$ equal to zero:

$$\frac{\partial}{\partial \mathrm{b}}\mathrm{E}[(\mathbf{H}-\mathbf{I})^2] = 0 \qquad (3)$$

by changing H and I in (3), the formula becomes:

$$\frac{\partial}{\partial \mathrm{b}}(\mathbf{G}+\mathbf{b}'\mathbf{P}\mathbf{b}-2\mathbf{G}\mathbf{b}') = 0 \qquad (4)$$

Deriving the formula above we have:

$$2\mathbf{b}\mathbf{P} - 2\mathbf{G} = 0 \qquad (4a)$$

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G} \qquad (4b)$$

with **b** being the vector that maximized the genetic progress. Replacing $\mathrm{cov}(\mathbf{G},\mathbf{I})$ with $\mathbf{b}'\mathbf{G}$ and σ_i with $(\mathbf{b}'\mathbf{P}\mathbf{b})^{1/2}$ in formula 2, we obtain:

$$\mathrm{R} = \mathbf{b}'\mathbf{G}(\mathbf{b}'\mathbf{P}\mathbf{b})^{-1/2}\mathrm{i} \quad (5)$$

Then, if we considered the different economic importance of traits, it become necessary to introduce a new element in the equation (3), that is a vector (**a**) reppresenting the weights of each trait. According with Wolfová et al., 2001 vector **a** can be standardized by dividing the economic weight of the traits by the respective genetic standard deviation as $\mathbf{a_{t\_std}} = \mathbf{a_t}/\boldsymbol{\sigma}_{\mathbf{at}}$. Thus **H** can be described as a function $\mathbf{H} = \mathbf{a_t}'\mathbf{g}$, called linear aggregate genotype (Hazel, 1943), that is the function of the additive genetic values of the traits of interest, each one showing a specific economic weight. Therefore, in this form the function **H** includes the expected genetic progress for each trait. When economic weight are used, the variance of **H**

becomes $\mathrm{cov}(\mathbf{a'g}, \mathbf{a'g})$ that is equal to $\mathbf{a'Ga}$ and correlation between **H** and **I** $\mathrm{cov}(\mathbf{b'y}, \mathbf{a'g})$ becomes $\mathbf{a'Gb}$. Similarly, to minimize the error of $\mathrm{E}[(\mathbf{H} - \mathbf{I})^2]$, that's equal to $\frac{\partial}{\partial \mathrm{b}} \mathrm{E}[(\mathbf{H} - \mathbf{I})^2] = 0$, the equation 4a becomes:

$$\mathbf{a'Ga} + \mathbf{b'Pb} - 2\mathbf{a'Gb} = 0 \quad (6)$$

That after some math become:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{Ga} \quad (7)$$

Note that the correlation between I and true genetic value (G) remain the same. Thus, the latter genetic progress formula is equal to the previous one (5).

Restricted selection index

Restricted selection index is adopted to preserve the genetic value of specifics phenotypes during the time, i.e. setting genetic progress of these traits to 0. Three different approaches have been developed in literature (Kempthorne and Nordskog, 1959; Tallis, 1962; Gjedrem, 1970), all of them leading to the same results. However, Tallies (1962) and further integration given by Ceron Rojas and Crossa (2018) are the most intuitive. The basic idea of these two studies is that maintain the genetics progress to zero, it is equivalent to assume a null correlation between selection index and response to selection for the traits considered, i.e., $(\mathrm{Cov}(\mathbf{H}, \mathbf{I}) = 0)$. Traits under constrain are collected in matrix **C**, where $\mathrm{Cov}(\mathbf{C}, \mathbf{I}) = 0$ which means $\mathbf{Cb}' = 0$. **C** is constructed as $\mathbf{C}' = \mathbf{U}'\mathbf{G}$, where **U** is an incident matrix including as many **1's** as many traits have to be restricted, and **0's** for the other traits. In this situation, the objective of the selection index is twofold: minimize the error $\mathrm{E}[(\mathbf{H} - \mathbf{I})^2]$, and ensure a null genetic progress for the traits of interest, that means $\mathrm{Cov}(\mathbf{C}, \mathbf{I}) = 0$. To solve both objectives it

become necessary to apply the Lagrange multipliers, that allow maximizing (or minimizing) the value of a given function $\mathrm{f}(\mathrm{x},\mathrm{y},\dots)$ under another function of restriction: $\mathrm{g}(\mathrm{x},\mathrm{y},\dots)=\mathrm{c}$:

$$\nabla \mathrm{f}(\mathrm{x}_0,\mathrm{y}_0) = \lambda \nabla \mathrm{g}(\mathrm{x}_0,\mathrm{y}_0) \qquad (8)$$

It can be rearranged in

$$\mathcal{L}(\mathrm{x},\mathrm{y},\lambda) = \mathrm{f}(\mathrm{x},\mathrm{y}) - \lambda(\mathrm{g}(\mathrm{x},\mathrm{y}) - \mathrm{c}) \quad (9)$$

Where $\boldsymbol{\lambda}$ is the vector of Lagrange multipliers.

Thus considering $\mathrm{E}[(\mathbf{H}-\mathbf{I})^2]$ as $\mathrm{f}(\mathrm{x},\mathrm{y})$ and $\mathrm{Cov}(\mathbf{C},\mathbf{I}) - 0$ as $\mathrm{g}(\mathrm{x},\mathrm{y}) - \mathrm{c}$, were *c* is equal to 0, thus equation (9) becomes:

$$\mathcal{L}(\mathbf{b},\boldsymbol{\lambda}) = \mathrm{E}[(\mathbf{H}-\mathbf{I})^2] - \boldsymbol{\lambda}(\mathrm{Cov}(\mathbf{C},\mathbf{I}) - 0\,) \ (10)$$

As demonstrated above $\mathrm{E}[(\mathbf{H}-\mathbf{I})^2]$ is equivalent to $\mathbf{a'Ga} + \mathbf{b'Pb} - \mathbf{2a'Gb}$ and $\mathrm{Cov}(\mathbf{C},\mathbf{I})$ is equal to $\mathbf{Cb'}$.

$$\mathcal{L}(\mathbf{b},\boldsymbol{\lambda}) = \mathbf{a'Ga} + \mathbf{b'Pb} - 2\mathbf{a'Gb} - \boldsymbol{\lambda}(\mathbf{Cb'} - 0\,)(11)$$

Setting the partial derivative of **b** and **v** equal to zero

$$\frac{\partial}{\partial \mathrm{b,v}}\ \mathbf{a'Ga} + \mathbf{b'Pb} - 2\mathbf{a'Gb} + \mathbf{v'Cb} - 0 = 0 \qquad (12)$$

The derivative results from **b** and **v** of (12) are the following (**v** correspond to the vector of Lagrange multiplayer, precedently called **λ**):

$$\frac{\partial}{\partial \mathrm{b}}\ \mathbf{Pb} + \mathbf{Cv} - \mathbf{Ga} = 0 \qquad (13)$$

$$\frac{\partial}{\partial \mathrm{v}}\mathbf{C}'\mathbf{b} = 0 \qquad (13a)$$

In matrix notation the two equations (11) and (11a) are:

$$\begin{bmatrix}\mathbf{0} & \mathbf{C}' \\ \mathbf{C} & \mathbf{P}\end{bmatrix}\begin{bmatrix}\mathbf{v} \\ \mathbf{b}'\end{bmatrix} = \begin{bmatrix}\mathbf{0} \\ \mathbf{Ga}\end{bmatrix} \quad (14)$$

The solution of the matrix $\begin{bmatrix}\mathbf{0} & \mathbf{C}' \\ \mathbf{C} & \mathbf{P}\end{bmatrix}$ must be obtained by inversion (Rojas and Crossa, 2018), that is:

$$\begin{bmatrix}\mathbf{0} & \mathbf{C}' \\ \mathbf{C} & \mathbf{P}\end{bmatrix}^{-1} = \begin{bmatrix}(-\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1} & (\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{P}^{-1} \\ \mathbf{P}^{-1}\mathbf{C}(-\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1} & -\mathbf{P}^{-1}\mathbf{C}(-\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{P}^{-1} + \mathbf{P}^{-1}\end{bmatrix} \quad (15)$$

and therefore:

$$\begin{bmatrix}\mathbf{v} \\ \mathbf{b}\end{bmatrix} = \begin{bmatrix}(-\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1} & (-\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{P}^{-1} \\ \mathbf{P}^{-1}\mathbf{C}(-\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1} & -\mathbf{P}^{-1}\mathbf{C}(-\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{P}^{-1} + \mathbf{P}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{0} \\ \boldsymbol{Ga}\end{bmatrix} \qquad (15a)$$

Solving (13a) for $\mathbf{b}$ could be considered as the equation that maximizes the function under the $\mathbf{C}'\mathbf{b}$ restriction:

$$\mathrm{b} = 0 + \mathbf{Ga}(-\mathbf{P}^{-1}\mathbf{C}(-\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{P}^{-1} + \mathbf{P}^{-1}) \qquad (16)$$

Gathering by P^{-1}:

$$\mathrm{b} = \mathbf{P}^{-1}\mathbf{Ga}\,(-\mathbf{P}^{-1}\mathbf{C}(-\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1}\mathbf{C}' + \mathbf{I}) \qquad (16a)$$

and replacing $(-\mathbf{P}^{-1}\mathbf{C}(-\mathbf{C}\mathbf{P}^{-1}\mathbf{C})^{-1}\mathbf{C}' + \mathbf{I})$ with matrix **K**.

$$\mathbf{b} = \mathbf{K}\mathbf{P}^{-1}\mathbf{Ga} \quad (17)$$

For a matter of clarity, the vector **b** identify in equation (17), is called $\mathbf{b_r}$ to distinguish it with vector **b** (the vector that maximized the $\mathrm{E}[(\mathbf{H}-\mathbf{I})^2]$, thus:

$$\mathbf{b_r} = \mathbf{KP^{-1}Ga} \quad (18)$$

Since $\mathbf{b} = \mathbf{P^{-1}Ga}$ equation (16) can be written as $\mathbf{b_r} = \mathbf{Kb}$. Note that matrix **K** is a linear transformation that reduces the space in which **b** is projected. The space was reduced as many are the restriction applied, obtaining a vector $\mathbf{b_r}$ that is **b** after restriction for the traits of interest.. The genetic progress (**R**) is obtained replacing **b** with $\mathbf{b_r}$ in formula (8a), that is $\mathbf{b_r} = \mathbf{P^{-1}Ga}$ and then using the new weights in formula (5):

$$\mathbf{R} = \mathbf{b_r'G}(\mathbf{b_r}'\mathbf{Pb_r})^{-\frac{1}{2}}\,\mathrm{i} \quad (19)$$

In this study, we analytically demonstrate how it is possible to obtain the economic weights (reported in vector **a**) when the restriction to variation for some traits included within the selection indices is applied. This passage is not described in the literature, but it has considerable practical importance.

The new economic weights can be obtained by considering equation (19) and (8a) as equal. Thus, we obtain $\mathbf{b_r} = \mathbf{Kb}$ and inverting the formula (8a) and we get $\mathbf{a} = \mathbf{G^{-1}Pb}$. Then the values of new economic weights, that we call $\mathrm{a_r}$, are obtained substituting b with Kb in the previous equations:

$$\mathbf{a}_{\mathbf{r}} = \mathbf{KG^{-1}Pb_r} \qquad (20)$$

The term $\mathrm{a_r}$, represents the vector of the new weights to provide to the traits for obtaining a null genetic progress for some target traits and a positive genetic gain for the others. It is

possible to express the new economic weights in a scale from 0 to 1 by dividing each weight for the sum of the absolute values of the others.

Finally, it is possible to check if the genetic progress with the new $\mathbf{a_r}$ is what desired, by using this new vector of economic weights in the general formula for genetic progress without restriction (5) after divided the $\mathbf{a_r}$ for the genetic standard deviation of each trait t if this operation was done also previously $\mathbf{a_{rt}}/\sigma_{at}$; see (Wolfová et al., 2001)).

Application

In this study actual genetic and phenotyps correlations estimated for a dual purpose local breed (Mancin et al., 2021), were considered (Table 1). To simplify the number of tratis, milk yield (kg), fat yield (kg), protein yield (kg), somatic cell score (SCS) and muscularity (as factor score; see Mancin et al., 2021 for computation) were considered. This study example considered a restriction to zero for the last two traits, SCS and muscularity. The initial economic weights, the final economic weights obtained after restriction of two traits and the corresponding responses to selection (ΔG) are reported in Table 2. The main equations described above have been used through the example and recalled at the corresponding steps.

Table 1. Traits included in the example with their phenotypic mean (Mean) heritability (h^2, on diagonal), and phenotypic (above diagonal) and genetic correlations (below diagonal) among the traits.

Trait	ID	Mean	MY	FY	PY	SCS	MUSC
Milk yield (kg)	MY	2.210	0.219	0.049	0.081	0.0630	-2.056
Fat yield (kg)	FY	2.600*	0.250	0.178	0.002	0.002	-0.054
Protein yield (kg)	PY	1.895*	0.411	0.0112	0.1255	0.0035	-0.077
Somatic cell score (points)[1]	SCS	0.379	-0.792	-0.014	-0.026	0.1332	-0.290
Muscularity (points)[1]	MUSC	9.144	1.484	-0.002	-0.002	0.0317	0.328

[1] Traits to which a restriction to the genetic progress was applied

*Variances have been multiplied by 10^3

Table 2. Traits included in the example with their initial economic weights (a) and the final economic weights (a_r) applied after a restriction for the genetic progress of two traits, and the corresponding response to selection (ΔG and ΔG_r).

Trait	a	ΔG	a_r	ΔG_r
Milk yield (kg)	0.000	0.844	0.000	0.513
Fat yield (kg)	0.300	0.030	0.518	0.025
Protein yield (kg)	0.400	0.051	0.465	0.041
Somatic cell score (points)[1]	0.000	0.079	-0.008	0.000
Muscularity (points)[1]	0.300	-0.878	0.010	0.000

[1] Traits to which a restriction to the genetic progress was applied

The procedure is presented below as R code (R Core Team, 2017).

Step 1: definition of matrices and vectors of data. A selection intensity of i=1.755 has been set, corresponding to a percentage p=0.1 of parents selected. Matrices of phenotypic and genetic (co)variances (**P** and **G**) for the target traits have been also set. Traits considered were:

milk yield (**MY),** fat yield (**FY**), protein yield (**PY),** somatic cell score (**SCS**) and muscularity (**MUSC)**. The economic weights of the traits have been reported in the vector **a**, and they mimic a real situation: 0.3 and 0.4 for fat and protein as milk quality traits and 0.3 for muscularity, to preserve the dual purpose attitude. Milk yield has an economic weight of zero because it is indirectly selected due its high genetic correlations with fat and protein yields. SCS has an initial weight of zero because the trait is aimed to be then restricted to obtain a null genetic progress. The economic weights of traits have been standardized to calculate the genetic progress.

```
        # selection intensity
i = 1.755


traits = c("MY", "FY" ,"PY","SCS", "MUSC")
# matrix of phenotypic covariances
P = matrix(c(10.1072, 0.2507,  0.4118, -0.7921, -1.4849,
          0.2507,  0.0105,  0.0112, -0.0137, -0.0017,
          0.4118,  0.0112,  0.0205, -0.0265, -0.0018,
         -0.7921, -0.0137, -0.0265,  2.7916,  0.0317,
          1.4849, -0.0017, -0.0018,  0.0317, 27.3200),
          ncol=5)



# matrix of genetic covariances
G = matrix(c(  2.2353,  0.0494,  0.0815,  0.0630, -2.0559,
            0.0494,  0.0019,  0.0023,  0.0018, -0.0540,
            0.0815,  0.0023,  0.0042,  0.0035, -0.0769,
```

```
            0.0630,  0.0018,  0.0035,  0.3768, -0.2906,
           -2.0559, -0.0540, -0.0769, -0.2906,  9.0014),
            ncol=5)

colnames(G) = traits; row.names(G) = traits
colnames(P) = traits; row.names(P) = traits
print(P)

        ##          MY      FY      PY     SCS    MUSC
## MY    10.1072  0.2507  0.4118 -0.7921  1.4849
## FY     0.2507  0.0105  0.0112 -0.0137 -0.0017
## PY     0.4118  0.0112  0.0205 -0.0265 -0.0018
## SCS   -0.7921 -0.0137 -0.0265  2.7916  0.0317
## MUSC -1.4849 -0.0017 -0.0018  0.0317 27.3200

        print(G)

        ##          MY      FY      PY     SCS    MUSC
## MY     2.2353  0.0494  0.0815  0.0630 -2.0559
## FY     0.0494  0.0019  0.0023  0.0018 -0.0540
## PY     0.0815  0.0023  0.0042  0.0035 -0.0769
## SCS    0.0630  0.0018  0.0035  0.3768 -0.2906
## MUSC -2.0559 -0.0540 -0.0769 -0.2906  9.0014

        # genetic standard deviations of traits
ds_G = sqrt(diag(G))
```

```
# initial economic weights of traits
a = c(0,0.5,0.5,0,0.0)


# standardized economic weights
a_std = a/(ds_G)
```

Step 2. Multivariate genetic progress calculated without restrictions for any traits.

The equations [4], [5b] and [6] have been applied.

```
        # Inverse of P matrix
Pinv = solve(P)


# Coefficient that maximize the equations Pb = Ga
b = Pinv %*% G %*% (a_std)
print(b)

        ##            [,1]
## MY    0.02889588
## FY    1.50905975
## PY    1.52049245
## SCS   0.04759244
## MUSC -0.04268001

        # calculate standard deviation of the index
ds_I = sqrt(t(b) %*% P %*% b)


# calculate genetic progress
```

```
gp = (i/ds_I) %*% t(b) %*% G


print(round(gp,5))

        ##          MY      FY      PY     SCS     MUSC
## [1,] 1.27378 0.03666 0.05638 0.14471 -2.36117
```

Step 3. Application of the restriction. Milk yields traits (MY, FY and PY) have shown a positive genetic progress, whereas SCS had a positive increase (that means a detrimental effect on uddr health) and MUSC resulted in worsening the mean vale by selection. Two restrictions to genetic progress were then applied to SCS and to MUSC, to prevent their negative effect on udder health and on reduction of muscularity. The procedure moved from the definition of the matrix **C** including the genetic variances for SCS and MUSC and their covariances with the other traits, in column. This matrix has been then included within the matrix of restriction K then used for finding out the new coefficients $\mathbf{b}_r$. This step applied the equations [12], [14] and [6] including the new coefficients.

```
        # create a matrix of restriction for SCS and MUSC
C = G[c('SCS','MUSC'),]
C = t(C)
print(C)

        ##          SCS    MUSC
## MY     0.0630 -2.0559
## FY     0.0018 -0.0540
## PY     0.0035 -0.0769
```

```
## SCS   0.3768 -0.2906
## MUSC -0.2906  9.0014

        # create an identity matrix with the dimension of the matrix correlations:
I = diag(rep(1,ncol(G)))


# calculate matrix K (matrix of restrictions)
K = I - (Pinv %*% ((C) %*% solve(t(C) %*% Pinv %*% (C)) %*% t(C)))


# coefficient that maximizes restriction
br = K %*% Pinv %*% G %*% (a_std)


#calculate genetic progress
gp = ((i/(ds_I))%*%t(br)%*%G)


print(round(gp,3))

        ##       MY    FY    PY SCS MUSC
## [1,] 0.36 0.017 0.029   0    0
```

Step 4. Obtaining the new economic weights for traits arising from the restrictions in the selection index. The equation [15] has been used in this case. The new economic weights have been then divided by the sum of their absolute value to obtain a sum of zero. Looking at the results, it is possible to observe that applying the restriction for SCS and MUSC is equivalent to provide a negative economic weight of -0.008 to SCS, an economic weight of 0.010 to MUSC, and weights of 0.518 and 0.465 for fat and protein.

```
        # calculated economic weight according with the restrictions
new_a = solve(G)%*%(P)%*%(br)


a = c(new_a)/(sum(abs(new_a)))
print(round(a,3))

        ## 0.000  0.588  0.395 -0.007  0.010
```

Step 5. Check the genetic progress. Aiming to check that inserting the new economic weights in the traditional equation for the genetic progress provides the same result than applying the restriction, it is possible to insert these new economic weights in the equation of genetic progress [6] after divided the economic weight for the genetic standard deviation of each trait.

```
        a_std = a/(ds_G)


b = Pinv %*% G %*% (a)
print(b)

        ##             [,1]
## MY   -0.0043681974
## FY    0.0754342010
## PY    0.1605745558
## SCS  -0.0004963555
## MUSC  0.0008106308

        # calculate standard deviation of the index
ds_I = sqrt(t(b) %*% P %*% b)
```

```r
# calculate genetic progress
gp = (i/ds_l) %*% t(b) %*% G

print(round(gp,3))
```

```
##         MY    FY    PY SCS MUSC
## [1,] 0.513 0.025 0.041   0    0
```

RESULTS AND DISCUSSION

Finding the economic weights that make a genetic progression equal to zero in each trait is helpful in dairy cows breeding (but also in beef cattle and in other livestock species), in which many different traits are included in the selection index.

Particularly, this approach could be convenient for traits that had an intermediate optimum value. In dairy and beef cattle indeed, but also in other livestock species, selection indexes also include a wide number of morphological characters with an intermediate optimum, such as leg traits or udder conformation (Jeyaruban et al., 2012). In these traits genetic improvement is aimed at maintaining the phenotype at intermediate values, and this could be only realized by guaranteeing that the trait has not a positive (or negative) increase. Other traits that should be maintained at the existing value while selecting for production improvement are the functional traits showing negative genetic correlations with production, such as birth weight vs. growth. In the study of Winder et al., (1990), in Red Angus cattle, the genetic gain of birth weight was restricted to zero aiming to preserve the trait, despite the negative genetic correlations (around -0.7) with the relative growth rate. Other functional traits typically showing negative genetic correlations with production are fertility and longevity (Oltenacu and Broom, 2010). Depending on the breeds' attitude and on their history, these traits are aimed to be maintained (as in dual purpose cattle) or positive selected due to their detriment occurred over the last decades of selection for just improving production (as in specialized dairy and beef breeds; Miglior et al., 2005).

This technique could be also implemented on traits that have a low or not calculable economic value but are related to the typical characteristics of the breed or to its adaptability to the territory in which it is reared (Krupová et al., 2016). This is the case of traits like feed

efficiency (Fuerst-Waltl et al., 2016) or important morphological traits (Mancin et al., 2021. These traits could be lost during the selection process for productive traits such as milk yield, therefore a solution consists of guarantee at least a non-negative genetic progress by applying a restriction while increasing the production.

The restriction of the genetic gain could be also applied to traits which positive increase corresponds to a detriment of some functional characteristics, such as the somatic cells' traits like SCS, which positive variation means a loss in udder health. A restriction to zero as a solution to prevent positive trends for SCS has been proposed e.g., in Alpine Grey (Mancin et al., 2021) and Rendena cattle (Sartori et al., 2018).

The restriction could be also a solution for productive traits with lesser economic importance than others and showing negative genetic correlations with the first one. This situation is likely to occur in some dual-purpose breeds, in which the two aptitudes, e.g., milk and beef, show negative genetic correlations, and a much greater economic weight is assigned to one aptitude. Under these circumstances, the genetic trend of the other aptitude may be negative, and restriction to zero could be a valuable solution to prevent a negative variation. This is the case of muscularity (MUSC) in the practical example reported above, showing a negative genetic correlation (around -0.4) with milk yield traits, as observed in many further studied on dual purpose breeds (Mancin et al., 2021, Sartori et al., 2018). In order to obtain the greatest positive increase as possible for milk without a detriment in muscularity, a restriction to the genetic gain of MUSC has been applied, as also proposed in Mancin et al. (2021).

According to Wolfová et al. (2005), it should be an uncorrected priori assumptions to consider some traits like MUSC and SCS as restricted traits, because they have own specific

economic value. However, in some situation as local breeds no previous studies about economic assessment of rearing system of the target breed (Alpine Grey in this example) were done. According to Krupová et al. (2016), it could be considered as a preliminary process to establish a suitable selection index satisfying the needs of any breed associations.

In conclusion, this study aimed to explain the theory behind the creation of single and multivariate selection indices and use of restriction for some traits following in a selection index, and how to derive new economic weights for each trait under restriction of genetic progress for some target traits. The method presented could be useful in all situations in which negative correlations occurs among target selected traits, allowing simultaneous increase and/or not worsening of the all-target traits, with beneficial effects especially in situation where many traits are required to be accounted for selection.

BIBLIOGRAPHY:

Céron-Rojas, J. J., and Crossa, J. (2018). *Linear Selection Indices in Modern Plant Breeding*. Available at: https://link.springer.com/content/pdf/10.1007%2F978-3-319-91223-3.pdf.

Cunningham, E. P., and McClintock, A. E. (1974). Selection in dual purpose cattle populations: effect of beef crossing and cow replacement rates. *Ann.Genet.Select.Anim.* 6, 227–239. doi:10.1186/1297-9686-6-2-227.

Falconer, D. S., and Mackay, T. F. C. (1996). Introduction to Quantitative Genetics. *Longmans Green, Harlow, Essex, UK* Ed 4., 464.

Fuerst-Waltl, B., Fuerst, C., Obritzhauser, W., and Egger-Danner, C. (2016). Sustainable breeding objectives and possible selection response: Finding the balance between economics and breeders' preferences. *J. Dairy Sci.* 99, 9796–9809. doi:https://doi.org/10.3168/jds.2016-11095.

Hazel, L. N. (1943). The Genetic Basis for Constructing Selection Indexes. *Genetics* 28, 476–490.

Jeyaruban, G., Tier, B., Johnston, D., and Graser, H. (2012). Genetic analysis of feet and leg traits of Australian Angus cattle using linear and threshold models. *Anim. Prod. Sci.* 52, 1–10. doi:10.1071/AN11153.

Krupová, Z., Krupa, E., Michaličková, M., Wolfová, M., and Kasarda, R. (2016). Economic values for health and feed efficiency traits of dual-purpose cattle in marginal areas. *J. Dairy Sci.* 99, 644–656. doi:10.3168/jds.2015-9951.

Mancin, E., Sartori, C., Guzzo, N., Tuliozi, B., and Mantovani, R. (2021). Selection Response Due to Different Combination of Antagonistic Milk, Beef, and Morphological Traits in the Alpine Grey Cattle Breed. *Animals* 11. doi:10.3390/ani11051340.

Mazza, S., Guzzo, N., Sartori, C., and Mantovani, R. (2016). Genetic correlations between type and test-day milk yield in small dual-purpose cattle populations: The Aosta Red Pied breed as a case study. *J. Dairy Sci.* 99, 8127–8136. doi:10.3168/jds.2016-11116.

Miglior, F., Muir, B. L., and Van Doormaal, B. J. (2005). Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88, 1255–1263. doi:10.3168/jds.S0022-0302(05)72792-2.

Oltenacu, P. A., and Broom, D. M. (2010). The impact of genetic selection for increased milk yield on the welfare of dairy cows. *Anim. Welf.* 19, 39–49.

R Core Team (2017). R: A Language and Environment for Statistical Computing.

Sartori, C., Guzzo, N., Mazza, S., and Mantovani, R. (2018). Genetic correlations among milk yield, morphology, performance test traits and somatic cells in dual-purpose Rendena breed. *Animal* 12, 906–914. doi:10.1017/S1751731117002543.

Winder, J. A., Brinks, J. S., Bourdon, R. M., and Golden, B. L. (1990). Genetic analysis of absolute growth measurements, relative growth rate and restricted selection indices in red Angus cattle. *J. Anim. Sci.* 68, 330–336. doi:10.2527/1990.682330x.

Wolfová, M., Nitter, G., Wolf, J., and Fiedler, J. (2001). Impact of crossing system on relative economic weights of traits in purebred pig populations. *J. Anim. Breed. Genet.* 118, 389–402. doi:https://doi.org/10.1046/j.1439-0388.2001.00304.x.

Wolfová, M., Wolf, J., Přibyl, J., Zahrádková, R., and Kica, J. (2005). Breeding objectives for

beef cattle used in different production systems: 1. Model development. *Livest. Prod. Sci.* 95, 201–215. doi:10.1016/j.livprodsci.2004.12.018.

Xie C.,Xu. S., Restricted multistage selection indices. Genetics Selection Evolution, BioMed Central, 1997, 29 (2), pp.193-203.

6. SELECTION RESPONSE DUE TO DIFFERENT COMBINATION OF ANTAGONISTIC MILK, BEEF, AND MORPHOLOGICAL TRAITS IN THE ALPINEGREY CATTLE BREED

STATUS: PUBLICATED ON ANIMALS
https://doi.org/10.3390/ani11051340

Selection response due to different combination of antagonistic milk, beef, and morphological traits in the Alpine Grey cattle breed

Enrico Mancin, Cristina Sartori, Nadia Guzzo, Beniamino Tuliozi, and Roberto Mantovani

ABSTRACT

Selection in local dual-purpose breeds requires great carefulness because of the need to preserve peculiar traits and also guarantee positive genetic progress for milk and beef production to maintain economic competitiveness. A specific breeding plan accounting for milk, beef, and functional traits is required by breeders of the Alpine Grey cattle (AG), a local dual-purpose breed of Italian Alps. Hereditability and genetic correlations among all traits have been analyzed for this purpose. After that, different selection indexes were proposed to identify the most suitable for this breed. Firstly, a genetic parameters analysis was carried out with different datasets. The milk dataset contained 406,918 test day records of milk, protein, and fat yields and somatic cells (expressed as SCS). In the beef dataset were included performance test data conducted on 749 young bulls. Average daily gain, in vivo estimated carcass yields and SEUROP were the phenotypes obtained from the performance tests. The morphological dataset included 21 linear type evaluations of 11,320 first party cows. Linear type traits were aggregated through factor analysis and three factors were retained, while head typicality (HT) and rear muscularity (RM) were analyzed as single traits. Heritability estimates (h^2) for milk traits ranged from 0.125 to 0.219.

Analysis of beef traits showed h^2 greater than milk traits, ranging from 0.282 to 0.501. Type traits showed a medium value of h^2 ranging from 0.238 to 0.374. Regarding genetic correlation, SCS and milk traits were strongly positively correlated. Milk traits had a negative genetic correlation with the factor accounting for udder conformations (-0.40) and with all performance test traits and RM. These latter traits showed also a negative genetic correlation with udder volume (-0.28). The HT and the factor accounting for rear legs traits were not correlated with milk traits, but negatively correlated with beef traits (-0.32 with RM). We argue that the consequence of these results is that the use of the current selection index, which is mainly focused on milk attitude, will lead to a deterioration of all other traits. In this study we propose more appropriate selection indexes that account for genetic relationships among traits, including functional traits.

INTRODUCTION

The extension of milk and beef markets has contributed to the gradual decline of the appeal of local domestic breeds due to their low productive performance compared with specialized breeds [1].

Despite this, in recent years local breeds have received an increased interest, as compared with cosmopolite breeds, they have better preserved functional characteristics (health, fertility, longevity, and rusticity). Local breeds are also often associated with the production of labelled foods (protected designation of origin/protected geographical indication), especially cheeses. [2]. Not considering only the economic aspects, these breeds have a link with the territory supporting rural/local economy and represent an effective resource of biodiversity [3]. In addition, local breeds are more adaptable to environmental changes than specialized breeds and they can also be bred in marginal areas and low-income environments [4]. Many local breeds have a dual-purpose attitude for milk and meat, but with differing emphases depending on traits' local economic importance.

For these reasons, it is fundamental to ensure accurate breeding plans (aimed at improving the traits of interest) for local breeds, as previous studies have shown a negative genetic correlation between milk and meat traits [5,6].

The Alpine Grey is an autochthonous cattle breed of the central Alpine arc, widespread in Tyrol (Austria), South Tyrol (Italy), and neighboring Switzerland. Each country maintains its own herd book and independent breeding plans [7]. The Alpine Grey is generally well adapted to live and produce both milk and meat under challenging environments based on Alpine pastures. The present Italian Alpine Grey population accounts for 17,373 heads in 1,737 farms (www.fao.org/dad-is/; update: 26 Oct. 2020) distributed mainly in the provinces of Bolzano and Trento (85%). Milk production amounts to 5,339 kg of milk per lactation with 3.75% of fat and 3.39% of protein, respectively. The average daily gain (ADG) of young bulls can reach 1.2 kg/day, and the carcass yields about 58% (ANAGA; www.grigioalpina.it). The breeding system for this Gray Alpine is generally constituted by small farms housing cattle during the winter months and releasing them to pasture in summer. The actual breeding goal to improve milk and meat traits is based on a selection index that assigns an economic weight of 24% to fat yield and 46% to protein yield. A further 20% of the economic weight is attributed to young bulls' ADG, and the remaining 10% to rear muscularity (RM), which is evaluated as a type trait on primiparous cows. Therefore, the present breeding plan does not account for further functional

and morphological traits that characterize the breed. Notwithstanding, these traits are required with increasing interest by breeders to maintain thetypicality and rusticity of the breed.

Therefore, the aim of this study was to estimate the genetic correlations between milk, beef, and functional and morphological traits in the local Alpine Grey breed. This was done by analyzing the genetic response to selection under different scenarios in which different weights were applied to current and novel traits to be accounted for in the selection index. Particularly, traits investigated were milk, fat, protein, and somatic cell score. This latter trait was derived from the test-day (milk) dataset, whereas type traits were obtained from the scoring of primiparous cows and beef traits fromyoung bulls at performance test.

MATERIALS AND METHODS

Data editing*:*

All data were provided by the National Breeder Association of Alpine Grey cattle (ANAGA). The study used three different datasets, including milk, beef, and morphological traits. The milk datasetcontained information on milk, fat, and protein yields (MY; FY, and PY, respectively; kg/d) and somatic cell counts (no./ml), with an average interval of 4 weeks between test day collection.

Dataset of morphological evaluations contained 21 type traits routinely scored on primiparous cows when aged about three years (36.9 ± 5.0 months). The dataset on beef attitude was obtained from performance test data and contained the average daily gain (ADG), an in vivo estimate of carcass yield (CY) and muscularity traits (SEUROP scale) carried out by skilled classifiers on young bulls aged about 12 months.

Milk dataset initially contained 1,134,032 individual test-day (TD) productions routinely collected from 1997 to 2018 following the Italian official milk recording system. The number of somatic cells/ml was converted into the normally distributed somatic cells score (SCS) according to [8]. As first data editing, the TD records with missing values, and the ones recorded when days in milk (DIM) was under 5 d or over 305 d from calving were removed. In additions, only TD belonging tolactation from 1 to 3 were retained. Values for MY, FY, and PY outside the mean ± four standard deviations within parity and lactation phase (considering 15 d intervals) were taken away from the data set as outliers. Among the remaining records, only those belonging to cows with age at calvingbetween 21 and 44 months at first parity, between 32 and 60 months at second parity, and between 44 and 76 months at third parity were retained for analysis. Furthermore, only lactations with a firstTD carried out within 45 days from

calving and including at least four records were kept for furtheranalysis: the reason for this was that functional controls of the cows are limited during the first 45 days, with the lactation peak occurring later in this breed [5]

Lastly, only records belonging to herd-TD with at least two observations could enter the final dataset. At the end of the editing process, a final dataset with 406,918 TD records belonging to 58,041 lactations and 29,219 cows was used. The pedigree file contained 49,389 animals, tracingback up to the sixth generation (complete generations).

Morphological traits are routinely measured once in the lifetime, around the time of first calving. Data initially consisted of 14,669 observations of 21 type traits scored on a scale of 1 to 50 points by trained classifiers during 2010 – 2018. The edited dataset contained 11,318 final records belonging to the same number of cows and 32,494 animals in the pedigree file. Records allowed to the final dataset were scored between 5 to 305 days in milk (DIM) and considered cows with age at first calving between 21 and 45 months. An exploratory factor analysis was carried out on all 21 type traits applying the varimax rotation [5,9]. Varimax rotation allows for a better interpretation ofthe biological meaning of each factor. The process consists in adjusting (rotating) the coordinates obtained from Principal Components Analysis (PCA). This adjustment is based on maximizing the variance shared among components, increasing the squared correlation of items related to one factor, while decreasing the correlation to any other factor.

Then, further analysis retained three main factors describing: udder volume (UV; Factor 2, that is F2 in Figure 1), udder conformation (UC; Factor 3, that is F3 in Figure 1) and rear legs (RL; Factor 7, that is F7 in Figure 1), as these three were the main factors of interest to breeders in terms of the genetic index, as they are latent factors connected with the production of milk (F2), with the healthof the udder (F3) or with the aptitude for grazing (F7).

These factors showed eigenvalues greater than one. They were named based on the biological meaning of the linear type traits showing the loading coefficients highest than an absolute thresholdof |0.45|, a way of proceeding also applied in other previous studies [5,6].

The milk trait dataset and the morphological dataset were combined to perform correlation analysis.The two datasets had in common 9,145 animals representing about 30% of the animals in the milk dataset.

The performance test dataset contained 749 records collected from 1988 to 2018 and belonging to the same number of bulls grouped by age and accounting for 6,266 animals in the pedigree file. The final dataset included only contemporary groups of young bulls consisting of at least three animals. The average daily gain (ADG) was obtained as a linear regression of monthly weight on age.

Further analysis considered only regressions with a coefficient of determination of at least 0.95. The number of weight-age couples used for the linear regression was 12 for each bull, and the average age at the beginning and the end of the performance test were respectively 50 ± 12 d and 356 ± 11d . The in vivo visual appreciation of fleshiness, evaluated using the SEUROP scale, and the in vivo carcass yields scored independently by two evaluators, were obtained at about 375 ± 16 d of age. For each trait, the analysis considered the average of the two evaluations. The in vivo SEUROP score considering the grades S, E, U, R, O, and P, from the best to worst conformation, was further subdivided into + or − subclasses as in [10]. The scores were then transformed in a linear scale from 80 (corresponding to a grade of P) to 130 (corresponding to S), adding or subtracting 3.33 points to the full class when necessary; that is, for an R+ grade, the score was 103.33, whereas for the U− it was 106.67. The whole numeric interval, ranging from 76.67 to 133.33, was considered as continuous. The carcass yield was expressed as a percentage and was an *in vivo* appraisal of the predicted carcass incidence at slaughter.

Models:

Milk traits were analyzed using the following test-day model:

$$y_{ijklmno} = HTD_i + LN_j + GL_k + \sum_{r=1}^{3} \varphi_r \times AP\text{-}LN_l + \sum_{r=1}^{3} \psi_r \times MP\text{-}LN_m + Pe_n + a_n + e_{ijklmno}$$

where $y_{ijklmno}$ is the individual test-day oth record (milk, fat, protein, and SCS) of the nth cow; HTD_i is the fixed effect of the herd-test-day (90 012 levels); LN_j represents the fixed effect of lactation number (3 levels, corresponding to the first three lactations); GL_k is the fixed effect of kth gestation length class (18 classes with 1 meaning no gestation and further classes accounting for 15-d intervals, from 1 to 240 d of gestation); AP-LN is the fixed effect of lth age at parity within lactation (42 classes in total); MP-LN is the fixed effect of the mth month of parity (36 classes corresponding to single months of a year within each j lactation); Pe is the random permanent environmental component, N (0, σ^2_{pe}); a is the additive genetic component, N (0, σ^2_a); and $e_{ijklmno}$ is the random residual term, N (0, σ^2_e). Fourth-order Legendre polynomials described the shape of the lactation curve for the fixed effects of AP-LN and MP-LN, with φ and

ψ as fixed regression coefficients for the Legendre polynomial of order r varying between 0 and 3. Factor analysis for all the morphological traits was carried out to reduce the number of traits and avoid redundant morphological measurements (see also Data Editing section). The traits included in the analysis are described in Table 1. The factor analysis was performed using the "psych" package of R [11]. The varimax rotation method was used [12]. Latent variables with eigenvalues $\geqslant 1$ were retained for further analysis. The factor score originated from every latent variable was considered as a new trait. According to traits of main interest expressed by the national breeders' association, only three of seven latent factors were considered for subsequent analyses. They were called as Factor 2 (F2), corresponding to udder volume; Factor 3 (F3), that is udder correctness, and Factor 7 (F7), rear legs (Figure 1). Together with these factors, the subsequent analysis also included the linear scores for rear muscularity (RM) and head typicality (HT), respectively considered as a beef trait and a functional trait for the breed.

The five morphological traits, two linear and three factor scores, were analyzed with the following model:

$$y_{ijklm} = HY_i + C_j + AC_k + DIM_l + a_m + e_{ijklm};$$

where y_{ijkl} is one of the five morphological traits; HY_i, C_j, AC_j, and DIM_k are respectively the fixed effects of the herd-year (i = 3 318 levels); the classifier (j = 67 levels); the age at calving (k = 12 classes: <21 months, from 21 to 45 using 2-months intervals); and the days in milk (l= 20 classes from 5 to 305 days after calving and using 15-days intervals a_m is genetic random additive effect of animals N $(0,\sigma^2_a)$; and e_{ijklm} is the random residual term, N $(0, \sigma^2_e)$.

Regarding the beef traits, the following animal model was implemented:

$$y_{ij} = GP_i + a_j + e_{ij},$$

where y_{ij} is a performance test phenotype for ADG, SEUROP, or CY; GP represents the categorical fixed effect of the contemporary group (i = 142 levels); a_j is the random additive genetic effect of the young bull j; and e_{ij} is the random residual term.

Variance component estimates and model assumptions

To estimate the (co)variance components, a Gibbs sampling algorithm was used, and the analysis was performed with the *gibbs3f90* program [13].The program generated a total number of 480,000 samples and considered an initial burn-in of30,000; one of every 150 chains was retained. A Gaussian distribution for all effects was considered. Flat priors were used for

all fixed effects, and null means and normal distributed priors were used for permanent environment, additive genetic, and residual terms, with this matrix notations:

$$a \sim N(0, G \otimes A); pe \sim N(0, Pe \otimes I); e \sim N(0, R \otimes I);$$

where **A** represents the relationship matrix obtained from pedigree, and **I** is an identity matrix. Heritability was obtained from variance components estimated by applying single-trait models, while genetic and phenotypic correlations from bi-traits models were obtained by merging the three different datasets in pairs. The covariance matrices used in the bi-traits analysis were as follows:

$$G = \begin{vmatrix} \sigma^2_{a1} & \sigma_{a1a2} \\ \sigma_{a1a2} & \sigma^2_{a2} \end{vmatrix}; Pe = \begin{vmatrix} 0[\sigma^2_{Pe1}] & 0[\sigma_{Pe1Pe2}] \\ 0[\sigma_{Pe1Pe2}] & 0[\sigma^2_{Pe2}] \end{vmatrix}; R = \begin{vmatrix} \sigma^2_{e1} & 0[\sigma_{e1e2}] \\ 0[\sigma_{e1e2}] & \sigma^2_{e2} \end{vmatrix};$$

where **G** is the matrix of additive genetic (co)variances σ^2_{a1}, σ^2_{a2}, σ_{a1a2} of traits 1 and 2, **Pe** is the matrix of permanent environmental (co)variances σ^2_{pe1}, σ_{pe1pe2}, σ^2_{pe2}, and **R** the matrix of residual (co)variances σ^2_{e1}, σ^2_{e2} and σ_{e1e2} of traits 1 and 2. When different datasets were merged, residual (co)variance was set to zero because the traits were recorded in different moments. In single traits analysis, **Pe** was not considered (i.e., beef and morphological datasets) because obtained only once in life. Nevertheless, when morphological or beef traits were analyzed with milk traits, a covariance σ_{pe1pe2} was included to provide a better estimate of the permanent environment component for milk yields traits, according to [14]. From a biological point of view this σ_{pe1pe2} represents the relationship between traits due to the common environment represented by each individual

Estimated selection response

A final step consisted in calculating the theoretical multivariate response to selection (R) under different weights for each trait considered as a breeding goal. The response to selection (R) is the change of the phenotypic mean during a generation for a specific or a group of selected traits. The theoretical multivariate response to selection [15] was calculated according to [16] using the following formula:

$$\mathbf{R=(i/\sigma_i)\cdot b'\cdot P^{-1}}$$

where i is the selection intensity set to 1.755 as in [6] corresponding to a proportion of 0.10 selected animals in the whole population, assuming a normal distribution; σ_i is the SD of the selection index, obtained as $\sigma_i = (b' P b)^{1/2}$; **b** is the vector of the weights for selection index and **b'** its transpose, with $b = P^{-1} Ga_s$. In this formula **P** and **G** are the phenotypic and the

genetic (co)variance matrices, respectively, and a_s is the vector including the economic weights of traits. In this vector **P** and **G** have the same meaning as in in the formula above, and as is the vector including the standardized economic weights of traits. As in [16], the relative emphasis of the traits in the selection index was intended as a proportion a of the trait's standardized economic value (i.e., as = a× σ_a) compared with the sum of all standardized values of all the traits accounted in the index. A final standardized response to selection (R_{dsi}) was calculated as $R_{dsi} = R / \sigma_i$.

Eight different scenarios (Si) to estimate the selection response were simulated.

The first scenario(S1) considered the current selection emphasis given to traits under routinely selection practices: 70% for milk traits (24% to fat yield and a 46% to protein yield) and the remaining 30% attributed to ADG (20%) and RM (10%) beef traits. All the other traits (that are all the traits mentioned above of the three datasets) had a selection emphasis of zero since they are not included in the selection index. They are indirectly selected due to the genetic correlations they have with the traits under direct selection. Scenario 2 and 3 (S2 and S3) had the same selection weights of S1, but in S2 the genetic gain for SCS was restricted to zero according to previous studies [6,16]. This restriction was done to prevent an increase in SCS since it could imply a detriment in udder health conditions.

Similarly, in S3, the genetic gain for both SCS and RM were restricted to zero. The RM variation was restricted to zero to prevent a worsening of the traits due to the negative genetic correlations occurring with milk yield traits, as reported below. In this latter case, 30% of weight attributed to beef traits was entirely shifted to ADG. The S4 and S5 provided less emphasis on milk attitude (65% for both scenarios), and the remaining 35% was divided in different manners. The S4 assigned 15% of the weight to RM, 5% to SEUROP, and 5% to CY, meaning a total of 25% of the weight on beef traits, and attributed 7% of the weight to F3-UC and 3% to HT. In S5, less credit to morphological traits was given respect to S4, i.e., 3.5% of the weight to F3-UC and 1.5% to HT, with a corresponding increase of RM to 20% of the weight. In S6 e S7, the milk traits' weight was further reduced to 55%. The morphological traits F3-UC and HT received the same weight as in S4 (for S6), and S5 (for S7), while the beef traits SEUROP and CY received a 10% of weight each in both S6 and S7. Last, in S8, milk traits were set to 70%, and beef traits to 30%, specifically 20% to RM and 5% each to SEUROP and CY, but restriction to zero were imposed for SCS and morphological traits (F3-UC, F7-RL, and HT) to prevent a detriment in their genetic variation.

RESULTS

Descriptive statistics and factor analysis

Mean, standard deviation, maximum and minimum values for all the studied traits are shown in Table 1. Milk, fat, and protein TD yields indicated a daily production of 16.3 kg/d, 0.62 kg/d, and 0.56 kg/d, respectively. Regarding SCS, a mean of 2.33 was found, i.e., approximately 62,760 cells/ml, suggesting an excellent value for mammary health in Grey Alpine. Almost all morphological traits present an average value close to 28 points, except the teats position – rear view (23.5 points). Young proved bulls presented a mean ADG of 1.15 kg/d, with a carcass yield of 56% and an average SEUROP conformation score of 103 points, corresponding to an R+ score.

Table 1. Descriptive statistics of analyzed traits.

Traits	Mean	SD	Minimum	Maximum
Milk traits:				
- Milk yield (kg/d)	16.30	5.36	0.60	45.20
- Fat yield (kg/d)	0.62	0.21	0.02	2.163
- Protein yield (kg/d)	0.56	0.17	0.02	1.49
- Somatic cell score (points)	2.33	1.86	-3.64	10.84
Linear Type traits (points; scale 1-50):				
- Strength/Robustness	29.15	6.67	Tight & weak	Large & strong
- Thinness	26.42	5.46	Heavy & coarse	Thin & sharp
- Shoulders	28.80	5.65	Loose	Smooth and adherent
- Top line	28.86	5.99	Weak	Straight & Strong
- Rear legs - side view	27.66	4.80	Straight	Sickle-Hocked
- Rear legs - rear view	29.09	5.54	Cow-hocked	Correct
- Foot angle	26.17	4.89	Narrow	Wide
- Pastern	27.02	5.11	Weak	Straight & strong
- Fore udder strength	27.68	5.54	Loose	Tight
- Fore udder length	26.73	5.35	Short	Long
- Rear udder height	26.83	5.41	Short	Tall
- Rear udder width	27.14	6.03	Narrow	Broad
- Suspensory ligament	28.43	5.19	Weak	Strong
- Udder depth	30.47	5.44	Deep	Shallow
- Udder symmetry	24.18	2.75	Not levelled front	Not levelled rear
- Teats position - rear view	23.50	3.58	Far	Close
- Teats Position - side view	26.74	3.92	Far	Close
- Teats length	26.18	5.07	Short	Long
- Front muscularity	28.61	6.13	Scarce	Developed
- Rear muscularity	27.60	5.61	Scarce	Developed
- Head typicality	26.44	6.21	Poor	Very good
Performance test traits:				
- Average daily gain (kg/d)	1.15	0.11	0.74	1.50
- SEUROP score (points)	103.3	4.09	90.0	120.0
- Carcass yield (%)	56.15	1.23	51.0	60.0

Figure 1 reports the factor analysis's main results that indicate a quite clear biological interpretation of latent factors based on loading coefficients with an absolute value greater than 0.45. Indeed, muscularity, udder volume (UV), udder conformation (UC), general aspect, feet correctness, teats, and rear legs (RL) have been identified after varimax rotation pattern as the biological meaning of factors F1, F2, F3, F4, F5, F6, and F7, respectively (Figure 1). These latent factors express an amount of variance of 2.68, 2.38, 2.03, 1.91, 1.62, 1.25, and 1.20, respectively (Table 1). The F1 included the following traits (minimum loading coefficient of |0.45|: strength/robustness with a loading coefficient of 0.86, rear muscularity (0.90), and fore muscularity (0.91). The F2 included some morphological traits regarding udder volume, such as: fore udder length with a loading coefficient of 0.67, and rear udder attaches (height; 0.83; and width; 0.86). The F3 was related to udder conformation and accounted for: fore udder strength (0.64), suspensory ligament (0.66), udder depth (0.81), and udder symmetry (0.47). The F4 contained traits describing the general aspect of the individual: thinness (0.63), shoulders (0.73), dorsal line (0.69), and head typicality (0.58). The F5, accounting for the Leg correctness, comprised pastern and foot angle (both with a loading coefficient of about 0.85). The F6, teats, included traits related to teats evaluation, i.e., teats length (0.76), and both side (0.52) and rear view (-0.54) of teats position, although with opposite sign. Last, F7 accounted for rear legs viewed by side (0.77) and back view (-0.48). F2, F3 and F7 was retained for further genetic analysis. Overall, these three factors accounted for 27.3% of the total variance of traits (Table 2) and allowed a clear interpretation of the latent variables. Each of them was further analyzed as a factor score, which resumes the information of the traits included using a standardized phenotypic variable [5]. In addition to the three factor scores, single linear type traits of rear muscularity (RM) and head typicality (HT) were analyzed.

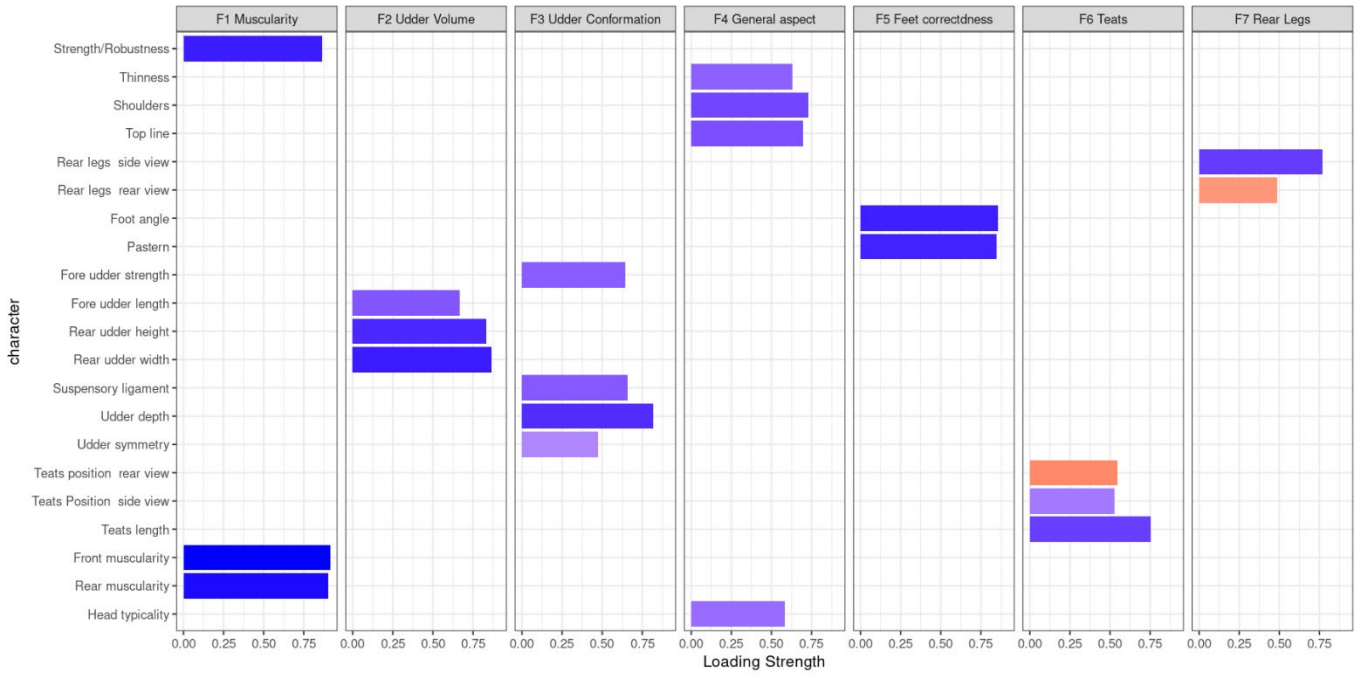
In fact, the first one represents a trait presently under selection, and the second one is a trait in which breeders shows a strong interest because it is part of the breed's typicality: a qualitative assessment of the cranial shape according to the Gray Alpine Herd Book (http://www.grigioalpina.it/wp-content/uploads/2015/10/Norme-tecniche.pdf, April 2021).

Table 2. Variance explained (Var) and percentage of the total variance explained (Var %) by the factors after rotations.

	Variance components[1]				
Analyzed traits	σ^2_a	σ^2_p	h^2	HPD 5[2]	HPD 95[3]
Milk traits:					
- Milk yield	2.211	10.100	0.219	0.181	0.336
- Fat yield	2.600[4]	14.525[4]	0.178	0.117	0.125
- Protein yield	1.895[4]	15.160[4]	0.125	0.166	0.201
- Somatic cell score (SCS, points)	0.379	2.834	0.133	0.119	0.148
Morphological aspects traits:					
- Udder volume factor (F2-UV)	0.244	0.793	0.309	0.254	0.364
- Udder conformation factor (F3-UC)	0.300	0.894	0.325	0.274	0.388
- Rear legs factor (F7-RL)	0.208	0.869	0.238	0.181	0.214
- Head typicality (HT)	13.001	34.601	0.374	0.304	0.417
Beef traits:					
- Rear muscularity (RM)	9.144	27.354	0.328	0.279	0.385
- Average daily gain (ADG, kg/d)	2.631	9.221	0.282	0.094	0.494
- SEUROP (points)	0.529	1.392	0.376	0.184	0.567
- Carcass yield (CY, %)	9.180	18.152	0.501	0.310	0.697

[1] σ^2_a is the additive genetic variance; σ^2_P is the phenotypic variance; h^2 is the heritability
[2] HPD5 is the highest posterior density region at 5%
[3] HPD95 is the highest posterior density region at 95%
[4] Variances have been multiplied by 10^3

Figure 1. Loading coefficient (LC) of individual morphological traits within the seven latent factorsextracted from the factor analysis (i.e., with eigenvalue >1) after the varimax rotation. Only LC ≤−0.45 or ≥0.45 have been reported. Blue bars represent positive loading coefficients, red bars negative loading coefficients.



Genetic parameters and genetic correlations

Table 3 reports the variance components and genetic parameters estimated in single-trait models. Note that not all traits possess all three components, i.e., genetic, environmental and residual: for example, udder volume factor, since it is tested once, does not possess σ^2_{pe} component. Compared to morphological and beef traits, milk traits showed generally lower heritability values, ranging from 0.218 (milk yield) to 0.133 (SCS). Morphological traits had medium-high heritability, i.e., near to 0.30, with head typicality that presents the highest h^2 in this group of traits (0.374), while F3-UC (udder conformation) showed the lowest value of heritability (0.238). Beef traits measured early in life on young bulls at performance testing station showed the highest h^2. SEUROP and CY (carcass yield) reached a value of 0.376 and 0.501, respectively. Conversely, AGD presented a medium heritability value of 0.282.

Table 3. Estimates of variance components and heritability (h^2) of analyzed traits as the means and HPD of the marginal posterior densities. Note that not all traits possess all three components (e.g., udder volume factor does not have a σ^2_{pe}, i.e., permanent environment component).

	Variance components[1]				
Analyzed traits	σ^2_a	σ^2_p	h^2	HPD 5[2]	HPD 95[3]
Milk traits:					
- Milk yield	2.211	10.100	0.219	0.181	0.336
- Fat yield	2.600[4]	14.525[4]	0.178	0.117	0.125
- Protein yield	1.895[4]	15.160[4]	0.125	0.166	0.201
- Somatic cell score (SCS, points)	0.379	2.834	0.133	0.119	0.148
Morphological aspects traits:					
- Udder volume factor (F2-UV)	0.244	0.793	0.309	0.254	0.364
- Udder conformation factor (F3-UC)	0.300	0.894	0.325	0.274	0.388
- Rear legs factor (F7-RL)	0.208	0.869	0.238	0.181	0.214
- Head typicality (HT)	13.001	34.601	0.374	0.304	0.417
Beef traits:					
- Rear muscularity (RM)	9.144	27.354	0.328	0.279	0.385
- Average daily gain (ADG, kg/d)	2.631	9.221	0.282	0.094	0.494
- SEUROP (points)	0.529	1.392	0.376	0.184	0.567
- Carcass yield (CY, %)	9.180	18.152	0.501	0.310	0.697

[1] σ^2_a is the additive genetic variance; σ^2_{pe} is the permanent environmental variance, σ^2_e is the residual variance; h^2 is the heritability.
[2] HPD5 is the highest posterior density region at 5%
[3] HPD95 is the highest posterior density region at 95%
[4] Variances have been multiplied by 10^3

Table 4 showed the genetic and phenotypic correlations between each trait pair considered in the study (full table with HPD is available in Supplementary Material). High genetic correlations (>0.75) were observed within milk traits, except SCS, which was mildly correlated from the genetic point of view with milk yield traits. High genetic correlations were also found between beef traits; mainly, SEUROP and CY presented a genetic correlation greater than 0.90. Regarding the genetic correlations within morphological traits, low values were generally observed, except for some negative correlations between F2-UV and F3-UC (-0.208) and between rear muscularity and F2-UV (-0.319). On the other hand, a medium but positive genetic correlation was obtained between rear muscularity and F3-UC (0.346). Considering the genetic correlations between different groups of traits, F2-UV had positive correlations with all milk yield traits, whereas F3-UC had a negative association with milk yield. F2-UV also had negative correlations with beef traits SEUROP and CY. Another trait related to the beef attitude is the morphological trait Rear muscularity of primiparous cows (RM). This trait had a medium but negative genetic relationship with milk yield traits (from -0.158 to -0.458). Interestingly, despite ADG carcass yields, and SEUROP werestrongly correlated with RM, they were not so negatively correlated as RM with milk yield traits.

Phenotypic correlations follow the same trends as the genetic ones, but their absolute values were slightly lower, especially among different groups' traits.

Table 4. Genetic (above the diagonal), and phenotypic (below the diagonal) correlations among milk traits, SCS, morphological and beef traits analysed. (within brackets). Traits that do not include zero in their HPD are reported in **bold**. Full table with HPD is reported in the Supplementary Material.

TRAITS[1]	MY	FY	PY	SCS	F2 UV	F3 UC	F7 RL	HT	RM	ADG	SEUROP	CY
MY		0.758 (0.732 0.781)	0.845 (0.829 0.860)	0.069 (-0.001 0.141)	0.330 (0.235 0.421)	-0.448 (-0.546 - 0.354)	0.060 (-0.066 0.182)	-0.091 (-0.191 0.011)	-0.458 (-0.547 - 0.365)	-0.071 (-0.362 0.187)	-0.24 (-0.491 0.001)	-0.156 (-0.363 0.051)
FY	0.768 (0.752 0.802)		0.824 (0.804 0.844)	0.067 (-0.008 0.144)	0.286 (0.185 0.383)	-0.326 (-0.429 - 0.224)	0.045 (-0.086 0.175)	-0.136 (-0.241 - 0.033)	-0.413 (-0.514 - 0.316)	-0.092 (-0.339 0.145)	0.029 (-0.196 0.233)	-0.103 (-0.390 0.140)
PY	0.905 (0.850 0.960)	0.766 (0.244 0.247)		0.088 (0.014 0.164)	0.289 (0.188 0.388)	-0.423 (-0.521 - 0.322)	0.099 (-0.031 0.228)	-0.169 (-0.274 - 0.062)	-0.397 (-0.493 0.299)	-0.066 (-0.280 0.139)	0.175 (-0.102 0.433)	-0.156 (-0.385 0.073)
SCS	-0.149 (-0.250 - 0.050)	-0.08 (-0.120 - 0.020)	-0.111 (-0.195 - 0.060)		0.246 (0.120 0.369)	0.149 (0.009 0.284)	0.190 (0.005 0.240)	-0.109 (-0.223 0.004)	-0.158 (0.274 - 0.039)	-0.184 (-0.452 0.043)	-0.008 (-0.219 0.182)	-0.259 (-0.475 - 0.054)
F2-UV	0.24 (0.211 0.244)	0.172 (-0.080 - 0.051)	0.211 (-0.119 - 0.085)	-0.001 (-0.006 0.028)		-0.208 (-0.340 - 0.073)	0.097 (-0.062 0.260)	0.129 (-0.005 0.259)	-0.319 (-0.441 - 0.190)	-0.121 (-0.395 0.144)	-0.351 (-0.536 - 0.129)	-0.359 (-0.736 - 0.040)
F3-UC	-0.122 (-0.137 - 0.101)	-0.067 (-0.001 0.028)	-0.104 (0.003 0.037)	0.012 (-0.006 0.028)	0.003 (0.010 0.054)		0.098 (-0.067 0.255)	0.079 (-0.057 0.209)	0.346 (0.224 0.463)	-0.128 (-0.444 0.157)	0.067 (-0.210 0.334)	0.061 (-0.258 0.381)
F7-RL	0.02 (0.001 0.038)	0.014 (-0.213 - 0.147)	0.021 (-0.238 - 0.157)	0.012 (-0.023 0.056)	0.033 (-0.326 - 0.224)	0.01 (0.243 0.344)		0.075 (-0.053 0.197)	-0.324 (-0.460 - 0.183)	0.148 (-0.176 0.426)	-0.156 (-0.457 0.157)	-0.159 (-0.573 0.213)
HT	-0.02 (-0.093 - 0.002)	-0.012 (-0.072 0.035)	-0.023 (-0.135 0.070)	-0.022 (-0.113 0.011)	0.07 (-0.036 0.014)	0.021 (-0.040 0.016)	0.085 (-0.014 0.040)		0.075 (-0.053 0.197)	-0.189 (-0.477 0.081)	0.208 (-0.051 0.463)	0.171 (-0.129 0.481)
RM	-0.134 (-0.349 - 0.265)	-0.079 (-0.066 0.006)	-0.086 (-0.100 - 0.013)	0.001 (-0.095 - 0.012)	-0.120 (0.114 0.227)	0.128 (-0.006 0.109)	-0.177 (0.151 0.262)	0.085 (0.031 0.080)		0.656 (0.343 0.909)	0.798 (0.556 0.981)	0.849 (0.540 0.991)
ADG	-0.014 (-0.069 0.036)	-0.018 (-0.045 0.053)	-0.015 (-0.021 0.085)	-0.046 (-0.11 0.093)	-0.034 (-0.437 - 0.097)	-0.037 (-0.169 0.264)	0.043 (-0.313 0.111)	-0.064 (-0.045 0.405)	0.182 (0.408 0.719)		0.839 (0.594 0.996)	0.545 (0.156 0.977)
SEUROP	-0.057 (-0.122 0.000)	0.006 (-0.088 0.031)	0.034 (-0.106 0.022)	-0.002 (-0.257 - 0.031)	-0.133 (-0.211 - 0.014)	0.025 (-0.094 0.129)	-0.052 (-0.151 0.068)	0.087 (-0.049 0.183)	0.276 (0.184 0.338)	0.621 (0614 0.741)		0.928 (0.831 0.990)
CY	-0.047 (-0.219 0.034)	-0.024 (-0.088 0.031)	-0.041 (-0.106 0.022)	-0.07 (-0.257 - 0.031)	-0.109 (-0.211 - 0.014)	0.019 (-0.094 0.129)	-0.041 (-0.151 0.068)	0.06 (-0.049 0.183)	0.241 (0.184 0.338)	0.545 (0.994 1.263)	0.825 (1.607 1.788)	

[1]MY = Milk yield; FY = Fat yield; PY = Protein yield; SCS = Somatic cell score; F2-UV = Udder volume factor; F3-UC = Udder conformation factor; RL- F7 = Rear legs factor; HT = Head typicality; RM = Rear muscularity; ADG = Average daily gain; SEUROP = *in vivo* SEUROP score; CY = *in vivo* Carcass yield

Genetic trends and response to different selection scenarios

Table 5 reports the economic weights assigned to each scenario (a) and the weight obtained after restriction to zero(b). In the current selection scheme (S1), great standardized selection responses were obtained for all milk traits (including F2-UV; Figure 2). The protein yield (PY), as expected, had the maximum value (0.47), corresponding to a genetic progress of 0.067 kg of protein per generation (data not shown). All other traits (morphological and beef)

showed a negative genetic trend (Figure 2; S1). F3-UC presented the worst selection response (-0.32). ADG and SEUROP resulted in the only beef traits with non-negative selection response (0.29 and 0.05, respectively).

When restrictions were applied to SCS (S2), corresponding to null genetic progress for this trait, the expected genetic gain for milk traits slightly declined: for PY from 0.47 to 0.40 and from 0.32 to

0.29 for FY. About morphological traits, F3-UC showed a further decrease from -0.32 to -0.36. Similarly, RM resulted in less negative genetic variation (from -0.16 to -0.04 standardized units) than in S1, almost null genetic progress. In the scenario in which restriction was applied on RM (S3) a similar situation than in S2 was seen, although with a small reduction of standardized genetic gain for milk and a small increase for all beef traits was observed, i.e., a non-negative variation for RM, positive increase for ADG, SEUROP, and CY (Figure 2, S3). On the other hand, the expected genetic progress for all morphological traits remained almost unchanged with respect S1 and S2. Indeed, F2-UV still showed a small increase, F3-UC remained negative, such as HT and F7-RL, although with a lower magnitude than for F3-UC. In S4 and S5, despite the small reduction of fat and protein yields and beef traits in favor of morphological traits, a small but favorable increase of FY and PT was still observed, because of the lower incidence of beef traits, but SCS increased negatively. However, in both scenarios, F3-UC and HT genetic gain resulted less negative than in the previous ones (Figure 2), while F7-RL showed a small negative increase. In S5, notably, RM resulted non-negative. In S6 and S7, there was a further reduction of milk traits favoring beef traits but maintaining the same weights for morphological traits as in S4 and S5. Thus, both FY and PY were reduced, and beef traits increased compared to the previous scenarios. S6 and S7 showed the best genetics progress for beef traits (>0.20, >0.30, >0.45, and >0.40 for RM, ADG, SEUROP and CY, SEUROP, respectively; Figure 2). These were the first scenarios in which RM had a positive selection response. On the contrary, milk traits presented the worse genetic gain as compared to other scenarios. In both S6 and S7, the SCS increased negatively, but slowly than in S4 and S5. Response to selection of F3-UC was less negative than in the previous scenarios, while F7-RL resulted more negatively affected (Figure 2). Last, in S8, milk yield genetic gain was maintained at the same level as in S6 and S7, but morphological traits resulted in a non-negative variation, and beef traits improved, particularly RM in primiparous cows. Also,

SEUROP and CY of young performance-tested bulls showed a positive increase (Figure 2). In general, F2-UV, not weighted in any scenario, followed the same trend as milk yield, i.e., increasing when milk yield increased, showing a selection response over to 0.20 standardized units. Notwithstanding, reducing milk traits' economic weight has contributed to an adverse selection response for F2-UV in scenarios S6, S7, and S8.

Table 5. Economic weights of traits as applied before (a) and after (b) the restriction for the genetic progress of target traits1. The sum to 1 of the economic weights of traits considers the absolute values of the weights.

a)

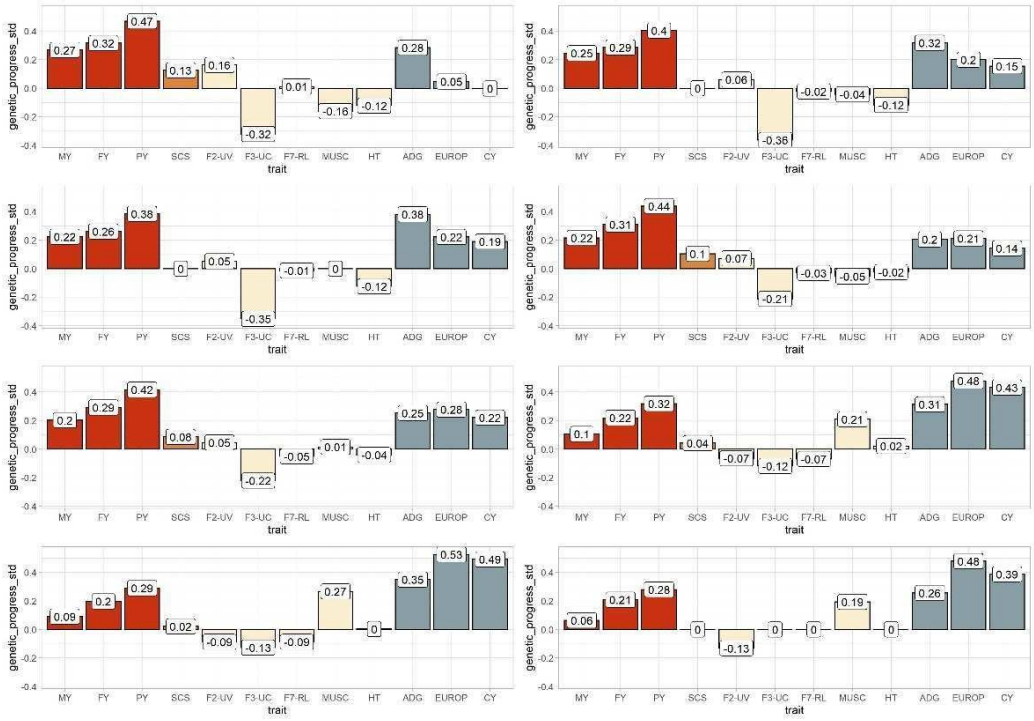
Scenario	MY	FY	PY	SCS	F2-UV	F3-UC	F7-RL	HT	RM	ADG	SEUROP	CY	Milk[2]	Morph.[3]	Beef[4]
S1	0	0.24	0.46	0	0	0	0	0	0.1	0.2	0	0	0.7	0	0.3
S2	0	0.24	0.46	0[5]	0	0	0	0	0.1	0.2	0	0	0.7	0	0.3
S3	0	0.24	0.46	0[5]	0	0	0	0	0[5]	0.3	0	0	0.7	0	0.3
S4	0	0.217	0.433	0	0	0.07	0	0.03	0.15	0	0.05	0.05	0.65	0.1	0.25
S5	0	0.217	0.433	0	0	0.035	0	0.015	0.2	0	0.05	0.05	0.65	0.05	0.3
S6	0	0.18	0.37	0	0	0.07	0	0.03	0.15	0	0.1	0.1	0.55	0.1	0.35
S7	0	0.18	0.37	0	0	0.035	0	0.015	0.2	0	0.1	0.1	0.55	0.05	0.4
S9	0	0.24	0.46	0[5]	0	0[5]	0[5]	0[5]	0.2	0	0.05	0.05	0.7	0	0.3

b)

Scenario	MY	FY	PY	SCS	F2-UV	F3-UC	F7-RL	HT	RM	ADG	SEUROP	CY	Milk[2]	Morph.[3]	Beef[4]
S1	0	0.24	0.46	0	0	0	0	0	0.1	0.2	0	0	0.7	0	0.3
S2	0	0.186	0.356	-0.225[5]	0	0	0	0	0.077	0.155	0	0	0.768	0	0.232
S3	0	0.175	0.335	-0.220[5]	0	0	0	0	0.052[5]	0.218	0	0	0.730	0	0.270
S4	0	0.217	0.433	0	0	0.07	0	0.03	0.15	0	0.05	0.05	0.65	0.1	0.25
S5	0	0.217	0.433	0	0	0.035	0	0.015	0.2	0	0.05	0.05	0.65	0.05	0.3
S6	0	0.18	0.37	0	0	0.07	0	0.03	0.15	0	0.1	0.1	0.55	0.1	0.35
S7	0	0.18	0.37	0	0	0.035	0	0.015	0.2	0	0.1	0.1	0.55	0.05	0.4
S8	0	0.153	0.294	-0.134[5]	0	0.160[5]	0.052[5]	0.015[5]	0.128	0	0.032	0.032	0.581	0.227	0.191

[1]Traits: MY = Milk yield; FY = Fat yield; PY = Protein yield; SCS = Somatic cell score; F2-UV = Udder volume factor; F3-UC = Udder conformation factor; RL- F7 = Rear legs factor; HT = Head typicality; RM = Rear muscularity; ADG = Average daily gain; SEUROP = in vivo SEUROP score; CY = in vivo Carcass yield 2Milk traits: MY, FY, PY, SCS; 3Morphological traits: F2-UV, F3-UC, F7-RL, HT; 4Beef traits: RM, ADG, SEUROP, CY; 5Restriction applied to target traits

Figure 2. Standardized genetics progress (y axes) for 12 traits studied considering 8 different scenarios (from S1 to S8) attributing different weights to specific traits in the possible selection index. Red bars represent milk, fat, and protein, orange bars SCS, light-yellow morphological traits (both factors and single trait analyzed), and grey bars beef traits (type rear muscularity or performance test traits). Traits' abbreviations are reported in Table 1.



DISCUSSION

Heritability

The Grey Alpine represents a perfect example of a local breed with a dual-purpose attitude, as it shows good productive performances for both milk and beef traits. In our study heritability of milk yield was lower than other traits analyzed, but this is because milk yield was analyzed as test-day records, that are recognized to give lower heritability, because of the high environmental variance. Moreover, it is also due to the nature of the data, i.e., longitudinal observations instead of data recorded once in life. Fat yields showed lower heritability than protein yields, although both PY and FY had about the same additive genetic variance. However, FY's residual variance was almost double compared to that of PY, reducing the heritability. Many studies have reported that fat is much more affected than protein by external

factors, such as the feeding regimen [17,18], and thatis in agreement with the greater residual variance for FY than for PY. On the other hand, the Grey Alpine showed heritability valued for milk traits similar to those reported for other dual propose/local breeds, as Italian Simmental, Rendena, and Valdostana. The Italian Simmental presented a heritability value of 0.18, 0.13, and 0.17 for milk, fat, and protein yields, respectively [19]. In Rendena local breed, heritability levels for these traits were 0.188, 0.157, and 0.165 [6], and similar values were also found in Valdostana breed, for which heritability estimates were 0.198, 0.132, and 0.169 [20]. In general, heritability for milk traits resulted slightly greater than in specialized breeds, like Holstein (e.g., 0.108 for MY in Italian Holstein; [21]). This could be related to the fact that in dual-purpose cattle, milk traits have been subjected to less selective pressures over time than dairy cattle [3].

Although many studies considered SCS as a low heritability trait (h^2 = 0.08 on average), especially in Holstein cows [22], a slightly greater value of 0.133 was found in this study. Still, the lower selection pressure could be identified as the possible cause of such greater than expected estimates. However, for other local cattle breeds of the Alpine area, Rendena, and Valdostana (Aosta Chestnut), heritability estimates were closer to those observed in cosmopolitan breeds, i.e., 0.08 [20].

Factor analysis allowed to characterize any factor with an explicit biological meaning due to the orthogonalization of loading coefficients, performed by varimax rotation, that maximizes factor independence [12]. According to [23], factor loadings are one of the best approaches for selection when a lot of different traits can be easily combined because of their collinearity. Factor loadings indeed allow summarizing the information from a multi-trait analysis by concentrating the traits into single information, avoiding the use of highly correlated measures. In this study, F2 was entirely explained by all the udder attaches (length of the fore udder and length and width of the rear udder); a greater value of the factor loading indicates a wider dimension of udder attach, directly linking the factor to the volume of the udder. The F3 included the other udder traits connected to udder “health”, like the strength of the suspensory ligament, the udder depth, and the udder symmetry. These three traits describe the mammary apparatus's conformation and assume an increasing value of F3 with an increased score of the three traits. Last, F7 describes the posterior rear legs conditions, including the side and the back view of rear legs. These two traits, characterized by

intermediate optimum values, entered the factor with opposite sign (i.e., positive the rear legs side view, and negative the rear legs back view), because the sickle-hocked defect, associated with a greater score, is often associated to the cow-hocked defect, assuming the lowest scores in Alpine Grey morphological evaluation (www.grigioalpina.it/?lang=en, Date of access: 30 March 2021).

Overall, morphological traits showed a medium-high value of h^2, as widely reported in the literature[24-26], including dual-purpose cattle [5]. Recent studies [27,28] demonstrated that the higher heritability values could be due to the greater number of gene clusters involved in biological processes relevant for udder morphology.

A wide range of studies has been carried out for morphological traits in specialized breeds, and generally, the heritability resulted lower than those estimated in this research. Regarding morphological traits evaluated in dairy cattle (udder volume, udder conformation also the leg), heritability values of 0.14, 0.08, and 0.07 were found in Holstein; also, a 0.18 for udder volume was reported [30,31] and as a mean of 0.22 for other traits regarding udder conformation [30,31]. For beef conformation traits, an heritability of 0.40 has been reported in both Brown Swiss and Red &White breeds [32], but in Italian dual purpose breeds values similar to those of the present study were found. For muscularity, udder volume and udder conformation heritabilities of 0.314, 0.166, and 0.169 were found in Valdostana [5], whereas Rendena showed higher heritability values of 0.359, 0.260, 0.267, possibly due to the different nature of the factorial score in these breeds as compared to the Alpine Grey. On the other hand, in beef cattle, similar heritability estimates have been reported for head typicality [25,26].

The high heritability estimates observed in this study for performance test traits compared to the morphological and milk traits were commonly observed even in other dual-purpose or beef breeds, like the Piedmontese (e.g., heritability of 0.47 for ADG) [6,32–35].

Genetic correlations

expected, milk yield traits shoed strong phenotypic and genetic correlations among them, as both protein and fat productions depend on the amount of milk produced. Somatic cell score showed a low-positive genetic correlation with milk traits, confirming previous findings [36], where the independence of traits was demonstrated by genomic analysis indicating the

presence of different genes and loci under the traits [37,38]. Udder volume (F2) showed a positive correlation with all milk traits, including SCS (a positive correlation was also reported in other studies [39]). On the contrary, F3 (udder conformation) presented a negative correlation with F2, and consequently, with all other milk yield traits. The genetic improvement for milk production leads to an indirect increase of udder volume that causes damage to its conformation. Similarly [6,30,31] found a negative genetic correlation of about -0.3 between udder conformation and udder volume for Italian Brow Swiss, Rendena and Valdostana cattle. An impressive result was discovered by analyzing the genetic correlation of SCS with F2-UV and F3-UC, which resulted positive, suggesting a detriment in udder health for increasing udder volume and conformation values. A similar result was also found in Rendena breed for udder volume [5], but it was the opposite for udder conformation. Rear legs (F7) and head typicality had genetic correlations not different from zero, either with milk or beef traits, considering that in all cases, the HPD95% included zero. The only positive correlation was observed between SCS and F7-RL, meaning that an increase in inflammatory udder status is associated with an impairment of rear legs and possibly a general unhealthy animal status [40]. Milktraits and F2-UV had a negative correlation with rear muscularity, whereas the positive medium correlation between F3-UC and RM can be explained considering the negative correlations that both these traits showed with milk yield because they both resemble a typical aspect observable more frequently in muscular cows. Positive correlations between muscularity and udder correctness have been previously reported [5]. Muscularity scored in cows showed a strong positive genetic correlation with all performance test traits, which agrees with other studies [6,10,4], despite a negative genetic correlation sometimes found between muscularity and ADG [33,41] The negative correlations between milk and beef traits have been identified in the different asset of genes involved in metabolism regulation, catabolism of collagen, and myogenesis compared to milk synthesis [42]. Other studies have reported this negative genetic correlation, estimating a similar correlation coefficient to our study, e.g., in Brown Swiss and Swiss Simmental cattle [39] or Italian Simmental [40]. On the other hand, slightly lower negative genetic correlations were found by Croué et al. [43], comparing the postmortem SEUROP with milk, fat, and protein yields in French dual-purpose cattle breeds (Montbeliarde, Normande, and Simmental).

Genetic response under different selection scenarios

Multivariate response to selection was calculated to properly account for different traits in the aggregate selection index of the breed. This index is made by assigning a different economic weightto EBVs of each target trait [44].

A proper knowledge of the true genetic relationships among target traits, and of the expected response under different selection pressures can help properly drive selection decisions and therefore the genetic trend of traits.

The present selection scheme (S1) produces the greatest growth of milk traits in terms of standardized genetic progress due to the high weight accounted by fat and protein yields in the selection index (70% of total). The positive selection response for milk yield is due to the favorable and strong genetic correlations with the FY and PY. The present scheme produces a negative increase of the SCS, negatively affect udder conformation, the muscularity measured in primiparous cows, and head typicality. S1 produces also a positive increase in udder volume, ADG measured on bulls, and, to a less extent, SEUROP scores on young bulls. A steady-state is detectable for rear legsand estimated CY on performance-tested young bulls.

The considerable selection response attainable for F2-UV, despite not being directly accounted for in the selection index in S1, is due to the strong genetic correlation of this composite trait with milk yield traits, since a large udder volume allows an increased milk production. In the current selection index, all the beef traits except ADG showed a negative (RM) to almost null (SEUROP and CY) selection response, because the economic weight for beef attitude was attributed only to ADG (i.e., 30%). Overall, the present selection index does not reflect the goal of selection for the dual-purpose attitude in the Alpine Grey cattle breed. A different situation was observed in the current selection response of another Italian dual-purpose cattle, the Rendena, in which an economic weight for beef attitude is due to all the traits accounted for in this study [5]. Nevertheless, negative response forRM is also produced in this breed due to the strong antagonistic correlations with milk traits. The positive response for SCS, observed in this study as also in [5] is undesirable for selection since it means a detriment in udder health.

The restriction for maintaining unchanged SCS in S2 produces a slightly negative effect on milk yields and slightly increases the response for the beef traits measured in performance-tested bulls. Despite this, the negative genetic correlations between muscularity and milk traits still led to an adverse effect of RM selection, limiting a proficient selection for the breed's dual-purpose attitude. Despite the neutral selection response for SCS, udder conformation (F3-UC) worsens with respect to S1, highlighting the need for further investigations on these traits that are both indirectly linkedto udder health.

When a restriction toward unchanged SCS and muscularity was analyzed (S3), a decline in milk yield progress was observed, underscoring the importance to reduce milk yield growth despite a selection goal more oriented toward the dual-purpose attitude. Although CY and SEUROP in young bulls were not directly selected, an increase in the standardized genetic response for these traits is detectable because of the positive correlation with muscularity.

In the fourth scenario (S4), an increase in milk yield response is observed, despite a 5% reduction of its weight, but beef trait response was like S2 and S3. However, SCS and morphology generally showed a negative response to selection, slightly more negative in the fifth scenario (S5), where they received a lesser economic weight, in favor of beef traits. Such traits on the other hand increased, particularly muscularity in primiparous cows which was slightly positive for the first time. In subsequent scenarios, milk yield was negatively affected due to the reduced economic weight in the selection index (as in S6 and S7) or the introduction of constraints in non-negative growth of traits negatively correlated to milk, fat, and protein yields. In the last three scenarios, muscularity was greatly increased as compared to the previous scenarios. S6 and S7 were the scenarios in which beef had more emphasis in the selection process. On the other hand, S8, notwithstanding the 70% of the weight on milk traits, produced results like S6 and S7, due to the greater incidence of beef and morphology. These latter scenarios should be considered technically more balanced toward the dual-purpose attitude, although they could be considered not as economically convenient due to the high commercial value of milk as compared to other traits.

In all scenarios proposed, the udder volume (F2) was always not directly selected due to its high correlation with fat and protein, and milk yield. Nevertheless, this trait showed a negative selection response in the last three scenarios due to the negative correlations with

beef traits. Regarding other morphological traits (F7-RL and HM), following breeders' suggestions, it could be important to ensure a non-negative or slightly negative genetic progress for these traits, as observed in most cases. Regarding performance test traits, high economic weight for ADG, as in S1, is meaningless. ADG has a less economic interest than CY and SEUROP (ANAGA, personal communication) that are also less negatively correlated with milk traits. If non-negative genetic progress for SCS, but especially for F3-UC and RM, is considered a priority, a necessary reduction of milk traits' genetic progress occurs.

In conclusion, due to the complex structure of genetic correlations among traits and the large number of negative genetic correlations, a selection index including all the important aspects for thebreed (milk, beef, and morphology) can be considered the best compromise. As expected, milk and beef traits have negative genetic correlation, and the situation becomes more complicated if morphological traits are included in the selection index.

However, the present selection index (S1) produces a detriment of beef attitude in the medium-long term and a loss of some peculiar and important functional characteristics in the breed. For these reasons, a selection index more oriented toward the beef attitude, without worsening some morphological characteristics appreciated by breeders, can be considered more appropriate, despite the reduction of the expected response for milk yield.

CONCLUSIONS

In conclusion, due to the complex structure of genetic correlations among traits and the large number of negative genetic correlations listed above, selection index including all the important aspects for the breed (milk, beef, and morphology) can be considered the best compromise. As expected, milk and beef traits have negative genetic correlation, and the situation became more complicated if morphological traits are included in the selection index. However, the present selection index (S1) produces a detriment of beef attitude in the medium-long term and a loss of some peculiar and important functional characteristics in the breed. For these reasons, a selection index more oriented toward the beef attitude, without worsening some morphological characteristics appreciated by breeders, can be considered more appropriate, despite the reduction of the expected response for milk yield.

The best scenario cannot be uniquely identified, because the economic values of a standard deviation of different traits are not equally predictable. This is particularly true for morphological traits for which intrinsic economic value is often hard to measure (i.e., F7), or it has a null value, but it is of great importance for maintaining the typicality of the breeds (i.e., HT). In this regard, the authors however suggest scenario 7 as the most suitable for selection in the alpine gray breed. In fact, scenario 7 allows a slight genetic progress for both productive traits (i.e, milk and meat), while preserving the dual attitude of the breed. Furthermore, this scenario guarantees the maintenance of the functional characteristics of this breed.

Author Contributions: Conceptualization, R.M.; methodology, R.M., E.M.; formal analysis, E.M.; investigation, C.S. and E.M..; resources, R.M.; data curation E.M. and N.G.; writing— original draft preparation, E.M.; , B.T writing—review and editing R.M., C.S, B.T. All authors haveread and agreed to the published version of the manuscript.

Funding: This research was funded by the Alpin Grey National Breeders Association (ANAGA) within the DUALBREEINDG project (CUP J51J18000000005)

Acknowledgments: Authors would like to thank the technicians of ANAGA breeders' associations for data

Conflicts of Interest: The authors declare no conflict of interest.

References

1.Tisdell, C. Socioeconomic causes of loss of animal genetic diversity: Analysis and assessment. Ecol. Econ. **2003**, 45, 365–376, doi:10.1016/S0921-8009(03)00091-0.

2.Krupová, Z.; Krupa, E.; Michaličková, M.; Wolfová, M.; Kasarda, R. Economic values for health and feed efficiency traits of dual-purpose cattle in marginal areas. J. Dairy Sci. **2016**, 99, 644–656, doi:10.3168/jds.2015-9951.

3.Gandini, G.C.; Villa, E. Analysis of the cultural value of local livestock breeds: A methodology. J. Anim. Breed. Genet. **2003**, 120, 1–11, doi:10.1046/j.1439-0388.2003.00365.x.

4.Hoffmann, I. Adaptation to climate change--exploring the potential of locally adapted breeds.Animal **2013**, 7 Suppl 2, 346–362, doi:10.1017/S1751731113000815.

5.Mazza, S.; Guzzo, N.; Sartori, C.; Mantovani, R. Genetic correlations between type and test-day milk yield in small dual-purpose cattle populations: The Aosta Red Pied breed as a case study. J. Dairy Sci. **2016**, 99, 8127–8136, doi:10.3168/jds.2016-11116.

6.Sartori, C.; Guzzo, N.; Mazza, S.; Mantovani, R. Genetic correlations among milk yield, morphology, performance test traits and somatic cells in dual-purpose Rendena breed. Animal **2018**, 12, 906–914, doi:10.1017/S1751731117002543.

7.Forabosco, F.; Mantovani, R.; Meneghini, B. European and Indigenous Cattle Breeds in Italy; Schiel & Denver Publishing Limited, **2011**; ISBN 9781849030748.

8.Ali, A.K.A.; Shook, G.E. An Optimum Transformation for Somatic Cell Concentration in Milk. J. Dairy Sci. **1980**, 63, 487–490, doi:10.3168/jds.S0022-0302(80)82959-6.

9.Mantovani, R.; Cerchiaro, I.; Contiero, B. Factor analysis for genetic evaluation of linear type traits in dual purpose breeds. Ital. J. Anim. Sci. **2005**, 4, 31–33, doi:10.4081/ijas.2005.2s.31.

10. Guzzo, N.; Sartori, C.; Mantovani, R. Analysis of genetic correlations between beef traits in young bulls and primiparous cows belonging to the dual-purpose Rendena breed. animal **2019**, 13, 694–701, doi:10.1017/S1751731118001969.

11. Revelle, W. psych: Procedures for Psychological, Psychometric, and Personality Research **2016**.

12. Kaiser, H.F. The varimax criterion for analytic rotation in factor analysis. Psychometrik **1958**, 23, 187–200, doi:10.1007/BF02289233.

13. Misztal, I.; Tsuruta, S.; Lourenco, D.; Aguilar, I.; Legarra, A.; Vitezica, Z. Manual for BLUPF90 family of programs. Univ. Georg. Athens, USA **2018**, 125.

14. Careau, V.; Wolak, M.E.; Carter, P.A.; Garland, T. Limits to behavioral evolution: The quantitative genetics of a complex trait under directional selection. Evolution (N. Y). **2013**, 67, 3102– 3119, doi:10.1111/evo.12200.

15. Lande, R. Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry. Evolution (N. Y). 1979, 33, 402, doi:10.2307/2407630.

16. Kause, A.; Mikkola, L.; Strandén, I.; Sirkko, K. Genetic parameters for carcass weight, conformation and fat in five beef cattle breeds. Animal **2014**, 9, 35–42, doi:10.1017/S1751731114001992.

17. Mousseau, T.A.; Roff, D.A. Natural selection and the heritability of fitness components. Heredity (Edinb). **1987**, 59, 181–197, doi:10.1038/hdy.1987.113.

18. Van Soest, P.J. Ruminant Fat Metabolism with Particular Reference to Factors Affecting Low Milk Fat and Feed Efficiency. A Review. J. Dairy Sci**. 1963**, 46, 204–216, doi:10.3168/jds.S0022- 0302(63)89008-6.

19. GURR, M.I. Factors affecting the composition of cow's milk. Nutr. Bull. **1985**, 10, 139–152,doi:10.1111/j.1467-3010.1985.tb01206.x.

20. Sartori, C.; Guzzo, N.; Mantovani, R. Genetic correlations of fighting ability with somatic cells and longevity in cattle. *Animal* **2020**, *14*, 13–21, doi:10.1017/S175173111900168X

21. Frigo, E.; Samorè, A.B.; Vicario, D.; Bagnato, A.; Pedron, O. Heritabilities and genetic correlations of body condition score and muscularity with productive traits and their trend functions in Italian Simmental cattle. Ital. J. Anim. Sci. **2013**, 12, 240–246, doi:10.4081/ijas.2013.e40.

22. Kheirabadi, K.; Razmkabir, M. Genetic parameters for daily milk somatic cell score and relationships with yield traits of primiparous Holstein cattle in Iran. J. Anim. Sci. Technol. **2016**, 58, 1– 6, doi:10.1186/s40781-016-0121-5.

23. Russell, D.W. In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. Personal. Soc. Psychol. Bull. **2002**, 28, 1629–1646, doi:10.1177/014616702237645.

24. Battagin, M.; Sartori, C.; Biffani, S.; Penasa, M.; Cassandro, M. Genetic parameters for body condition score, locomotion, angularity, and production traits in Italian Holstein cattle. J. Dairy Sci. **2013**, 96, 5344–5351, doi:10.3168/jds.2012-6352.

25. Gutiérrez, J.P.; Goyache, F. Estimation of genetic parameters of type traits in Asturiana de los Valles beef cattle breed. J. Anim. Breed. Genet. **2002**, 119, 93–100, doi:10.1046/j.1439- 0388.2002.00324.x.

26. Mantovani, R.; Cassandro, M.; Contiero, B.; Albera, A.; Bittante, G. Genetic evaluation of type traits in hypertrophic Piemontese cows. J. Anim. Sci. **2010**, 88, 3504–3512, doi:10.2527/jas.2009- 2667.

27. Marete, A.; Lund, M.S.; Boichard, D.; Ramayo-Caldas, Y. A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. PLoS One **2018**, 13, 1–17, doi:10.1371/journal.pone.0199931.

28. Chessa, S.; Nicolazzi, E.L.; Nicoloso, L.; Negrini, R.; Marino, R.; Vicario, D.; Ajmone Marsan, P.; Valentini, A.; Stefanon, B. Analysis of candidate SNPs affecting milk and

functional traits in the dual-purpose Italian Simmental cattle. Livest. Sci. **2015**, 173, 1–8, doi:10.1016/j.livsci.2014.12.015.

29. Olasege, B.S.; Zhang, S.; Zhao, Q.; Liu, D.; Sun, H.; Wang, Q.; Ma, P.; Pan, Y. Genetic parameter estimates for body conformation traits using composite index, principal component, and factor analysis. J. Dairy Sci. **2019**, 102, 5219–5229, doi:10.3168/jds.2018-15561.

30. Dube, B.; Dzama, K.; Banga, C.B.; Norris, D. An analysis of the genetic relationship between udder health and udder conformation traits in South African Jersey cows. Animal **2009**, 3, 494–500, doi:10.1017/S175173110800390X.

31. De Haas, Y.; Janss, L.L.G.; Kadarmideen, H.N. Genetic and phenotypic parameters for conformation and yield traits in three Swiss dairy cattle breeds. J. Anim. Breed. Genet. **2007**, 124, 12– 19, doi:10.1111/j.1439-0388.2007.00630.x.

32. Jensen, J.; Mao, I.L.; Andersen, B.B.; Madsen, P. Genetic parameters of growth, feed intake, feed conversion and carcass composition of dual-purpose bulls in performance testing. J. Anim. Sci. **1991**, 69, 931–939, doi:10.2527/1991.693931x.

33. Aass, L. Variation in carcass and meat quality traits and their relations to growth in dual purpose cattle. Livest. Prod. Sci. **1996**, 46, 1–12, doi:10.1016/0301-6226(96)00005-X.

34. Sbarra, F.; Mantovani, R.; Bittante, G. Heritability of performance test traits in Chianina, Marchigiana and Romagnola breeds. Ital. J. Anim. Sci. **2009**, 8, 107–109, doi:10.4081/ijas.2009.s3.107.

35. Bonfatti, V.; Albera, A.; Carnier, P. Genetic associations between daily BW gain and live fleshiness of station-tested young bulls and carcass and meat quality traits of commercial intact males in Piemontese cattle. J. Anim. Sci. **2013**, 91, 2057–2066, doi:10.2527/jas.2012-5386.

36. Karacaören, B.; Jaffrézic, F.; Kadarmideen, H.N. Genetic parameters for functional traits in dairy cattle from daily random regression models. J. Dairy Sci. **2006**, 89, 791–798, doi:10.3168/jds.S0022-0302(06)72141-5.

37. Meredith, B.K.; Berry, D.P.; Kearney, F.; Finlay, E.K.; Fahey, A.G.; Bradley, D.G.; Lynn, D.J. A genome-wide association study for somatic cell score using the Illumina high-density bovine beadchip identifies several novel QTL potentially related to mastitis susceptibility. Front. Genet. **2013**, 4, 1–10, doi:10.3389/fgene.2013.00229.

38. Short, T.H.; Lawlor, T.J. Genetic Parameters of Conformation Traits, Milk Yield, and Herd Life in Holsteins. J. Dairy Sci. **1992**, 75, 1987–1998, doi:10.3168/jds.S0022-0302(92)77958-2.

39. Samoré, A.B.; Rizzi, R.; Rossoni, A.; Bagnato, A. Genetic parameters for functional longevity, type traits, somatic cell scores, milk flow and production in the Italian Brown Swiss. Ital. J. Anim. Sci. **2010**, 9, 145–152, doi:10.4081/ijas.2010.e28.

40. Ptak, E.; Jagusiak, W.; Zarnecki, A. 59th Annual Meeting of the European Association for Animal Production Vilnius, Lithuania – August 24-27, **2008** Session 15. Free communications in Animal Genetics Relationship between test day somatic cell score and conformation traits in Polish Holstein. 2008.

41. Frigo, E.; Samorè, A.B.; Vicario, D.; Bagnato, A.; Pedron, O. Heritabilities and genetic correlations of body condition score and muscularity with productive traits and their trend functions in Italian Simmental cattle. Ital. J. Anim. Sci. **2013**, 12, 240–246, doi:10.4081/ijas.2013.e40.

42. Maiorano, A.M.; Lourenco, D.L.; Tsuruta, S.; Toro Ospina, A.M.; Stafuzza, N.B.; Masuda, Y.; Filho, A.E.V.; Dos Santos Goncalves Cyrillo, J.N.; Curi, R.A.; De Vasconcelos Silva, J.A. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. PLoS One **2018**, 13, 1–24, doi:10.1371/journal.pone.0200694.

43. Croué, I.; Fouilloux, M.N.; Saintilan, R.; Ducrocq, V. Carcass traits of young bulls in dual- purpose cattle: Genetic parameters and genetic correlations with veal calf, type and production traits. Animal **2017**, 11, 929–937, doi:10.1017/S1751731116002184.

44. Hazel, L.N. The Genetic Basis for Constructing Selection Indexes. *Genetics* **1943**, *28*, 476– 490.

7. GENETIC SELECTION IN LOCAL REGGIANA BREED: ESTIMATION OF GENETIC PARAMETER AND PHENOTYPE PLASTICITY FOR MAIN PRODUCTIVE AND REPRODUCTIVE TRAITS

STATUS: ON SUBMISSION

Genetic Selection in Local Reggiana Breed: estimation of genetic parameter and phenotype plasticity for main productive and reproductive traits

INTRODUCTION

Local breeds refer to animals related to a limited territory (Derrouch et al., 2016). These breeds may present peculiar characteristics due to the adaptation to a specific environment (Bertolini et al., 2020). In recent decades there has been a general increase in awareness of the role of indigenous breeds, with a consequent increase in research conducted on the latter. (Sechi et al., 2007; Mastrangelo et al., 2017; Senczuk et al., 2020; Mancin et al., 2022). These studies revealed the great diversity and variability in terms of genes and biological pathways involved, in particular they have identified many genes implicated in disease resistance or associated with more general "adaptability" traits. (Ben-Jemaa et al., 2021; Mancin et al., 2022). Indeed, preserving genetic heritage of local breed can be particularly useful to ensure food security in an ever-changing environment (Boudalia et al., 2020) thanks to the more “genetic adaptability”. Furthermore, native breeds have a fundamental role into preserving the local ecosystem, thanks to the generation of ecosystem and socio-cultural services (Hiemstra et al., 2010; Ovaska and Soini, 2017; Leroy et al., 2018). However, since 80’s an unprecedented deterioration of local breeds animals have been observed (FAO and Platform for Agrobiodiversity Research, 2011), mainly due with the substitution of those breeds with more productive and specialized ones (Gandini et al., 2010). This can be prevented by increasing farm profitability through new strategies, such as enhancement of the farming system in a low-input context or through the enhancement of typical products (Gandini et al., 2010). The Reggiana cattle, a local breed raised on Reggio Emilia province, is strictly example. In fact, starting from the second post-war period, Reggiana suffered a drastic demographic decline, which led to only 800 head in the 1980s (www.razzareggiana.it, update: 26 December 2021). However, this negative trend stopped around the 90s, when the of high-quality single-breed cheese (*Parmigiano Reggiano delle Vacche Rosse*) has been created. Indeed, the high profitability of that product has ensured over time a sustained price for milk that compensated the lower productivity of this breed (www.razzareggiana.it, update: 26 December 2021). Since 1990 the population is progressively increase up to the 3,000 heads present today. Despite

this, an adequate breeding program on Reggiana is necessary to ensure competitiveness of that breeds, where along with increase in the productive traits, a maintenance of the original rusticity must be considered (Biscarini et al., 2015). An adequate breeding plan for local breeds should also consider the impact of the Genotype by Environment interaction (GxE) as native breeds are usually read in heterogenous environmental, i.e. mountain vs plans or herd with higher discrepancy in terms of technological inputs. Therefore, in case of a high level of GxE, the currents breeding plans should be re-designed based on different production systems (Beat) Furthermore, the GxE analysis can be useful to deepen the knowledge of phenotypic plasticity on local cattle as it has been hypothesized that the latter have greater resilience, however few studies have been conducted. Indeed, the GxE, or phenotype plasticity indicates the ability of an organism to cope in different environment.

Specifically, it represented the capability of genotype to change the phenotypic expression when theorganism is exposed to different environment (Huang et al., 2020), In dairy cattle, the GxE covers an important part of quantitative variation on the main productive and morphological traits (Tait- Burkard et al., 2018), while it seems slightly affect reproductive (Zhang et al., 2019) and longevity traits (Mwansa and Peterson, 1998). However, the amount of additive variance expressed by GxE for specific traits depend by many factors, as the type of environment descriptor considered and by the models used. For example, multivariate models are commonly used for series of character states (e.g., presence of silage or not) while in presence of continuous gradient (e.g., herd milk production) the reaction norms models are mostly used (RNN). The RNN is particularly suitable for dairy cattle, in which sire have large number of daughters that are in turn that raised in a variety of herd environments (Rauw and Gomez-Raya, 2015). The amount of additive variance expressed by GxE change also according to the type of breeds considered. Recently some studies (Toledo- Alvarado et al., 2017) compared the impact of less specialized breeds as Simental or Alpine Grey with more specialized ones (Holstein or Brow Swiss), and demonstrates that less specialized breeds are fewer sensitive to changes in herd productivity. (Martinez-Castillero et al., 2020), reported similar results for fertility traits using a bit-traits models based on herd productivity. However only few studies evaluated the impact of RNN on those breeds (Schmid et al., 2021) Sartori et al., 2022 in press. The estimation of environmental plasticity or adaptability of these local breeds can be a further process of enhancement of the latter, especially in a focused agriculture increasingly interested in collaborating with climate change (Mulder, 2016). Despite

this, estimated GxE is also fundamental since it may improve may increase the accuracy of environment-specific breeding values. In addition, is interesting to calculated the GxE it quote since may reduce the genetic progress of specific traits (Mulder and Bijma. 2007). Therefore, with an aims to developed a suitable selection plans for Reggiana breeds in that study we estimated variance components of milk and fertility traits and his environmental plasticity using as covariate the herd environment.

MATERIAL AND METHODS:

Data editing:

All data were provided by the National Breeder Association of Reggiana cattle (ANABoRaRe, Mancasale Reggio Emilia, Italy), following the official milk recording system.

Milk dataset:

Test day dataset (TD) contains information of 301,537 records routinely collected from 1991 to 2021, belonged to 13,467 animals and lactations. Milk dataset contains information about milk yields (MILK_Y, kg/d), percentage of fat (FAT_p, , kg/d) and protein (PRT_p, , kg/d), and somatic cell counts (no./mL). Only TD belonged to 1 to 5 parities has been used. Record with a day in milk (DIM) outside the interval of 5 d and 305 d were removed. The cows with an age at calving outside the interval of 21-44 months for first parity, 23-60 for the second, 44-76 for the third and 56-87 and 59 -110 for the fourth and fifth parity respectability, has been also deleted form dataset.

Additionally, only lactations with at least one TD starting before 45 days and at least four TD records has been retrieved for further analysis. The phenotype outside the mean ± four standard deviations within parity and lactation phase (considering 15 d intervals) has been removed from the dataset. Lastly, only records belonging to herd-TD with at least three observations was retained. Then, somatic cell counts (no./mL) were normalized in SCS according to (Ali and Shook, 1980), Protein and Fat yields was also derived as multiplication of MILK_Y per PRT_p and FAT_p respectability. The final dataset contained 115,432 TD records belonging to 16,134 and 6,921 cows. The pedigree file contained 8,792 animals, tracing back up to the 5^{th} generations.

Fertility:

Fertility traits were analyzed combing information from two data sources, the insemination dataset and the test-day dataset. The insemination dataset contained the inseminations events (n =53,201) of 11.936 cows collected from 1986 to 2020. For what concern TD database, same data editing reported above was performed, and in addition lactation that presented two and more herd for the same animals was removed, interval of two consecutive parity less than 8 months and over 10 months. After these checks, the test-day and the insemination datasets were merged and cleaned according to (Mancin et al., 2020). The merged dataset contained of 22,650 lactations information of 8,007 cows. Four fertility traits have been considered in this study: days open (DO), calving interval (CI), calving to first insemination (PFI) and number of inseminations to achieve the pregnancy (N_INS). The DO represented the interval between date of parity and the insemination in which pregnancy was achieved, CI was calculated as difference between two consecutive dates of parity, CFI is difference between date of parity and first insemination after the parity, NINS is the count of insemination necessary to achieve the pregnancy, it is considered as categorical trait (Tiezzi et al., 2012a). Category was divided by number of inseminations expect for insemination after 5 are considered as unique category (Tiezzi et al., 2012a). DO and N_INS records belong to the last lactations, i.e., lactations without a subsequent one of a live animals was considered as censored information. Since censored records represented nearly 2% of phenotypes, it was removedfor a matter of simplicity.

Note that the phenotype has different consistency, CI has the least amount of data 13,826 since two consecutive parity dates are needed, DO and CFI contain 17,350 phenotypes, while N_INS containscontain more phenotypes than all other traits 22,535 since it was also possible to calculate it on heifers.

Single traits models:

Univariate animal model was applied to estimate the variance components and heritability for theten phenotypes.

Milk traits:

Random regression test day models have been used for estimated variance components for MY,PRT_y, FAT_y PRT_p FAT_p and SCS

$$y_{ijklmno} = HTD_i + LN_j + GL_K + \sum_{r=1}^{3} 3\varphi r \times AP - LN_l + \sum_{r=1}^{3} 3\omega r \times MP - LN + Pe_n + a_n + e_{ijklmno} \ (1)$$

where y is the individual test-day record of the n^{th} cow; HTD_i is the cross-classified fixed effect of herd-test-day (17,628 levels); LNj represents the cross classified fixed effect of lactation number (3 levels, corresponding to the first three lactations); GLk is the cross classified fixed effect of kth class (18 classes with 1 meaning no gestation and further classes accounting for 15-d intervals, from 1 to 240 d of days after conception),AP-LN is the cross classified fixed effect of l^{th} age at parity within lactation (42 classes in total); MP-LN is the fixed effect of the m^{th} month of parity (36 classes corresponding to single months of a year within each j lactation).

While random effect is represented by permanent environmental component (Pe), and the additive genetic effect (a_n) both sampled from a normal distribution with different co-variances structure, for further information see Model computation paragraphs. Residuals ($e_{ijklmno}$) are also sample form a homogenous normal distribution.

To describe form of lactation curve Fourth-order Legendre polynomials are covariates on the effectof AP-LN and MP-LN, φ and ψ on the formula 1 are coefficients for the polynomial of order r varying between 0 and 3.

Fertility traits:

For what concerned fertility traits the following animal's model has been used:

$$y_{ijkn} = H_i + YS_j + LN_k + a_n + e_{ijkn} \ (2)$$

where y_{ijkl} represent one of the fourth fertility traits, H_i is the cross-classified fixed effect indicating herd effect, levels change according with the traits considered (form 153 in CI to 175 for NINS). YS_j is the years-season cross classified fixed effects extract form the date of parity (86 to 92 levels). The random effect additive and residual are sampled form the same distributions described above.

Environmental gradient

Environmental gradient was calculated form the solution of random cross-classified effect HYM (herd and the years and months of the control) for the MY (kg/d). Tests day random regression model are used:

$$y_{ijklmno} = HYM_j + LN_J + GL_K + \sum_{r=1}^{3} 3\,\phi\,r \times AP - LNl + \sum_{r=1}^{3} 3\omega r \times MP - LN + Pen + a_n + e_{ijklmno} \quad (3)$$

The model is like (1) expect that HTD_i has been replace by HYM_j. Additional to avoid the effect was sampled form a normal distribution with $N(0, I\sigma^2)$. To avoid bias and inaccurate estimationof the environmental gradients data HYM with at least four record was retrieved. Then HYM with higher standard error or with accuracy over than 50%). was maintained.

Bit-traits models*:*

Genetic correlations between all the phenotypes are calculated using bi-variate models where:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} pe_1 \\ pe_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (4)$$

to the phenotypic records or traits the liabilities of the categorical traits (N_INS); X_1 and X_2 are the incidence matrices for fixed, while W_1,W_2 incidence matrices of the random effects permanent environment and Z_1, and Z_2 are the ones of the additive genetic. Thevectors of the systematic effects are represented by b_1, b_2, whereas pe_1a pe_2 are those of the permanent environment effects. The vector of additive genetic effect is represented by a_1 and a_2, residual is represented by e_1 and e_2.

Reaction norm*:*

RNN consist of an implementation of single traits animal models, in which genetic sensitivity has been taken in account by a random regression of genetic additive effect on environmental gradientestimated on (3). In this case beyond the random genetic effect (a_0) it is reported (a_1) that is the quote represented by the GxE. In matrix from RN is reported as the following:

$$y = Xb + Wpe + Z_0a_0 + Z_1a_1 + e \quad (5)$$

Where Z_0 and Z_1 are matrix of the two additive effects, Z_0 is a matrix that connect a_0 to the phenotype, while $\boldsymbol{Z}_1$ is a matrix containing environmental gradient obtained in (3) and used Where Z_0 and Z_1 are matrix of the two additive effects, Z_0 is a matrix that connect a_0 to the phenotype, while $\boldsymbol{Z}_1$ is a matrix containing environmental gradient obtained in (3) and used as covariables

And it is assumed a distributions

$$\begin{bmatrix} a_{n0} \\ a_{n1} \end{bmatrix} \sim N\left(0, A \otimes \begin{bmatrix} {\sigma_{a_0}}^2 & \sigma_{a_0,a1} \\ \sigma_{a_0,a1} & {\sigma_{a_1}}^2 \end{bmatrix}\right) \quad (6)$$

Where $\boldsymbol{Z}_0$ is an incidence matrix presented as in (1-2), while According to (Zhang et al., 2019); residuals residual variances are usually fitted when residual variances differ among production environments. In this model residual vector are heterogenous distribution in followed a normal distribution, $N(0, \begin{bmatrix} I \otimes \sigma_{e1} & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & I \otimes \sigma_{en} \end{bmatrix})$ where a diagonal matrix with elements for observations in the i[th] quantile groups of the standard deviations. In the present study only 3 quantiles have been used, due to limited data size.

Model's computations and assumption:

To estimate the (co)variance components, a Gibbs sampling algorithm was used implemented in the gibbs3f90 for continuous traits and in thrgibbs3f90 for threshold traits (Aguilar et al., 2018). A total of 500,000 Gibbs samples chain was generated, with an initial burn-in of 100,000 and to avoid collinearity one of every 100 chains was retained, this was done using postgibbsf90, median and highest posterior density interval (HDP5 and HPD95) of that chain was reported in the results. Continuous traits was genereted form a normal probability function. Categorical traits were sample form a truncated normal distribution bounded by T delimiter based on values of observed variable (*y*). For example, assuming that the random *y* is composed by *n* levels, thus *n+1* delimiter: $T = \{t_0, t_1, \ldots, t_n, t_{n+1}\}$. Then assuming a liability scales of :

$$l_i = Xb + e$$

Were Xb are generic effect used in the models, thr conditional probability of *y* is under one of category (*l*) is:

$$\mathrm{P(yi\ =\ j|\ \beta, T)\ = P(tj-1 < l \leq t\ |\ \beta, T\ ) = \Phi[Tj\ -\ X\beta]\ -\ \Phi[Tj-1\ -\ X\beta\ ]},$$

Φ(.) is the standard cumulative normal distribution function, with Xb we generally mean all effect presented in the models. In the Gibbs sampling Bounded uniform prior were used for all fixed effects, and null means and normal distributed priors were used for permanent environment, additive genetic, and residual effect, with this matrix notations:

$$a \sim N(0, G \otimes A);\ pe \sim N(0, Pe \otimes I); e \sim N(0, R \otimes I)(7)$$

where **A** represents the relationship matrix obtained from pedigree, and **I** is an identity matrix. In single traits **G** **P**e and **R** are represented by a scalar, $(\sigma_a^2,\ \sigma_{pe,}^2 \sigma_e^2)$, representing additive genetic variances, permanent environment variances and residuals one's respectability, while in the bi-traits analysis were as follows:

$$G = \begin{vmatrix} \sigma_{a1}^2 & \sigma_{a1a2} \\ \sigma_{a1a2} & \sigma_{a2}^2 \end{vmatrix}; Pe = \begin{vmatrix} \sigma_{pe1}^2 & \sigma_{pe1pe2} \\ \sigma_{pe1pe2} & \sigma_{pe2}^2 \end{vmatrix};\ R \begin{vmatrix} \sigma_{e1}^2 & \sigma_{e1e2} \\ \sigma_{e1e2} & \sigma_{e2}^2 \end{vmatrix}(8)$$

where **G** is the matrix of additive genetic (co)variances σ_{a1}^2, σ_{a1a2}, σ_{a1a2} of traits 1 and 2. **Pe** is the matrix of permanent environmental (co)variances σ_{pe1}^2, $\sigma_{\mathrm{pe1,pe2}}$, σ_{pe2}^2, and R the matrix of residual (co)variances σ_{e1}^2, σ_{e1e2} and $\sigma_{\mathrm{e2}}^2$of traits 1 and 2. Note that when different datasets were merged (i.e. milk and fertility traits), residual (co)variance was set to zero because the traits were recorded in different moments. Estimated heritability calculated in the single traits analysis was calculated as $h^2 = \frac{\sigma_a^2}{\sigma_p^2}$, where σ_p^2 is the total phenotypic variance express as $\sigma_p^2 = \sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2$, in the reaction norm due to the heterogenous variances of the residuals σ_e^2 was calculated as the mean of five quantile groups variances. The estimated of correlation (genetic and residual) were calculated as $r_a = \frac{cov(x,y)}{\sigma_{ix}\sigma_{iy}}$, where i refers to the genetic, and phenotypic; x and y refer the different phenotyped while; cov stands for the estimated covariance between the traits; and σi x, and σi y, are the standard deviation of traits

Validation of Random Regression models:

To evaluate the effectiveness of inclusion of reaction norm Likelihood Ration Test (LRT) was used as in (Zhang et al., 2019). RNN models was compared with reduced models that is equal to RNN without considering random regression. Prediction accuracy of EBV of RNN was also compared with reduced models, using LR cross validations. LR cross validation was used for young bulls born after 2008 (70 animals). In these phenotypes of bulls's daughter has been removed (20% of data) (Legarra and Reverter, 2018), using similar statistic of (Mancin et al., 2021b).

RESULTS AND DISCUSSION:

Descriptive statistic:

Mean, coefficient of variation (CV), minimum and maximum values for the ten phenotypes were reported in Table 1 and their distributions were reported on S1. In literature few studies reported the performance of productive and reproductive traits of Reggiana breed. (Gandini et al., 2007) is one of the few studies. Moreover, since this study was conducted 16 ago, it may be interesting to make a comparison on the state of selection and not of Reggiana, in terms of phenotypic progress. No differences were observed for fertility traits, while a slight increase was found for protein and fat percentage: PRT_P increase from 3.38 % to 3.70%, while FAT_P increase from 3.21% to 3.45%.

An increase of milk production of about 450 kg of milk per lactation was also observed. The increase in milk production together with the increase in lipid and protein content could be attributed to an increase in production technologies, i.e. feeding, housing (Khanal et al., 2010), but also from an interest of farmers in selecting more productive animals net of a higher protein content of milk due to the incentive of "*Parmiggiano Reggiana delle Razze Rosse*".

Reggiana has daily milk production levels that are on average 19.3 kg/d, that is lesser than those reported by the Italian Breeders Association (AIA (Italian Breeders Association), 2016) in specialized Italian Holstein (31.3 kg/d), Italian Brown Swiss breeds (23.6 kg/d), and dual-purpose Italian Simmental cows (22.0 kg/d). However, compared with other local breeds Reggiana has greater milk production as Grey Alpine (16.30 kg/d) (Mancin et al., 2021a) or Rendena (16.5 kg/d) (Guzzo et al., 2019). On the other hand, Reggiana presented a high SCS value compared to the other local breeds mentioned, (3.22 points) which places it much closer to other cosmopolitan breeds such as Holsteins and Brown Swiss (Franzoi et al., 2020).

Table 1 proved the good Reggiana fertility parameters presented by Reggiana breed. Indeed, Reggiana has a DO of only 180 days and only 1.30 interventions for conception. Comparing them with studies conducted on other breeds (Toledo-Alvarado et al., 2017; Martinez-Castillero et al., 2020), the Reggiana presented significantly lower DO, CI and CFI respect to the other breeds as Holstein, Brown Swiss Italian Simental, it presented also better parameter than other local breeds as Gray Alpine. However, this discrepancy with respect to the Alpine Gray is mainly due to the non- seasonality of pregnancy of Reggiana which makes it have shorter intervals on average. Both traits have shown a slight progress over time, this more than from an increase in the genetic value is due to an increase in environmental conditions, such as the breeding effect, especially for the fertility traits (S1).

Table 1. Descriptive statistic of the ten phenotypes after data-editing

Type traits	traits	units	Mean	Min.[1]	Max.[2]	C.V. (%)[3]	N[4]
MILK TRAITS	**MILK_y**	Kg/day	19.110	0.200	90.000	0.376	115432
	FAT_y	Kg/day	0.660	0.008	5.359	0.429	115432
	PRT_y	Kg/day	0.637	0.006	3.456	0.341	115432
	FAT_p	%	3.701	0.053	16.930	0.233	115432
	PRT_p	%	3.450	0.160	10.650	0.112	115432
	SCS	Count	3.227	-3.644	10.893	0.569	115397
FERTILITY TRAITS	**DO**	Days	108.930	3	299	0.625	17465
	CI	Days	391.410	279	594	0.140	13897
	CFI	Days	80.580	3	199	0.525	17465
	N_INS*	number	1.330	1	5	0.583	22650

[1]Min: minimum values; [2]Max: maximum value;[3]Coefficient of Variation [4]N:number of phenotype; milk yields;*categorical traits. Milk yields (MILK_y), percentage of fat (FAT_p) and protein (PRT_p), Fat yields (FAT_y); Protein yields (PRT_y); somatic cells score (SCS); Days open (DO); Calving Interval (CI) ; Calving First Insemination (CFI); number of insemination (N_INS)

Variance components:

Single traits

Variance's components of the then phenotype is reported on table 2. Milk traits presented medium hereditability ranging from 27% from PRT_P to 7% for FAT_Y, while reproductive traits presented a low hereditability close to 2%.

Observing milk traits, Reggiana presented smaller h^2 values compared with other Italian cattle breeds. For example Italian Simmental presented a heritability value of 0.18, 0.13, and 0.17 for milk, fat, and protein yields, respectively (Frigo et al., 2013), Rendena breed has an heritability of 0.188, 0.157, and 0.165 (Sartori et al., 2018), in Valdostana they were 0.198, 0.132, and 0.169 (Sartori et al., 2020), while hereditability are presented in Grey Alpine with 0.219 for MILK_Y and 0.125 0.178 for FAT_Y and PRT_Y (Mancin et al., 2021a). The lower values found Reggiana can be attributed by two factors. One factor is relating to the short and incomplete pedigree compared with the others breeds, only four generation on pedigree were traced back. The second reason may be the lower genetic variance expressed by Reggiana

due to the bottleneck present in the 1980s which could have reduced the genetic variances (www.ANABoRaRE.it). However, these results are consistent with the ones present in the extensive literature conducted on milk (co)variances components.(Miglior et al., 2017). In our study and in those previously cited, the FAT_Y showed lower heritability than protein ones. Although the similar amount of additive genetic variance was present in both traits, the fat yield had nearly double the residual varices. The greater residual variance may be the cause of lower fat yields heritability. In fact some studies reported that lipid components on milk are more influenced by feeding regimen respect to protein (Van Soest, 1963; Gurr, 1985). The heritability of SCS in Reggiana, was similar to the ones reported in other countrywide spread dairy and/or dual-purpose breeds such as Holstein, Ayrshire, and Italian Brown Swiss (Reents, 1995; Ikonen et al., 2004; Dal Zotto et al., 2007).Similar values were also present in the other local breeds as Valdostana, Rendena, Grey Alpine (Sartori et al., 2018; Mancin et al., 2021a).

An h^2 of almost 0.02 have been identify for all fertility traits, ranging to 1.8% for N_INS to 2.3% for CFI. However, it is difficult to make a comparison with other local breeds, since i) few studies have investigated the hereditability of fertility traits in local breeds ii) probably there will not be many differences as extremely fertility is an extreme conserved trait. Indeed, results found in our studies were similar to those from previous studies conducted on specialized breeds ((González-Recio and Alenda, 2005; Tiezzi et al., 2012b; Liu et al., 2017)). In our study as in (Zhang et al., 2019)), N_INS presented the lower h^2 values among all fertility traits, while CFI has greater hereditability.

Table 2. Variance's components estimated under single traits models

Traits	Va	Vpe	Vres	h^2
MILK_y	2.995	8.205	13.692	0.120
	(2.170 3.785)	(7.520 8.822)	(13.590 13.830)	(0.089 0.151)
FAT_y	0.376	1.080	3.950*	0.069
	(0.2721 0.4823)	(0.986 1.170)	(3.918 3.986)	(0.051 0.089)
PRT_y	0.2571	0.837	1.5673*	0.097
	(0.1863 0.3300)	(0.7748 0.9016)	(1.5540 1.581)	(0.071 0.123)
FAT_p	0.0902*	0.044	0.4664	0.150
	(0.0776 0.1033)	(0.0366 0.0532)	(0.4624 0.4704)	(0.130 0.170)
PRT_p	0.024	0.014	0.051	0.273
	(0.0213 0.0277)	(0.0117 0.0158)	(0.0506 0.0515)	(0.242 0.305)
SCS	0.213	0.774	1.787	0.077
	(0.1506 0.2837)	(0.7147 0.8320)	(1.7680 1.7990)	(0.054 0.100)
DO	57.355	173.990	2742.827	0.019
	(22.7100 97.4700)	(117.5000 236.30)	(2668 2818)	(0.008 0.033)
CI	55.176	169.380	2570	0.0197
	(14.790 99.780)	(102. 231.3)	(2489 2650)	(0.06 0.0356)
CFI	29.383	57.395	1037.2	0.0261
	(11.880 46.610)	(35.400 82.810)	(1010 1067)	(0.011 0.042)
N_INS	0.020	0.047	1.033**	0.0193**
	(0.010 0.036)	(0.017 0.066)	(0.987 1.233)	(0.010 0.0233)

*express as liability . Milk yields (MILK_y), percentage of fat (FAT_p) and protein (PRT_p), Fat yields (FAT_y); Protein yields (PRT_y); somatic cells score (SCS); Days open (DO); Calving Interval (CI) ; Calving First Insemination (CFI); number of insemination (N_INS),Additive genetic variance (Va), Permanent enviroment variance (Vpe); residual variances (Vres);(h^2) hereditabilty

Bit-traits and genetic correlations

The estimated genetic and phenotypic correlation are reported on Table 3. As expected, production tare highly correlated. PRT_Y has a correlation of 0.88, with MILK_Y, and a correlation of 0.638 with FAT_Y, while in PRT_Y and FAT_Y correlation was 0.75. The extensive literature conducted on both local and specialized breeds, confirm the high correlations among traits (Dube et al., 2009; Mazza et al., 2016; Charton et al., 2018; Sabedot et al., 2018). In fact, FAT_y and PRT_y is directly dependent to the daily milk production. However, MILK_y is negatively correlated with FAT_y and PRT_Y, with a value of -0.378 - 0.531, respectability. It means that a selection focus only on milk production will reduce the solid content of milk. Indeed, in a genomic perspective the antagonism of Milk yields and milk solids has been demonstrated by the opposite effect on SNPs in linkage disequilibrium with DGAT1 (Jiang et al., 2019). In addition some studies identify the GABARAPL1 gene presented an antagonistic effect on milk yield and fat percentage (Pimentel et al., 2011; Nayeri et al., 2016). This negative correlation must be taken in account in selection plans on Reggiana were besides to selection for milk traits and selection for increase of percentage of protein and lipid must be considered. On the contrary, selecting the animals only for milk

productivity may result in a decline in the genetic parameters for the various percentages in the long terms (de Jager and Kennedy, 1987). Furthermore, a reduction in the protein and lipid content of milk leads to a less cheese-making milk (Guinee et al., 2007). This could be a particular problem in the Reggiana breed as the majority milk produced is destined for cheese.

SCS presented significative positive correlation with MILK_Y, which mean that increase of milk productivity lead a detriment of udder condition and a consequently augment SCS concentration (Kheirabadi and Razmkabir, 2016). However, SCS has a negative correlation with fat and protein percentages therefore a selection focus on increase percentage of solid content, as mentioned before, could be also beneficial for the udder health. Furthermore, the slight positive correlation with milk production (0.36), demonstrated the possible suitability of selection plans to focus and increase both milk production and quality (SCS and percentage of solid). All fertility traits are highly genetic correlated among them. DO and CI are extremely genetically correlated (r=0.984), in fact DO and CI only differs for parity length that is almost equal in all animals. These traits are also highly genetic correlated with CFI, about 0.89. N_INS presented overall lower correlation with these traits, it presented average correlation of 0.50 with DO CI, while no significant correlation has been observed with CFI. Same correlation pattern has been observed in (González-Recio and Alenda, 2005; Tiezzi et al., 2012a). Fertility traits have a positive correlation with milk traits, which means that an increasing of milk production is genetically connect with worse performance in terms of fertility. Milk yields have a correlation of about 0.52 with DO and CI, while 0.3 with CI and no significant correlation with N_INS. The other yields (FAT_Y and PRT_Y) traits presented similar correlation. In fact it is widely known that the more productive cows presented negative energetic balance and thus less energy to dedicated to the reproduction (Berry et al., 2016). However, (Strucken et al., 2012) claims that increasing the allele favorable to reproductive seems not affect the selection on milk production, since only tree marker associated with both traits are identify.

Table 3. Genetic correlations (upper diagonal) and phenotypic (lower diagonal), bold number are significant correlation, zero not included in the HPD95

	MILK_y	FAT_y	PRT_Y	FAT_p	PRT_p	SCS	DO	CI	CFI	N_INS*
MILK_y		**0.638**	**0.886**	**-0.378**	**-0.531**	**0.273**	**0.501**	**0.542**	**0.372**	0.415
FAT_y	**0.894**		**0.752**	**0.501**	-0.011	-0.270	**0.468**	**0.381**	**0.348**	**0.582**
PRT_y	**0.901**	**0.729**		-0.093	-0.071	0.050	**0.488**	**0.470**	**0.422**	0.343
FAT_p	-0.067	**0.392**	**0.016**		**0.585**	**-0.601**	0.100	-0.038	0.151	0.096
PRT_p	**-0.426**	**-0.046**	**-0.012**	**0.295**		**-0.496**	-0.117	-0.209	0.094	-0.224
SCS	**-0.027**	**-0.010**	**-0.019**	**0.013**	**0.022**		-0.015	-0.118	**0.518**	-0.748
DO	**0.071**	**0.031**	**0.022**	0.002	**-0.001**	-0.005		**0.984**	**0.885**	**0.524**
CI	**0.074**	**0.032**	**0.022**	0.001	**-0.002**	0.003	**0.994**		**0.892**	**0.654**
CFI	**0.071**	**0.032**	**0.022**	0.003	-0.002	0.016	**0.531**	**0.520**		0.097
N_INS*	**0.728**	**0.766**	**0.598**	0.121	0.013	-0.273	**0.662**	**0.133**	**-0.113**	

*Express as liability. Milk yields (MILK_y), percentage of fat (FAT_p) and protein (PRT_p), Fat yields (FAT_y); Protein yields (PRT_y); somatic cells score (SCS); Days open (DO); Calving Interval (CI) ; Calving First Insemination (CFI); number of insemination (N_INS)

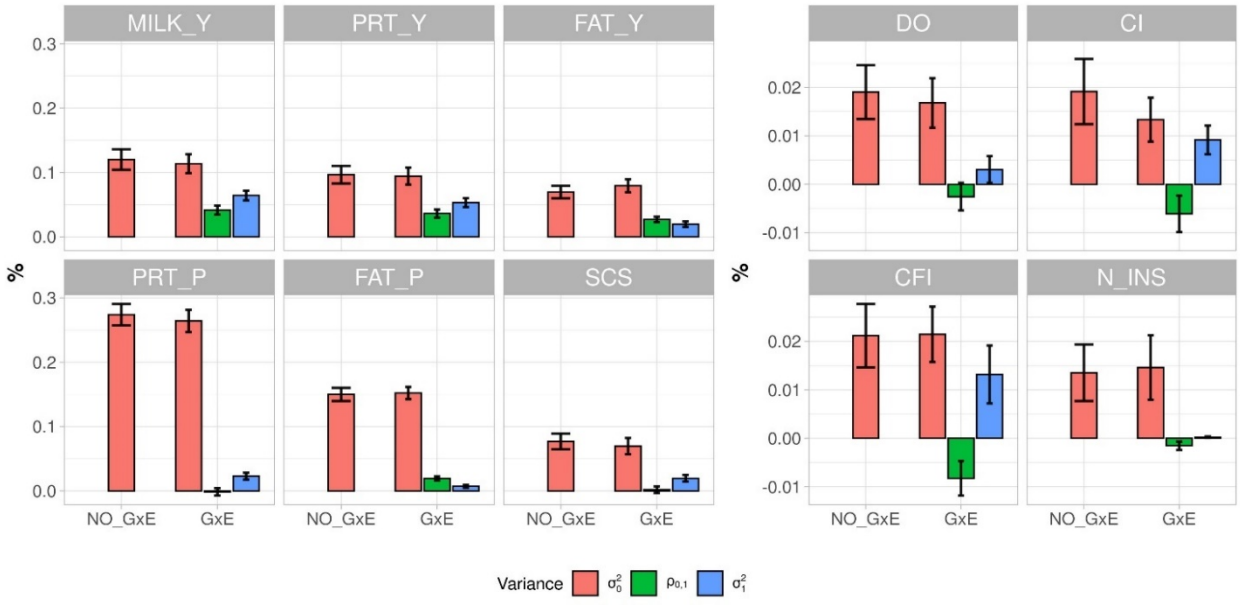
Reaction norm

Variance estimated under RNN are reported on Table 4, Figure 1 described the ration between $\sigma_0^2, \rho_{0,1,}\sigma_1^2$ and phenotypic variances. RNN analysis was possible to conduct on Reggiana despite the reduced populations numerosity. This may be due to the use of few sires and the homogeneous distribution of their dams in the different herd ranked by production. However, RNN analysis performed on fertility traits presented estimation with higher standard deviation while in milk traits not. This discrepancy are explained by the different number of phenotype in the two dataset (Misztal and Legarra, 2017). The Results identify in the current study agree to what reported on literature, namely that production traits expressed the higher quote of GxE respect the ones connected with longevity or fertility (Tait-Burkard et al., 2018). Milk traits presented higher quote of variance explained by GxE (σ_1^2), ranging form 5% (MILK and PRT_Y) to almost zero for PRT_p, while fertility presented lower values around 0.05%. In particular, yields traits presented higher GxE (slope), for MILK_y and PRT_Y. Small GxE quote are found for FAT_p, SCS ; while for PRT_p the quote was not significant. Study in which herd production was used as enviromental gradient presented simlar pattern Sartori 2022 and Smith et al. 2021. Additionally, similar results for milk production were identified in a study that used THI (Temperature Humidity Index) as an environmental covariate, but not always for fat and protein yield.(Negri et al., 2021, Mulim et

al, 2021). Excecpt for SCS and PRT_p, the milk traits had a significant positive correlation between slope and intercept. Ideed ,MILK_y and PRT_y presented similar correlation of about 0.5, while higher correlation are identify on FAT_p (r=0.58) and FAT_y (r=0.70). The higher correlations of FAT traits respect to the other traits was also reported in (Fikse et al., 2003; Shariati et al., 2007).

Fertility traits have only a low quote of genetic variances expressed by GxE. DO was the only fertilty traits with no significant quote of GxE. However, we believe this non-significance is due to poor convergence of gibbs sampling due to the small number of samples rather than a biological reason. Negative correlation between slope and intercept was identify in all fertility traits, as in (Zhang et al., 2019; Shi et al., 2021). DO,CI and CFI have a average correlation of -0.5 while N_INS presented higher correlations values (-0.89), similar trends are reported in (Zhang et al., 2019), where year production has been also used as environmental gradient. Despite this, the latter study presented a higher overall negative correlation, which could potentially mean that Reggiana exhibited greater phenotypic plasticity for fertility than Danish Holstein cattle. However, many factors may have influenced this discrepancy, starting with the sample size and different number of class used for heterogenous variances.

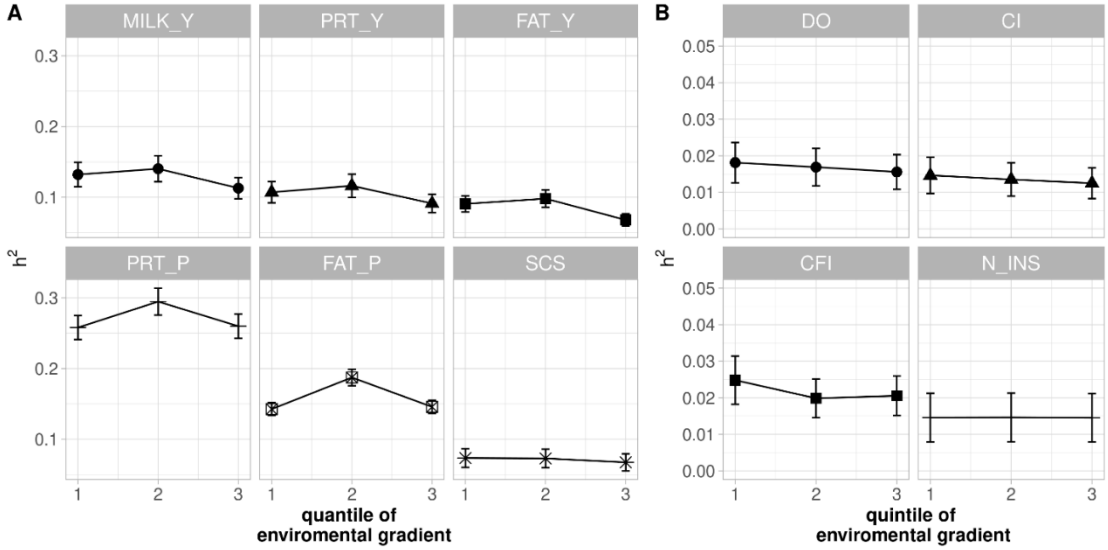
Figure 1. Bar plot described the ration between $\sigma_0^2, \rho_{0,1}, \sigma_1^2$ and phenotypic variance. Reduced models (NO_GxE) and Reaction Norm (GxE) models were compared, standard deviation of estimation has also been reported as black bar.



Milk yields (MILK_y), percentage of fat (FAT_p) and protein (PRT_p), Fat yields (FAT_y); Protein yields (PRT_y); somatic cells score (SCS); Days open (DO); Calving Interval (CI) ; Calving First Insemination (CFI); number of insemination (N_INS)

The positive correlation between interception for milk traits means that animals allocated in more productive environments can better express their genetic potential for milk, protein, and fat production, while it appears not to affect the percentage of protein and fat in milk. and SCS.Same trends have been observed for fertility traits where more productive environment presented shorter interval and lesser number of insemination events. This apparent contradiction has been explain in (Toledo-Alvarado et al., 2017), which demonstrated that a more productive herd means a more favorable environment ,with greater technological input, where animals are able to best express the genetic potential of animals. Compared to the other studies (Mwansa and Peterson, 1998; Wallenbeck et al., 2009; Rauw and Gomez-Raya, 2015; Zhang et al., 2019; Schmid et al., 2021) only 3 classes of heterogeneous residues were used here , this is due to a smaller number of heads and herd (environmental gradient)This caused in a more discontinuous estimates of heritability between herd groups Figure 2. As in (Zhang et al., 2019) residual variance components did not increase along with the quantile of herd effects production. The highest hereditability values are identified in the middle quantile for milk traits traits, while no clear patterns have been observed for fertility. On this point Calus et al. (2006) suggested that a higher-order RNM and alternative heteroskedastic error specifications might be used in analysis of G × E interactions in other studies (Cardoso and Tempelman 2012), however it was out of the aim of the current work.

Figure 2. Hereditability of the intercept of reaction norm, for milk traits (A) and fertility traits (B), calculated in the different quantiles. Bar plot are the standard error of estimation



Milk yields (MILK_y), percentage of fat (FAT_p) and protein (PRT_p), Fat yields (FAT_y); Protein yields (PRT_y); somatic cells score (SCS); Days open (DO); Calving Interval (CI) ; Calving First Insemination (CFI); number of insemination (N_INS)

Table 4 Variances components estimated under RNN models, number between bracket are the HPD-5 and HPD-95 interval, bold number in the corr column indicated a significant correlation

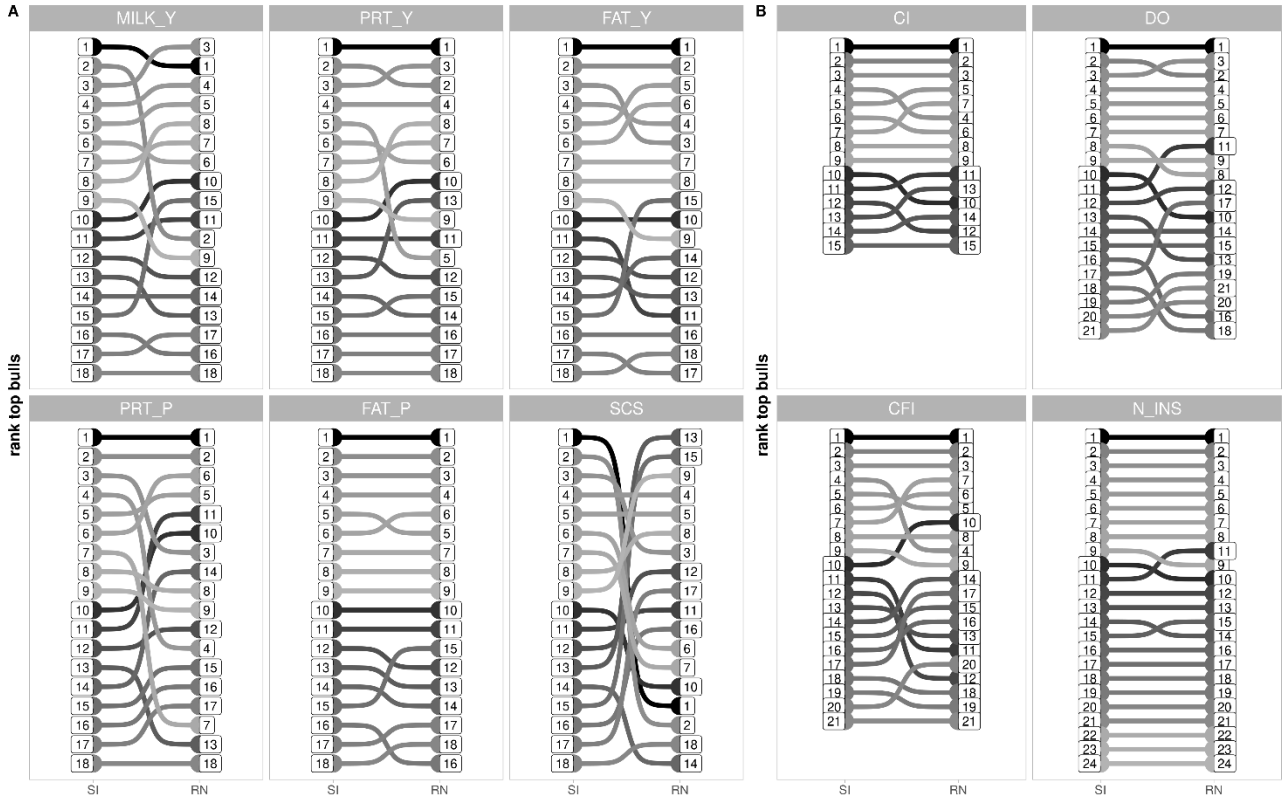
TRAITS	ga0	covga01	ga1	pe	res*	corr
MILK_y	3.035 (2.398 3.628)	1.115 (0.805 1.386)	1.724 (1.383 2.003)	7.3370 (6.781 7.720)	13.626 (13.520 13.707)	**0.4931 (0.3590 0.5837)**
FAT_y	0.444 (0.357 0.524)	0.1522 (0.110 0.185)	0.110 (0.073 0.144)	0.9674 (0.895 1.020)	3.951 (3.920 3.975)	**0.701 (0.527 0.795)**
PRT_y	0.2661 (0.207 0.320)	0.1034 (0.072 0.128)	0.1511 (0.118 0.177)	0.7579 (0.704 0.796)	1.558 (1.54 1.567)	**0.5193 (0.372 0.614)**
FAT_P	0.0926 (0.082 0.100)	0.0118 (0.0084 0.0145)	0.004 (0.002 0.006)	0.0408 (0.0350 0.0459)	0.4592 (0.4557 0.4618)	**0.6069 (0.4158 0.7569)**
PRT_P	0.023 (0.0205 0.0256)	0.000 (-0.0010 0.0005)	0.002 (0.001 0.003)	0.0129 (0.0113 0.0143)	0.0504 (0.0501 0.0507)	-0.0185 (-0.1359 0.0740)
SCS	0.1941 (0.1394 0.2401)	0.005 (-0.0173 0.0237)	0.053 (0.0323 0.073)	0.756 (0.700 0.793)	1.786 (1.773 1.796)	0.0530 (-0.1775 0.2381)
DO	49.60 (28.513 73.506)	-7.865 (-20.155 2.643)	6.454 (1.160 17.297)	186.45 (1396. 225.5)	2806. (2746 2856)	-0.577 (-0.9107 0.1193)
CI	36.055 (19.179 55.20)	-8.557 (-17.101 -1.604)	51.101 (27.601 74.732)	147.250 (99.986 185.210)	2613 (2549 2662)	**-0.4183 (-0.732 -0.0812)**
CFI	24.425 (15.375 33.937)	-9.5465 (-16.345 -4.352)	15.2900 (2.812 24.143)	52.8550 (32.145 67.460)	1078 (1055. 1099)	**-0.5486 (-0.8239 -0.2578)**
N_INS*	0.0153 (0.0068 0.0263)	-0.0016 (-0.0035 -0.0006)	0.0002 (0.0001 0.0005)	0.0568 (0.0383 0.0730)	1.0443 (0.9626 1.1527)	**-0.8921 (-0.9551 -0.6956)**

*N_INS expressed a liability scale; Milk yields (MILK_y), percentage of fat (FAT_p) and protein (PRT_p), Fat yields (FAT_y); Protein yields (PRT_y); somatic cells score (SCS); Days open (DO); Calving Interval (CI) ; Calving First Insemination (CFI); number of insemination (N_INS). Additive variance intercept (ga0); covariance between intercept and slope (covga01) slope (ga1), permanent environment (Pe); residual (res); correlation between ga0 and ga1 (cor)

Reaction norm vs Single traits: bulls-ranking and improvement of accuracy:

The spearman correlation of bulls' breeding values (re-ranking) obtained with and without GxE interaction common criterion to establish the necessity to use RNN in common evaluation practice. In our work, only bulls with an accuracy over 0.5 was considered (n=84). Non substantial re-ranking has been observed for milk traits, the highest reranking values was observed for MILK_y and PRT_y and PRT_p with values of 0.95, while FAT_y/p presented a spearman correlation of 0.97 and 0.99 respectability. SCS has an intermediate value of 0.96. No substantial values were also observed for fertility traits with values ranging from 0.97-0.99. Bulls re-ranking were plotted on figure 3, in those cases only young bulls (born after 2010) with at least two dams with two phenotypes was considered, the number of bulls differs according to the different dataset. Other studies investigated the impact of RNN models in bulls re-ranking, despite the variety number of environmental descriptor used and the majority did found any significant bulls-ranking for productive and reproductive traits, for example (Kearney et al.,2004; Craig et al., 2018)). Nevertheless a slight bulls re-rank for fertility traits was observed on (Ismael et al., 2016), however bit-traits models has been used.

Figure 3 Bulls re-ranking using reduced models (NO_GxE) and using reaction norm models (GxE),for milk (A) and fertility(B) traits.

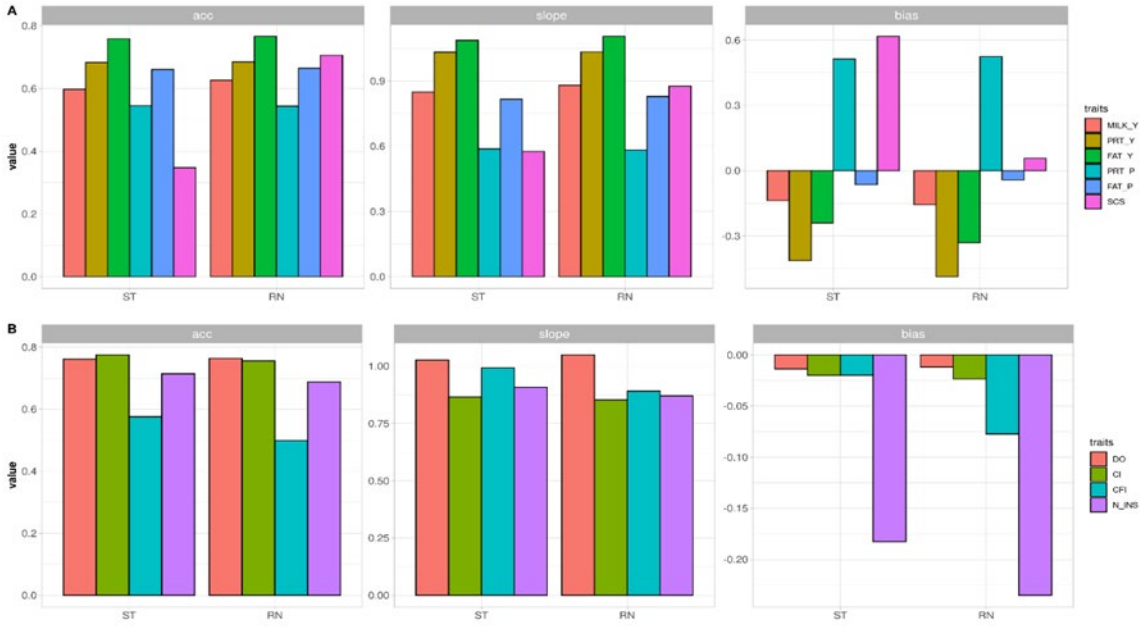


Milk yields (MILK_y), percentage of fat (FAT_p) and protein (PRT_p), Fat yields (FAT_y); Protein yields (PRT_y); somatic cells score (SCS); Days open (DO); Calving Interval (CI) ; Calving First Insemination (CFI); number of insemination (N_INS).

All RNN regression models presented a significative reduction DIC values compared with reduced models, with means that RNN presented better models fitting. Despite this, an approach to verify if this increase in model fitting presented a better EBV estimation was needed. On this point LR cross validation models were used. LR is a straightforward cross validation strategy and it provided bias dispersion and accuracy of the different prediction. The performance of genetic evaluations is inferred on a target group of animals, as in this case young bulls born after 2008 (n = 70). Slight increase of accuracy has been observed for yields traits (MILK_Y, PRT_Y and FAT_Y), a noticeable augment of accuracy has been observed for SCS. Models that do not account for GxE presented a slight overdispersion of breeding values prediction, especially from SCS, RNN on the other had seem to slight improve this dispersion close to an optimal value of one, especially for SCS, on the contrary reaction norm models

presented generally higher bias, farther to the optimum values of one. For what concerned fertility traits no substantial increment has been identify, it is interesting to note that RNN lead a deterioration of overall performance in CFI. However not clear reason responsible of this trend has been identify it may connect to the non-homogeneous number of daughters per environment of the bulls belonged to the validation cohort.

Figure 4. LR cross validation using reduced models (NO_GxE) and using reaction norm models (GxE),for milk (A) and fertility(B) traits.



Milk yields (MILK_y), percentage of fat (FAT_p) and protein (PRT_p), Fat yields (FAT_y); Protein yields (PRT_y); somatic cells score (SCS); Days open (DO); Calving Interval (CI) ; Calving First Insemination (CFI); number of insemination (N_INS).

CONCLUSION:

The current study constitutes a preliminary point for the adoption of adequate selection plans for the Reggiana breed, however it is of great interest for the genetic and phenotypic characterization of this breed since the genetic parameters (heritability and correlation) have never been estimated. The Reggiana had a lower heritability for milk traits, while the genetic correlation and fertility traits are in line with the other breeds. Interestingly, a modest GxE rate was observed for milk but also for fertility traits. However, we identify small impacts of GxE on model's accuracy and bulls re-ranking. The Reggiana like other more productive breeds seems to better express their genetic potential in more productive herds. However, the

Reggiana appears to be less influenced by GxE than other breeds however this could be due to too many factors despite the breed type.

BIBLIOGRAPHY:

Damien Derrouch, Margot Barbier, Julie Labatut, Nathalie Couix. 2016. Native breed: Definition. Dictionnaire d'Agroecologie, https://dicoagroecologie.fr/en/encyclopedia/native-breed/

Mulder, H. A., and Bijma, P. (2007). Effects of genotype x environment interaction on genetic gain in breeding programs. J. Anim. Sci. 83, 49–61. doi: 10.2527/2005.83149x

National Research Council (US) Committee on Technological Options to Improve the Nutritional Attributes of Animal Products. Designing Foods: Animal Product Options in the Marketplace. Washington (DC): National Academies Press (US); 1988. Factors Affecting the Composition of Milk from Dairy Cows. Available from: https://www.ncbi.nlm.nih.gov/books/NBK218193/

Kearney, J.F., Schutz, M.M., Boettcher, P.J., Weigel, K.A., 2004.Genotype X environment in-teraction for grazing versus confinement. I. Production Traits. Journal of Dairy Science87, 501–509.

Negri, R.; Aguilar, I.; Feltes, G.L.; Araujo Cobuci, J. Selection for Test-Day Milk Yield and Thermotolerance in Brazilian Holstein Cattle. Animals 2021, 11, 1–13

Mulim, H.A.; Luiz, P.; Carneiro, S.; Henrique, C.; Malhado, M.; Fernando, L.; Pinto, B.; Mourão, G.B.; Valloto, A.A.; Pedrosa, V.B. Genotype by environment interaction for fat and protein yields via reaction norms in Holstein cattle of southern Brazil. J. Dairy Res. 2021, 88, 16–22.

Aguilar, I., Tsuruta, S., Masuda, Y., Lourenco, D. A. L., Legarra, A., and Misztal, I. (2018). BLUPF90 suite of programs for animal breeding with focus on genomics. *10th World Congr. Genet. Appl. to Livest. Prod.*

Ali, A. K. A., and Shook, G. E. (1980). An Optimum Transformation for Somatic Cell

Concentration in Milk. *J. Dairy Sci.* 63, 487–490. doi:10.3168/jds.S0022-0302(80)82959-6.

Beat, B. Swiss Brown Swiss in different environments : Does GxE play an important role ? 1–3.

Ben-Jemaa, S., Senczuk, G., Ciani, E., Ciampolini, R., Catillo, G., Boussaha, M., et al. (2021). Genome-Wide Analysis Reveals Selection Signatures Involved in Meat Traits and Local Adaptation in Semi-Feral Maremmana Cattle . *Front. Genet.* 12, 669. Available at: https://www.frontiersin.org/article/10.3389/fgene.2021.675569.

Berry, D. P., Friggens, N. C., Lucy, M., and Roche, J. R. (2016). Milk production and fertility in cattle. *Annu. Rev. Anim. Biosci.* 4, 269–290. doi:10.1146/annurev-animal-021815-111406.

Bertolini, F., Schiavo, G., Bovo, S., Sardina, M. T., Mastrangelo, S., Dall'Olio, S., et al. (2020). Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano-Reggiano cheese production system. *Animal* 14, 921–932. doi:10.1017/S1751731119003318.

Biscarini, F., Nicolazzi, E., Alessandra, S., Boettcher, P., and Gandini, G. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. *Front. Genet.* 5, 1–16. doi:10.3389/fgene.2015.00033.

Boudalia, S., Said, S. Ben, Tsiokos, D., Bousbia, A., Gueroui, Y., Mohamed-Brahmi, A., et al. (2020). Bovisol project: Breeding and management practices of indigenous bovine breeds: Solutions towards a sustainable future. *Sustain.* 12, 1–9. doi:10.3390/su12239891.

Charton, C., Guinard-Flament, J., Lefebvre, R., Barbey, S., Gallard, Y., Boichard, D., et al. (2018). Genetic parameters of milk production traits in response to a short once-daily milking period in crossbred Holstein × Normande dairy cows. *J. Dairy Sci.* 101, 2235–2247. doi:10.3168/jds.2017-13173.

Craig, H. J. B., Stachowicz, K., Black, M., Parry, M., Burke, C. R., Meier, S., et al. (2018). Genotype by environment interactions in fertility traits in New Zealand dairy cows. *J. Dairy Sci.* 101, 10991–11003. doi:10.3168/jds.2017-14195.

Dal Zotto, R., De Marchi, M., Dalvit, C., Cassandro, M., Gallo, L., Carnier, P., et al. (2007).

Heritabilities and genetic correlations of body condition score and calving interval with yield, somatic cell score, and linear type traits in Brown Swiss cattle. *J. Dairy Sci.* 90, 5737–5743.

de Jager, D., and Kennedy, B. W. (1987). Genetic Parameters of Milk Yield and Composition and Their Relationships with Alternative Breeding Goals. *J. Dairy Sci.* 70, 1258–1266. doi:10.3168/jds.S0022-0302(87)80139-X.

Dube, B., Dzama, K., Banga, C. B., and Norris, D. (2009). An analysis of the genetic relationship between udder health and udder conformation traits in South African Jersey cows. *Animal* 3, 494–500. doi:10.1017/S175173110800390X.

FAO, and Platform for Agrobiodiversity Research (2011). *Biodiversity for Food and Agriculture*.

Fikse, W. F., Rekaya, R., and Weigel, K. A. (2003). Genotype x environment interaction for milk production in guernsey cattle. *J. Dairy Sci.* 86, 1821–1827. doi:10.3168/jds.S0022-0302(03)73768-0.

Franzoi, M., Manuelian, C. L., Penasa, M., and De Marchi, M. (2020). Effects of somatic cell score on milk yield and mid-infrared predicted composition and technological traits of Brown Swiss, Holstein Friesian, and Simmental cattle breeds. *J. Dairy Sci.* 103, 791–804. doi:https://doi.org/10.3168/jds.2019-16916.

Frigo, E., Samorè, A. B., Vicario, D., Bagnato, A., and Pedron, O. (2013). Heritabilities and genetic correlations of body condition score and muscularity with productive traits and their trend functions in Italian Simmental cattle. *Ital. J. Anim. Sci.* 12, 240–246. doi:10.4081/ijas.2013.e40.

Gandini, G., Avon, L., Bohte-Wilhelmus, D., Bay, E., Colinet, F. G., Choroszy, Z., et al. (2010). Motives and values in farming local cattle breeds in Europe: a survey on 15 breeds. *Anim. Genet. Resour. génétiques Anim. genéticos Anim.* 47, 45–58. doi:10.1017/s2078633610000901.

Gandini, G., Maltecca, C., Pizzi, F., Bagnato, A., and Rizzi, R. (2007). Comparing local and commercial breeds on functional traits and profitability: The case of reggiana dairy cattle. *J. Dairy Sci.* 90, 2004–2011. doi:10.3168/jds.2006-204.

González-Recio, O., and Alenda, R. (2005). Genetic parameters for female fertility traits and a fertility index in Spanish dairy cattle. *J. Dairy Sci.* 88, 3282–3289. doi:10.3168/jds.S0022-0302(05)73011-3.

Guinee, T. P., Mulholland, E. O., Kelly, J., and Callaghan, D. J. O. (2007). Effect of Protein-to-Fat Ratio of Milk on the Composition, Manufacturing Efficiency, and Yield of Cheddar Cheese. *J. Dairy Sci.* 90, 110–123. doi:https://doi.org/10.3168/jds.S0022-0302(07)72613-9.

Gurr, M. I. (1985). Factors affecting the composition of cow's milk: Implications for its nutritive value. *Nutr. Bull.* 10, 139–152.

Guzzo, N., Sartori, C., and Mantovani, R. (2019). Analysis of genetic correlations between beef traits in young bulls and primiparous cows belonging to the dual-purpose Rendena breed. *animal* 13, 694–701. doi:10.1017/S1751731118001969.

Hiemstra, S. J., De Haas, Y., Mäki-Tanila, A., and Gandini, G. (2010). *Local cattle breeds in Europe: Development of policies and strategies for self-sustaining breeds*. doi:10.3921/978-90-8686-697-7.

Huang, W., Carbone, M. A., Lyman, R. F., Anholt, R. R. H., and Mackay, T. F. C. (2020). Genotype by environment interaction for gene expression in Drosophila melanogaster. *Nat. Commun.* 11, 5451. doi:10.1038/s41467-020-19131-y.

Ikonen, T., Morri, S., Tyrisevä, A.-M., Ruottinen, O., and Ojala, M. (2004). Genetic and Phenotypic Correlations Between Milk Coagulation Properties, Milk Production Traits, Somatic Cell Count, Casein Content, and pH of Milk. *J. Dairy Sci.* 87, 458–467. doi:https://doi.org/10.3168/jds.S0022-0302(04)73185-9.

Ismael, A., Strandberg, E., Berglund, B., Kargo, M., Fogh, A., and Løvendahl, P. (2016). Genotype by environment interaction for the interval from calving to first insemination with regard to calving month and geographic location in Holstein cows in Denmark and Sweden. *J. Dairy Sci.* 99, 5498–5507. doi:https://doi.org/10.3168/jds.2015-10820.

Jiang, J., Ma, L., Prakapenka, D., VanRaden, P. M., Cole, J. B., and Da, Y. (2019). A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. *Front. Genet.* 10, 412.

doi:10.3389/fgene.2019.00412.

Khanal, A. R., Gillespie, J., and MacDonald, J. (2010). Adoption of technology, management practices, and production systems in US milk production. *J. Dairy Sci.* 93, 6012–6022. doi:https://doi.org/10.3168/jds.2010-3425.

Kheirabadi, K., and Razmkabir, M. (2016). Genetic parameters for daily milk somatic cell score and relationships with yield traits of primiparous Holstein cattle in Iran. *J. Anim. Sci. Technol.* 58, 1–6. doi:10.1186/s40781-016-0121-5.

Legarra, A., and Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method 01 Mathematical Sciences 0104 Statistics. *Genet. Sel. Evol.* 50, 1–18. doi:10.1186/s12711-018-0426-6.

Leroy, G., Hoffmann, I., From, T., Hiemstra, S. J., and Gandini, G. (2018). Perception of livestock ecosystem services in grazing areas. *Animal* 12, 2627–2638. doi:10.1017/S1751731118001027.

Liu, A., Lund Mogens Sandøand Wang, Y., Guo, G., Dong, G., Madsen, P., and Su, G. (2017). Variance components and correlations of female fertility traits in Chinese Holstein population. *J. Anim. Sci. Biotechnol.* 8, 1–9. doi:10.1186/s40104-017-0189-x.

Mancin, E., Sartori, C., Guzzo, N., and Mantovani, R. (2020). Non-genetic effects affecting fertility traits in local Reggiana cattle. *Acta Fytotech. Zootech.* 23, 341–349. doi:10.15414/afz.2020.23.mi-fpap.338-346.

Mancin, E., Sartori, C., Guzzo, N., Tuliozi, B., and Mantovani, R. (2021a). Selection Response Due to Different Combination of Antagonistic Milk, Beef, and Morphological Traits in the Alpine Grey Cattle Breed. *Animals* 11. doi:10.3390/ani11051340.

Mancin, E., Tuliozi, B., Pegolo, S., Sartori, C., and Mantovani, R. (2022). Genome Wide Association Study of Beef Traits in Local Alpine Breed Reveals the Diversity of the Pathways Involved and the Role of Time Stratification. 12, 1–22. doi:10.3389/fgene.2021.746665.

Mancin, E., Tuliozi, B., Sartori, C., Guzzo, N., and Mantovani, R. (2021b). Genomic prediction in local breeds: The rendena cattle as a case study. *Animals* 11, 1–19. doi:10.3390/ani11061815.

Martinez-Castillero, M., Toledo-Alvarado, H., Pegolo, S., Vazquez, A. I., de los Campos, G., Varona, L., et al. (2020). Genetic parameters for fertility traits assessed in herds divergent in milk energy output in Holstein-Friesian, Brown Swiss, and Simmental cattle. *J. Dairy Sci.* 103, 11545–11558. doi:https://doi.org/10.3168/jds.2020-18934.

Mastrangelo, S., Tolone, M., Sardina, M. T., Sottile, G., Sutera, A. M., Di Gerlando, R., et al. (2017). Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genet. Sel. Evol.* 49, 1–10. doi:10.1186/s12711-017-0360-z.

Mazza, S., Guzzo, N., Sartori, C., and Mantovani, R. (2016). Genetic correlations between type and test-day milk yield in small dual-purpose cattle populations: The Aosta Red Pied breed as a case study. *J. Dairy Sci.* 99, 8127–8136. doi:10.3168/jds.2016-11116.

Miglior, F., Fleming, A., Malchiodi, F., Brito, L. F., Martin, P., and Baes, C. F. (2017). A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. *J. Dairy Sci.* 100, 10251–10271. doi:10.3168/jds.2017-12968.

Misztal, I., and Legarra, A. (2017). Invited review : ef fi cient computation strategies in genomic selection. 731–736. doi:10.1017/S1751731116002366.

Mulder, H. A. (2016). Genomic selection improves response to selection in resilience by exploiting genotype by environment interactions. *Front. Genet.* 7, 1–11. doi:10.3389/fgene.2016.00178.

Mwansa, P. B., and Peterson, R. (1998). Estimates of GxE effects for longevity among daughters of Canadian and New Zealand sires in Canadian and New Zealand dairy herds. *Interbull Bull.*, 27.

Nayeri, S., Sargolzaei, M., Abo-Ismail, M. K., May, N., Miller, S. P., Schenkel, F., et al. (2016). Genome-wide association for milk production and female fertility traits in Canadian dairy

Holstein cattle. *BMC Genet.* 17, 1–11. doi:10.1186/s12863-016-0386-1.

Ovaska, U., and Soini, K. (2017). Local Breeds – Rural Heritage or New Market Opportunities? Colliding Views on the Conservation and Sustainable Use of Landraces. *Sociol. Ruralis* 57, 709–729. doi:10.1111/soru.12140.

Pimentel, E. C. G., Bauersachs, S., Tietze, M., Simianer, H., Tetens, J., Thaller, G., et al. (2011). Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. *Anim. Genet.* 42, 251–262. doi:https://doi.org/10.1111/j.1365-2052.2010.02148.x.

Rauw, W. M., and Gomez-Raya, L. (2015). Genotype by environment interaction and breeding for robustness in livestock. *Front. Genet.* 6, 1–15. doi:10.3389/fgene.2015.00310.

Reents, R. (1995). Zuchtwertschätzung auf Zellzahl. *Züchtungskunde* 67, 461–466.

Sabedot, M. A., Romano, G. de S., Pedrosa, V. B., and Pinto, L. F. B. (2018). Genetic parameters for milk traits, somatic cell, and total bacteria count scores in Brazilian Jersey herds. *Rev. Bras. Zootec.* 47. doi:10.1590/rbz4720160351.

Sartori, C., Guzzo, N., and Mantovani, R. (2020). Genetic correlations of fighting ability with somatic cells and longevity in cattle. *Animal* 14, 13–21. doi:10.1017/S175173111900168X.

Sartori, C., Guzzo, N., Mazza, S., and Mantovani, R. (2018). Genetic correlations among milk yield, morphology, performance test traits and somatic cells in dual-purpose Rendena breed. *Animal* 12, 906–914. doi:10.1017/S1751731117002543.

Schmid, M., Imort-Just, A., Emmerling, R., Fuerst, C., Hamann, H., and Bennewitz, J. (2021). Genotype-by-environment interactions at the trait level and total merit index level for milk production and functional traits in Brown Swiss cattle. *Animal* 15, 100052. doi:10.1016/j.animal.2020.100052.

Sechi, T., Usai, M. G., Miari, S., Mura, L., Casu, S., and Carta, A. (2007). Identifying native animals in crossbred populations: The case of the Sardinian goat population. *Anim. Genet.* 38, 614–620. doi:10.1111/j.1365-2052.2007.01655.x.

Senczuk, G., Mastrangelo, S., Ciani, E., Battaglini, L., Cendron, F., Ciampolini, R., et al. (2020). The genetic heritage of Alpine local cattle breeds using genomic SNP data. *Genet. Sel. Evol.* 52, 1–12. doi:10.1186/s12711-020-00559-1.

Shariati, M. M., Su, G., Madsen, P., and Sorensen, D. (2007). Analysis of milk production traits in early lactation using a reaction norm model with unknown covariates. *J. Dairy Sci.* 90, 5759–5766. doi:10.3168/jds.2007-0048.

Shi, R., Brito, L. F., Liu, A., Luo, H., Chen, Z., Liu, L., et al. (2021). Genotype-by-environment interaction in Holstein heifer fertility traits using single-step genomic reaction norm models. *BMC Genomics* 22, 1–20. doi:10.1186/s12864-021-07496-3.

Strucken, E. M., Bortfeldt, R. H., Tetens, J., Thaller, G., and Brockmann, G. A. (2012). Genetic effects and correlations between production and fertility traits and their dependency on the lactation-stage in Holstein Friesians. *BMC Genet.* 13, 108. doi:10.1186/1471-2156-13-108.

Tait-Burkard, C., Doeschl-Wilson, A., McGrew, M. J., Archibald, A. L., Sang, H. M., Houston, R. D., et al. (2018). Livestock 2.0 - Genome editing for fitter, healthier, and more productive farmed animals. *Genome Biol.* 19, 1–11. doi:10.1186/s13059-018-1583-1.

Tiezzi, F., Maltecca, C., Cecchinato, A., Penasa, M., and Bittante, G. (2012a). Genetic parameters for fertility of dairy heifers and cows at different parities and relationships with production traits in first lactation. *J. Dairy Sci.* 95, 7355–7362. doi:10.3168/jds.2012-5775.

Tiezzi, F., Maltecca, C., Cecchinato, A., Penasa, M., and Bittante, G. (2012b). Genetic parameters for fertility of dairy heifers and cows at different parities and relationships with production traits in first lactation. *J. Dairy Sci.* 95, 7355–7362. doi:https://doi.org/10.3168/jds.2012-5775.

Toledo-Alvarado, H., Cecchinato, A., and Bittante, G. (2017). Fertility traits of Holstein, Brown Swiss, Simmental, and Alpine Grey cows are differently affected by herd productivity and milk yield of individual cows. *J. Dairy Sci.* 100, 8220–8231. doi:10.3168/jds.2016-12442.

Van Soest, P. J. (1963). Ruminant Fat Metabolism with Particular Reference to Factors Affecting Low Milk Fat and Feed Efficiency. A Review. *J. Dairy Sci.* 46, 204–216.

doi:10.3168/jds.S0022-0302(63)89008-6.

Wallenbeck, A., Rydhmer, L., and Lundeheim, N. (2009). GxE interactions for growth and carcass leanness: Re-ranking of boars in organic and conventional pig production. *Livest. Sci.* 123, 154–160. doi:10.1016/j.livsci.2008.11.003.

Zhang, Z., Kargo, M., Liu, A., Thomasen, J. R., Pan, Y., and Su, G. (2019). Genotype-by-environment interaction of fertility traits in Danish Holstein cattle using a single-step genomic reaction norm model. *Heredity (Edinb).* 123, 202–214. doi:10.1038/s41437-019-0192-4.

8. GENOMIC PREDICTION IN LOCAL CATTLE BREEDS: RENDENA CATTLE AS CASE OF STUDY

STATUS: PUBLICATED ON ANIMALS
https://doi.org/10.3390/ani11061815

Genomic prediction in local breeds: the Rendena cattle as a case study

Enrico Mancin*, Beniamino Tuliozi, Cristina Sartori, Nadia Guzzo and Roberto Mantovani

ABSTRACT

The maintenance of local cattle breeds is key to select for efficient food production, landscape protection, and for conservation of biodiversity and local cultural heritage. Rendena is an indigenous cattle breed from the alpine North-East of Italy, selected for dual purpose, but with lesser emphasis to beef traits. In this situation, increasing accuracy for beef trait could prevent detrimental effect due to the antagonism with milk production. Our study assessed the impact of genomic information on Estimated Breeding Values (EBVs) in Rendena performance tested bulls. Traits considered were average daily gain, in vivo EUROP score, and in vivo estimate of dressing percentage. The final dataset contained 1,691 individuals with phenotypes, and 8372 animals in pedigree, 1743 of which genotyped. Using cross validation method three models were compared: i) Pedigree-BLUP (PBLUP); ii) single-step GBLUP (ssGBLUP), and iii) Weighted single-step GBLUP (WssGBLUP). Models including genomic information presented higher accuracy, especially WssGBLUP. However, the model with the best overall properties was the ssGBLUP, showing higher accuracy than PBLUP, and optimal value of bias and dispersion parameters. Our study demonstrated that integrating phenotypes for beef traits with genomic data can be helpful to estimate EBVs even in a small local breed.

INTRODUCTION

Rendena is dual-purpose cattle breed indigenous of the North-East of Italy. This breed is included within the “European Federation of Cattle Breeds of the Alpine System” (FERBA), an organization whose main purpose consists in the preservation and promotion of local cattle breeds of the alpine system (http://www.ferba.info, 20 April 2021). As it is the case with many indigenous breeds, a greater genetic diversity than specialized and cosmopolitan breeds is expected also for the Rendena [1]. This remarkable biodiversity is of great ecological importance and can be a beneficial factor for the survival of local population. Moreover, traditional breeds like Rendena provide additional benefits to local human population like economic advantages, ecosystem

services and also cultural benefits, such as preservation of cultural heritage and tradition of a specific area [1]. Rendena cattle shows also excellent values for traits concerning fertility and longevity, maintains a median milk production (5000 kg per lactation) and possesses a fairly good beef conformation [2]. Rendena cows are selected for both milk and meat, but with more emphasis on dairy production in the selection index [3], with dairy accounting for 65% and beef traits for 35% [4] Although beef attitude plays a less important role than milk in the selection index, an increase in the accuracy of the selection for this feature over time could prevent its detriment due to the antagonism with milk production [3]. Estimations of breeding values (EBVs) have until now mostly taken place using classical animal model analysis in Rendena through best linear unbiased predictor (BLUP;[3]). for traits related to milk, meat production and linear type traits However, several studies have shown how the use of genomic data can lead to an increase in prediction accuracy compared to using only pedigree information [5].

For a long time, two major limitations to the genomic selection approach on small populations such as Rendena have been the prohibitive cost of genotyping a sufficient number of Single Nucleotide Polymorphism (SNPs) per individual and the equations for EBVs' estimation, which were based on a multistep approach [6]. In fact, the drawback of the multistep approach in small populations is the scarce number of genotyped animals with phenotype to be used as reference population to ensure a good accuracy of prediction [6]. This is even more noticeable when sex-limited traits are considered [7]. To overcome this problem, methods such as the use of de-regressed proof [8,9] have been developed to allow the inclusion of animals whose only genotype is known, using progeny yield deviation adjusted for mates as pseudo-phenotype. However, this method presented some biases and lower accuracy whenever animals have few progenies with phenotype [10,11].

However, in the last few years both limitations preventing the use of the genomic selection approach in small breeds with limited diffusion such as Rendena have subsided. Firstly, the constant decline in prices of SNP platforms has allowed genomic selection to become much more cost-efficient. Secondly, equations such as the single step Genomic Best Linear Unbiased

Prediction (ssGBLUP) have been developed and found out to be suitable even in small-breed contexts [12]. The ssGBLUP evaluates simultaneously genotyped and non-genotyped animals by substituting the pedigree-based relationship matrix (A) present on BLUP, with a relationship matrix that combines pedigree and genomic information, usually called H [13]. Single step GBLUP represents a simple alternative to de-regressed proofs. Moreover, ssGBLUP offers the advantage of avoiding double counting contributions and it implicitly limits the bias of preselection for genotyped animals without phenotype [14–16]. Several studies have shown that ssGBLUP outperformed other methods in different livestock species in the context of genomic selection [17]. On the other hand, ssGBLUP might have its own drawback: genomic relationship matrix (G) included in single step assume that all SNPs explain the same amount of variance [18]. This may be a limit in the presence of traits influenced by many quantitative traits loci (QTL), such as some beef related traits like carcass weight and daily gain [19,20]. Indeed, some studies reported that SNP regression equations, in which prior assumption of SNPs effect and variance are modeled with different a priori assumption, outperformed the prediction of ssGBLUP [21]. On this point Zhang et al. [22] proposed to “relax” the assumption of the G matrix in which all SNPs equally contribute to the genomic variance of the traits, by adding specific SNPs weights. These methods are called weighted single step GBLUP (WssGBLUP), and in recent research it has shown to be effective by increasing the accuracy with respect to ssGBLUP for phenotypes like those related to the beef attitude [23]. In this study we investigated if the inclusion of genomic data in the estimate of breeding values for three key beef traits measured during performance tests in Rendena might increase their predictive accuracy with respect to traditional pedigree BLUP (PBLUP). In particular, the objective of this study was both to test different single-step GBLUP methods for beef traits and to measure their difference in accuracy using alternative weighting strategies, i.e., among different weighted single-step GBLUPs.

MATERIALS AND METHODS

Data availability

Phenotypic and pedigree data

Phenotypic and pedigree information were provided by the National Breeder Association of Rendena Cattle (www.anare.it). The phenotypes consisted in data recorded on Rendena young bulls during performance tests conducted from 1985 to the present. The phenotype used were the average daily gain (ADG) obtained by linear regression of weight on age recorded at least 11 times during the stay of bulls at the test performance test station, the mean in vivo fleshiness score (EUROP grade) and the mean in vivo estimate of dressing percentage (DP) evaluated by 3 skilled classifiers at the end of test, i.e., about 11 months of age. In vivo fleshiness score (EUROP grade) was linearly transformed as previously reported [4]. In the final dataset 1691 animals and as many phenotypes were present. The animals present in this dataset were born between 1985 and 2020. In addition, 8372 animals in pedigree were retrieved tracing back up to the 10th generation.

Table 1. Descriptive statistics of the target phenotypic data obtained from Rendena young bulls under performance test.

Taits[1]	**Number**[2]	**Mean**	**CV %**	**Min**	**Max**
ADG (kg/d)	1691 (690)	1024.00	12.06	474.00	1562
EUROP (points)	1691 (690)	99.05	3,84	80.00	111.10
DP (points)	1691 (690)	54.18	1.74	50.00	57.70

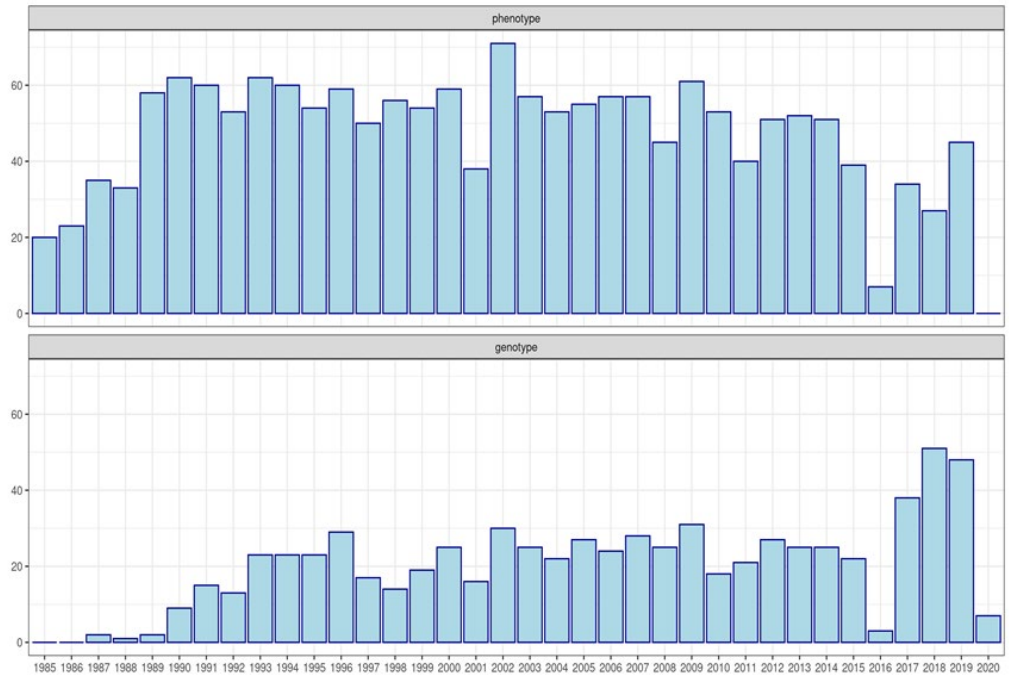
1 ADG= Average daily gain; EUROP = in vivo fleshiness score, DP = in vivo estimate of dressing percentage. 2 Number of animals with records and (genotype)

Genotype Data

Two genotyping platforms were used in this study, Illumina Bovine LD GGP v3, including 26 497 SNP markers (LD; no. = 1427) and Bovine 150K Array GGPv3 Bead Chip, comprising 138 974 SNPs (HD; no. = 554; Illumina Inc., San Diego, CA, USA). The higher density panel was used

only in 554 males whereas the remaining males (no; =174) and all the females (no. = 1253) were genotyped with the LD platform. Males genotyped with LD chips were all animals with at least one father and one full sib genotyped with the HD chip. The two panels shared about 60% of markers. Females with a call rate (CR) lower than 95% and male with a CR lower than 90% were discarded before the analysis. In addition, for both platforms SNPs with a minor allele frequency (MAF) <0.01 and call rate lower than 0.90 were removed with plink program [24]. Before genomic imputation, possible progeny conflicts were corrected with seekparentsf90 program [25]. The imputation of LD samples to HD density was performed with AlphaImpute2 program [26], that combines algorithms of population imputation with the use of imputation from pedigree information utilizing a sort of multi-locus iterative peeling [26]. To avoid excessive computational demand, we imputed one chromosome at each time. Threshold of loci inclusion to HD panels was set to 0.90 and a conservative genotype threshold for imputation of 0.99 was chosen. A further genomic quality control was then made for whole imputed panels (1953 individuals): SNPs with MAF lower than 0.05, with Hardy-Weinberg equilibrium lower than 0.15, and a call-rate under 0.90 were removed from dataset. In addition, animals with a call-rate under 0.90 were removed, in this case using preGSf90 program [25]. At the end of genotype editing, 1743 animals were retained for further analysis, 690 of which with both phenotype and genotype. The genotyped female are closed relatives with the male in the performances test, i.e, dams, or grad-dams.

Figure 1. Number of animals with phenotype (above) and number of animals with genotype (bottom) for all animals used in genetic or genomic prediction. X-axes represent the birth years and y-axis the number of animals per year.



Prediction Model

Pedigree Best Linear Unbiased Prediction (PBLUP):

The same fixed and random effects were used for all analyzed traits with the following model:

$$\mathrm{y} = \mathrm{Xb} + \mathrm{Za} + \mathrm{e} \qquad [1]$$

Where **y** is the vector of phenotypes, **X** represents the incident matrix for systematic fixed effects, **b** is the vector of fixed effects. Two cross-classified effects were used as in [4]: the contemporary group (142 levels), and the parity order of the cow (four classes: first parity; second parity; third to seventh parity; above the eighth parity included). **Z** is the incident matrix of random genetic additive effects, while a represents the vector of the additive genetic effects (EBVs) and e is the vector of residuals sampled from a distribution $N(0, I\sigma_e^2)$**,** where σ_e^2 is the residual variance. The additive genetic effect was sampled form a normal distribution with mean zero and variance

σ_a^2, and a covariance structure depending to the model used. In the PBLUP, model covariance of random genetic effect was sampled form a distribution $N(0, A\sigma_a^2)$, with **A** that represents the Identical by Descendent (IBD) matrix constructed from pedigree information. All genetic and genomic prediction models were carried out with the *blupf90* suite of programs [27].

The variance components used in all prediction scenarios were estimated under this model by univariate approach. In additions, genetic and residual correlation among traits was estimated with multi-traits models. Covariances' structures were **G**⊗**A,** and **R**⊗**I**, with **G**, and **R** that are 3x3 matrices respectively including the additive genetic and the residual (co)variances matrices, ⊗ is the Kronecker product, and **A** and **I** the additive relationships matrix and an identity matrix, respectively. Prior distributions for **G** and **R** matrices were independent inverse Wishart. Genetic and residual correlations (r_a) were calculated between trait pairs as ratio of the covariance on the square root of the product of the respective variances. Variances were estimated using Gibbs's sampling algorithm with gibbs3f90 program [27]. A chain of 200 000 iterations was used in both models. The first 5000 samples were discarded as burn-in. Samples were stored every 100 iterations to leave 1950 samples for inference.

Genomic Best Linear Unbiased Prediction (PBLUP):

In ssGBLUP inverse of the IBS matrix A^{-1} was replaced by the H^{-1} matrix as follows:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & (\alpha G + \beta A_{22})^{-1} - \mathrm{A}_{22}^{-1} \end{bmatrix} \quad [2]$$

where A^{-1} and A_{22}^{-1} represent the inverses of the IBD matrix for all individuals and only genotyped animals, respectively. To avoid singularity problems the bending coefficient α and β, were set to 0.95 and 0.05, respectively. A^{-1} was computed accounting for inbreeding to avoid inflation (bias) and to reduce distance between the two matrix as suggested elsewhere [28]. G is the genomic relationship matrix, built using the first method proposed in [18]:

$$G = \frac{MM'}{2\sum p_i(1-p_i)} [3]$$

where M is a matrix of SNP content centered by twice the current allele frequencies, and pi is the allele frequency for the ith SNP. In addition, variance components were re-estimated under this model, to evaluate variances changes by inclusion of genomic data. Therefore, G in ssGWAS is adjusted so the average diagonal and off-diagonal matches the averages of A22.

Weighted Single Step Genomic Best Linear Unbiased Prediction (WssGBLUP):

One The last method used for genetic prediction was WssGBLUP, that differs from ssGBLUP in the construction of $\mathbb{G}$. Particularly, $\mathbf{G_w}$ matrix was build using the following method [17]

$$G_w = \frac{MDM'}{2\sum p_i(1-p_i)} \qquad [4]$$

Where D is a diagonal matrix in which the elements of the diagonal correspond to the weight or effect of each SNP. Generally, SNPs' effects $(\hat{u})$ (are obtained as a function of the SNPs effect through a back-solving procedure from the EBVs solution obtained iteratively with the (W)ssGBLUP [29] as follows:

$$\hat{u} = \delta\alpha \frac{1}{2\sum p(1-p)} DM'[MDM']^{-1}\hat{a} \quad [5]$$

Where $\hat{a}$ is the vector of solutions of the genomic breeding values of the genotyped animals, and δ accounts for the difference in genetic base between the pedigree and genomic relationship.

An iterative algorithm following what reported in [16] was used. This algorithm consists in the subsequent steps:

1. Initial parameters are set $t = 1, D_{(t)} = I, G_{(t)} = \frac{MD_{(t)}M'}{2\sum p_i(1-p_i)}$.
2. GEBV ($\hat{a}$) are obtained using ssGBLUP algorithm.
3. Allele substitution effects for each SNP ($\hat{u}$) is reported in [5] with *postGSf90* [22].

4. Each $d_{i(t+1)}$ element of $D_{(t+1)}$, as $CT^{\frac{|\hat{u}_i|}{sd(\hat{u})}-2}$ is then calculated as in [18], where *CT* is a shrinkage factor determining how much the distribution of SNP effects departs from normality.

5. SNP weight are normalized by keeping genetic variance constant among iteration:

$$D_{(t+1)} = \frac{tr(D_{(1)})}{tr(D_{(t+1)})} tr(D_{(t+1)}).$$

6. G is then re-built with the new obtained weights as: $G_{(t+1)} = \frac{MD_{(t+1)}M'}{2\sum p_i(1-p_i)}$.

7. Further iterations are carried out up to convergence using WssGBLUP.

2.2.4. Weighted strategies

A further aim of this study was to identify optimal weight strategies to achieve higher accuracy and less biased genomic prediction. *NonlinearA* methods [18,30] was used as the weighting strategies. We focus on the effect of variance limitations (*limit*), and the shrinkage factor (*CT*). Other strategies, as linear weight [22,31] or Bayesian variable selection methods [32] were not applied on this study because of its excessive shrinkage that led to high biased prediction and incompatibility between **A** and **G** matrix, as reported in supplementary material S1.

Three values of *CT* were used in this study (1.105, 1.125 and 1.250), considering that values greater than one deviate proportionally from a normal distribution and exhibit grater shrinkage. By default, *postGSf90* program set maximum change in SNPs variance equal to $CT^{(5-2)}$, thus default limitations for the three parameters were automatically set to 1.350, 1.424 and 1.953 for *CT* equal to 1.105, 1.125 and 1.250. Other scenarios have been explored setting the maximum change on variance equal to 5.

LR cross validations

Estimators of bias, dispersion and accuracy were adopted to evaluate the different prediction models. LR cross-validation method was used on this behalf [33]. In this approach two

datasets (whole and partial) are used, and parameters described above are estimated in a set of focal individuals. The whole dataset contains all populations information while partial dataset includes a subset of phenotypic data up to a given date. In this study the 2015 was set as cut-off year and the focal individuals are the younger bulls with only genotype information (i.e., born after 2015; 109 animals). The focal individuals represented the young animals of interest for selection and in most of the cases they represented young "genomic" candidate to selection [33]. Simply speaking, focal individuals are the animals for which accuracy of prediction is of greater interest for selection. LR defined bias as: $bias = \overline{\hat{u}_p} - \overline{\hat{u}_w}$, where $\hat{u}_p$ is the estimate of individual EBV in the partial dataset and $\hat{u}_w$ is the estimate of individual EBV in the whole dataset. Bias equal to 0 stands for unbiased prediction. Due to different magnitudes of each trait, bias was also standardized by the genetic standard deviation of each trait analyzed. Dispersion was described by the slope of the regression between EBVs in whole dataset to EBVs in the partial one, i.e., $disp = \frac{cov(\hat{u}_w, \hat{u}_p)}{var(\hat{u}_p)}$, with an expectation of 1, i.e., *disp* <1 designate over-dispersion while *disp*>1 indicates an under-dispersion. In this study, we refer as accuracy (*acc*) the correlation of breeding values estimated in the two datasets [33]: $acc = \frac{cov(\hat{u}_w, \hat{u}_p)}{\sqrt{var(\hat{u}_p) var(\hat{u}_w)}}$. This estimator stands for the inverse of accuracy gain when the phenotype was added moving from partial dataset to the whole one. Low values of the "*acc*" estimator mean that the EBV estimate of focal group is mainly influenced by the addition of new phenotypic information as respect to the conditional kinship information. Thus, $E(acc) \approx \frac{acc_p}{acc_w}$. Furthermore, reliability, the squared accuracy, was obtained through the following approximated expression: $rel = \frac{cov(\hat{u}_w, \hat{u}_p)}{(1-\hat{F})\sigma_u^2}$, where $\hat{F}$ is the average population inbreeding coefficient and σ_u^2 is the genetic variance estimated in the whole dataset. The expected value for rel is equal to acc^2, and the adequacy of this estimator was proofed on appendix 1 in [34]. Note that no differences were observed in terms of variance components between the whole dataset and the focal groups thus for that reason adjustment by selected reliability proposed in [35] was not applied. In addition, according to [34], the increase in accuracy when genomic data are introduced was estimated as $inc = {\rho_{A,G}}^{-1} - 1$, where $\rho_{A,G} =$

$\frac{cov(\hat{u}_A,\hat{u}_G)}{\sqrt{var(\hat{u}_A)var(\hat{u}_G)}}$, $\hat{u}_A$ is the EBV estimated with PBLUP in the partial dataset and $\hat{u}_G$ is EBV estimated using genomic information in the partial dataset. In fact, using the same reasoning done for ***acc,*** $\rho_{A,G}$ quantifies the increase of the inverse of accuracy when genomic data are added because his expected valued is $\frac{acc_A}{acc_G}$. A further evaluation of the increase in accuracy due to genomic data was also obtained following [34], that suggested to adjust the increase in accuracy for the ratio of genetic variances of two models accounting or not for genomic information, i.e., $inc_{adj} = \frac{\sigma_A^2}{\sigma_G^2} inc$, where σ_A^2 is the genetic variance estimated with only pedigree information and $\sigma_G^2$is the variance when genomic information is included. For matter of simplicity, only *inc_adj* has been reported as parameter that identify the increase of accuracy. Note that EBV in the focal populations are normally distributed, thus condition under LR assumption were not violated.

RESULTS

Variance components

Heritability (h^2), genetic and residual correlations estimated using PBLUP are reported on table 2. All traits presented a medium to high heritability. EUROP was the trait with lowest heritability, 0.304, while ADG and DP showed a h^2 of 0.335 and 0.392, respectively. In addition, all traits' pairs, as expected, presented medium to high genetic and residual correlations. ADG presented a medium positive genetic correlation with the other two traits (0.38 on average), while DP and EUROP resulted strongly correlated (0.981) to be considered a unique trait.

Table 2. Mean of genetic (upper diagonal) and residual (lower diagonal) correlations and heritability (diagonal) between traits in Rendena population, estimated with PBLUP. Number in parenthesis are the lower and the upper 95% highest posterior density.

	ADG	EUROP	DP
ADG	0.335 (0.204 ± 0.335)	0.364 (0.100 ± 0.597)	0.398 (0.148± 0.6315)
EUROP	0.572 (0.660 ± 0.742)	0.304 (0.174 ± 0.446)	0.981 (0.962 ± 0.997)
DP	0.613 (0.517 ± 0.702)	0.792 (0.753 ± 0.836)	0.392 (0.248 ± 0.541)

ADG= Average daily gain, EUROP =and in vivo fleshiness score CY, DP = in vivo estimate of dressing percentage.

Table 3 reported estimated heritability, genetic and residual correlations using ssGBLUP. In this case both h^2 and correlations resulted similar to those estimated with the PBLUP. For what concerns h^2, ADG decreased of about 0.02, while EUROP increased of about 0.04 in ssGBLUP as compared to PBLUP. On the other hand, DP remained basically unchanged comparing the two approaches. Correlations presented almost the same values in both analyses, with the only exceptions of the genetic and residual correlations between ADG and EUROP that resulted increased in ssGBLUP of about 0.02 and 0.08, respectively.

Table 3. Mean of genetic (upper diagonal) and residual (lower diagonal) correlation and heritability (diagonal) between traits in Rendena population, estimated with ssGBLUP. Number in parenthesis are the lower and the upper 95% highest posterior density.

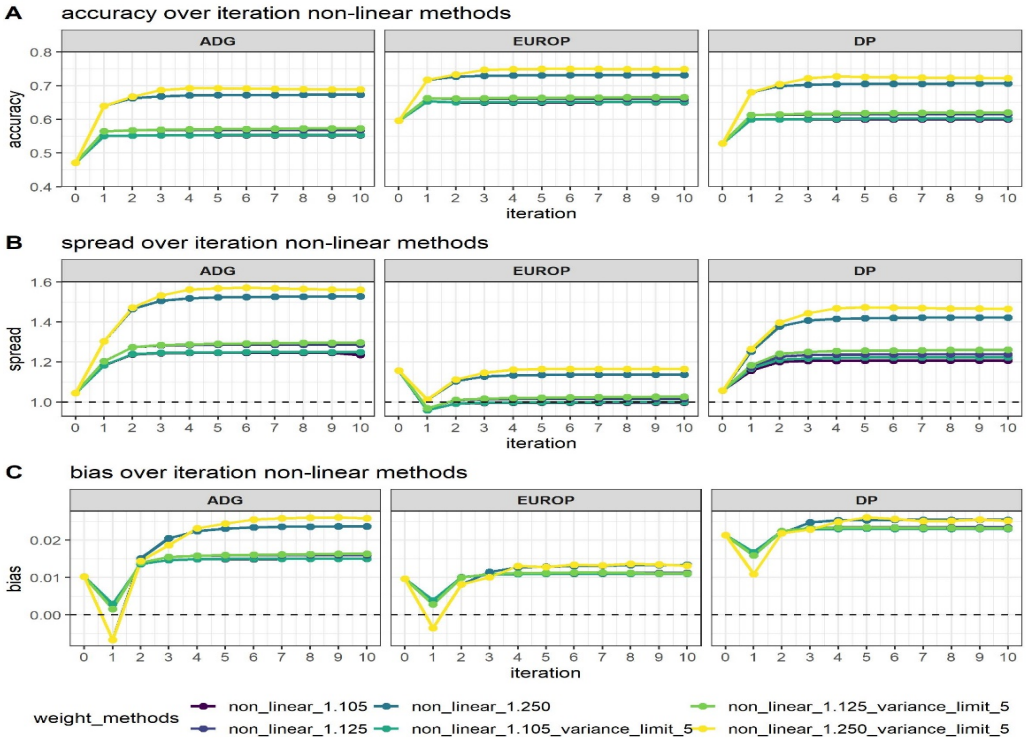
	ADG	EUROP	DP
ADG	0.313 (0.223± 0.489)	0.385 (0.153 ± 0.597)	0.392 (0.160± 0.622)
EUROP	0.651 (0.651 ± 0.718)	0.345 (0.216 ± 0.487)	0.985 (0.961± 0.999)
CY	0.616 (0.530 ± 0.671)	0.790 (0.753 ± 0.826)	0.396 (0.250 ± 0.530)

ADG= Average daily gain, EUROP =and in vivo fleshiness score CY, DP = in vivo estimate of dressing percentage

Weighting strategies

Figure 2 shows how different values of *CT* and the limitation of SNPs' variance can affect genomic prediction.

Figure 2. Accuracy (A), dispersion (B) and bias corrected by genetic standard deviations (C) of breeding value estimated using different weighting strategies along the 10 iterations process of the algorithm used in WssGBLUP. Dotted line in graph B and C represents the expected value.



As expected, higher accuracy (Figure 2 A) was reached in the WssGBLUP analyses with the increase of the number of iterations, although in most cases the asymptote was reached at the 2nd iteration, with the only exception of the CT 1.25 with the limit of maximum variance established at 5, that reached the maximum accuracy after 3-4 iterations. Variance limits did not affected accuracy using a CT of 1.105 or 1.125.

Bias (Figure 2 C) followed the same trends in all phenotypes; at first iteration bias was even lower than with ssGBLUP, but when iterations increased, bias rapidly increased. ADG presented higher biases, even if difference in magnitude was considered by standardizing values obtained. Even dispersion (spread; Figure 2 B) followed the same trends as accuracy with an increase after 2/4 iterations depending mainly on the value attributed to CT. For all traits, as the interactions increased, dispersion departed from the expected value of 1, although for EUROP the use of CT at 1.125 was maintained steadily close to 1.

In general, higher CT values (that is, greater departures from normality) presented better accuracy but more under-dispersion and biases. When CT changed from 1.105 to 1.125 accuracy increase of 2% in all phenotypes, and substantial increase of accuracy was observed moving to a CT value of 1.250 (+20 % on average). When the threshold for maximum SNPs variance was raised up to 5, accuracy increased slightly, especially from the 3rd to 10th iteration.

Figures 3, 4, and 5 show the percentage of variance explained by a sliding window of 20 non overlapping SNPs. These plots show how the different values of CT and limit influenced the shrinkage SNPs. Furthermore, observing the peaks in the Manhattan plots, it can be seen how these traits are potentially controlled by few QTLs. The high peak found on chromosome 22 for EUROP and DP can explain why these traits are highly genetically correlated.

Figure 3. Manhattan plots for Average Daily Gain (ADG) using different WssGBLUP strategies in iteration equal to 10; y-axes represent the percentage explained by each SNPs. Variance explained were calculated with sliding window approach.

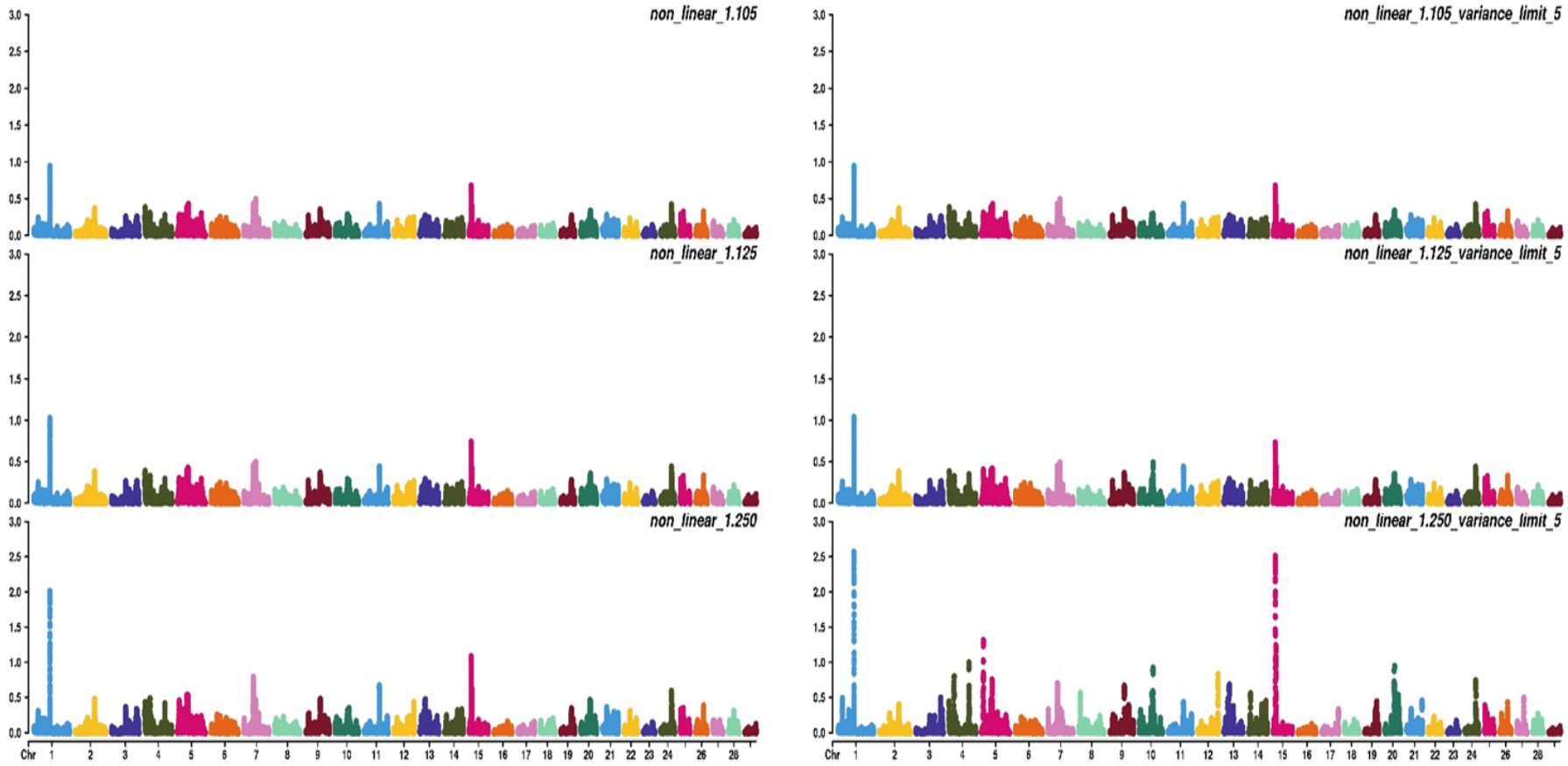


Figure 4. Manhattan plots for fleshiness score (EUROP) using different WssGBLUP strategies in iteration equal to 10; y-axes represent the percentage explained by each SNPs. Variance explained were calculated with sliding window approach.

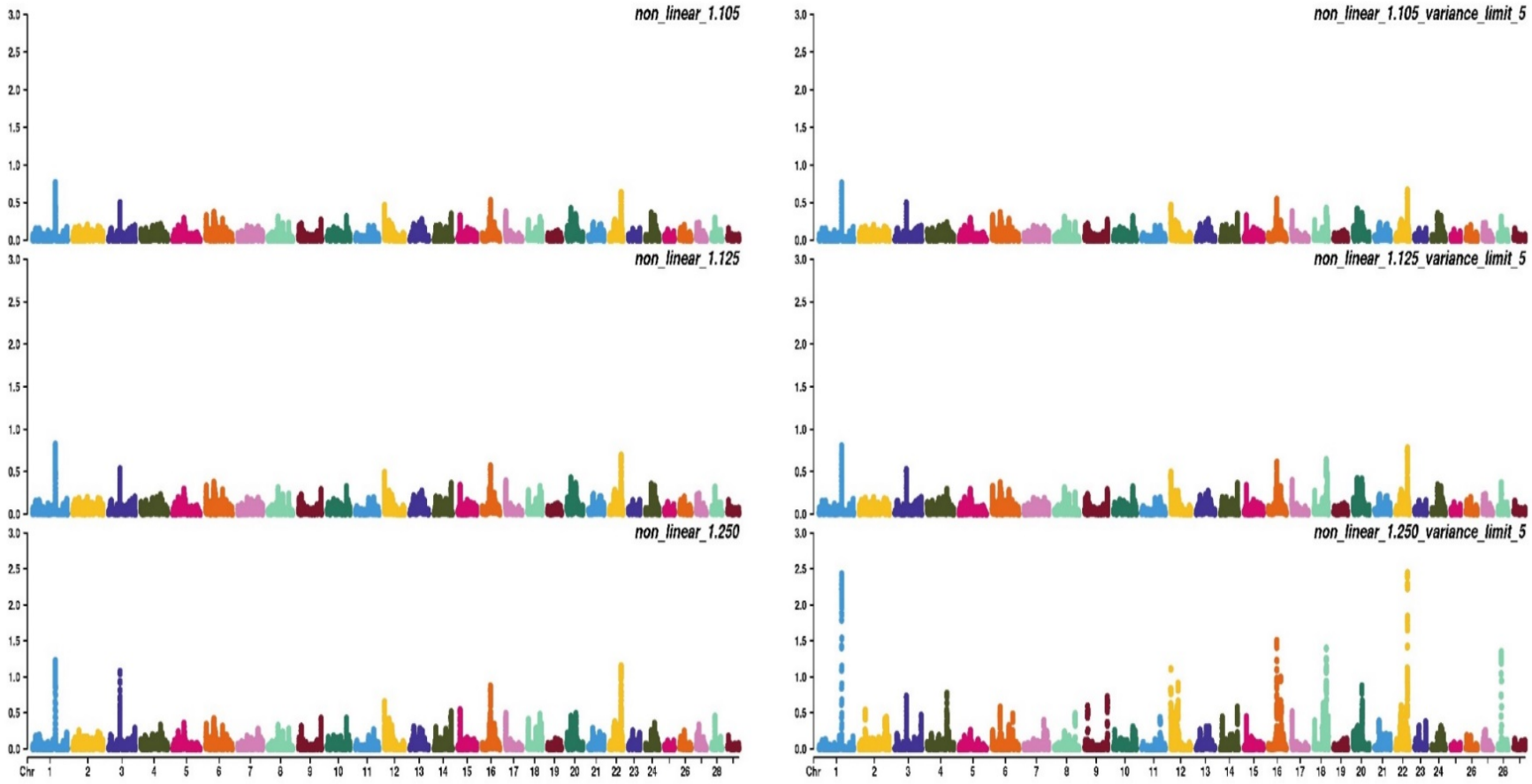
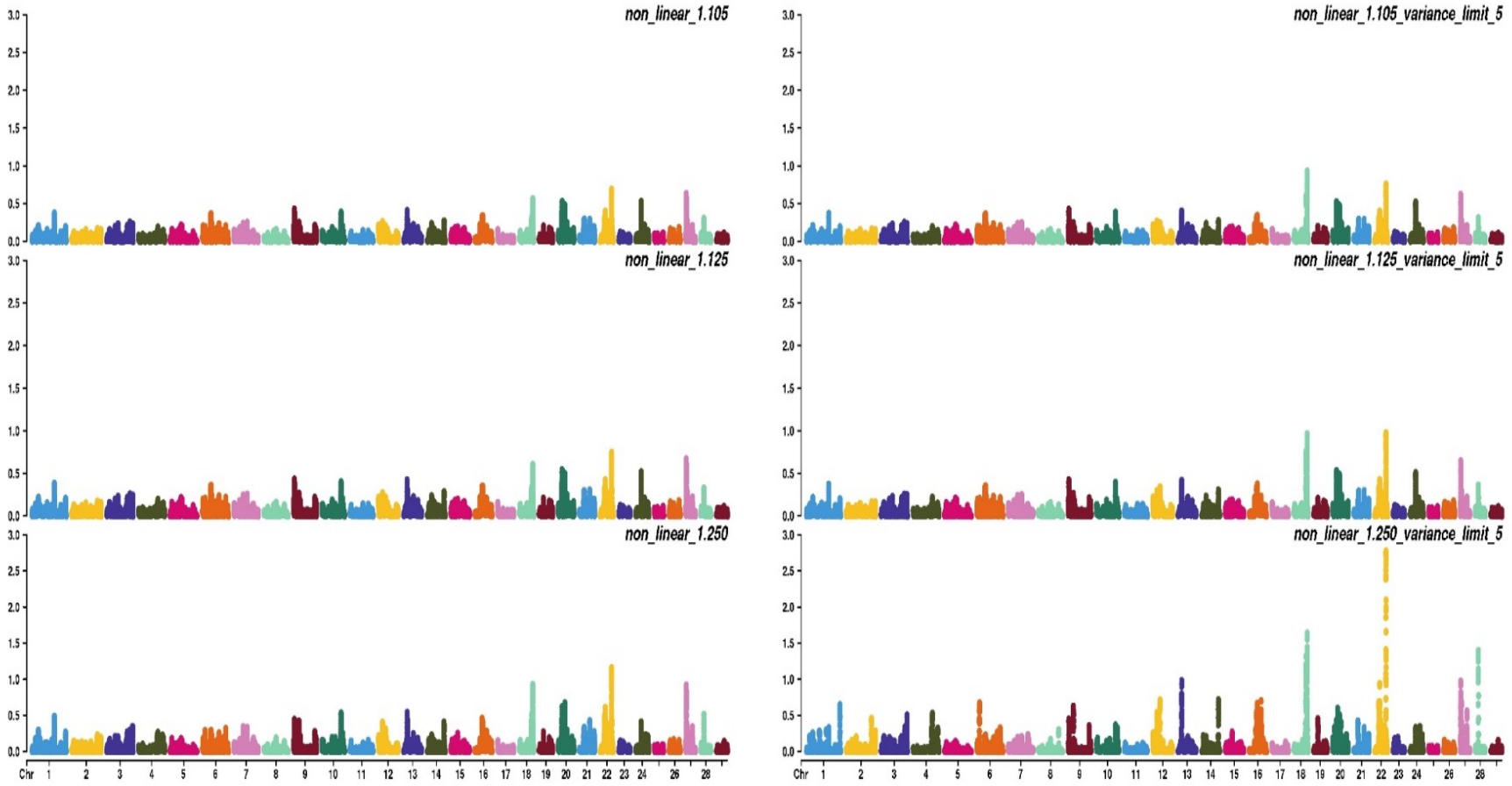


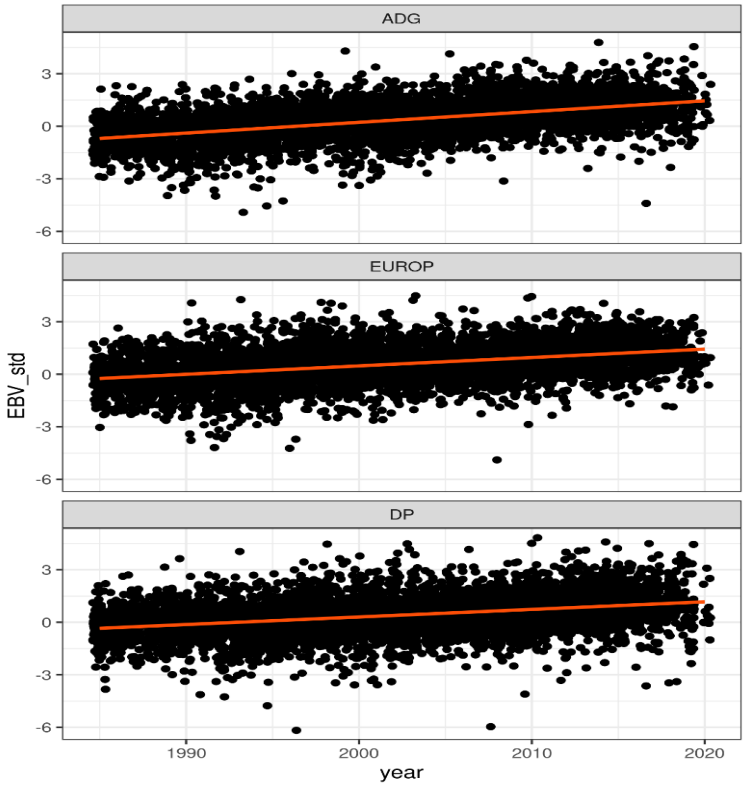
Figure 5. Manhattan plots for dressing percentage (DP) using different WssGBLUP strategies in iteration equal to 10; y-axes represent the percentage explained by each SNPs. Variance explained were calculated with sliding window approach.



Model comparison

From the previous analysis, for each phenotype, two weighting strategies were retrieved: the one presenting a value of bias close to the optimal value (WssGBLUP_1) and the one with highest accuracy (WssGBLUP_2) The weighting strategies that produced the lowest bias were associated with a CT = 1.105, default value for limit and iteration 1. On the other hand, as reported previously, CT = 1.250 and limit equal to 5 produced the best results in terms accuracy of prediction. For ADG and DP maximum accuracy value was found on iteration 4, while for EUROP iteration 7 was most successful (Figure 2). Table 4 shows the different performances of prediction of PBLUP, ssGBLUP and the two selected WssGBLUP obtained under the LR cross-validation method. EUROP presented the highest accuracy in all models considered, followed by DP and ADG. All traits presented bias value close to 0, although DP presented a slightly positive bias of about 0.02 on average. Generally, all models except WssGBLUP_2 showed very low biased prediction, also considering that estimated genetic progress per years is consistent, being positive and equivalent to 0.58, 0.42, and 0.33 standard deviations for ADG, EUROP and DP, respectively (Figure 6).

Figure 6. Standardize genetic progress per each year: x-axis indicates birth year of animals and y-axis the standardize EBV, from 1985 (when performance test stared) to 2020 (current data).



For what concerned dispersion parameter, in this study we found that ADG and DP were slightly under-dispersed, while EUROP was a little over-dispersed for PBLUP and WssGBLUP_1, showing values of dispersion < 1.

When only pedigree information was used, lower accuracy was observed for all traits: ADG presented a value of 0.366, EUROP of 0.464, and DP of 0.506. Lower reliability values were also found in this model. Interesting, PBLUP is the only prediction model in which a marginally negative bias was observed. PBLUP presented similarly biased values as WssGBLUP_1 and ssGBLUP for EUROP and DP (same absolute value but opposite sign), while for ADG it presented a greater absolute value of bias than for the other two methods.

When genomic information was added, a global increase in accuracy and reliability was observed. In ssGBLUP models an increase of accuracy of 0.106, 0.087, and 0.064 was observed for ADG, EUROP and DP, respectively. Reliability estimators showed the same trend. The ssGBLUP had higher accuracy values with respect to PBLUP, and it also presented bias and dispersion closest to optimal value. As it can be seen from Figure 2 and Table 4, higher accuracy and reliability values were observed as SNPs shrinkage increased (that is, for higher values of CT): however, in parallel, more under-dispersion and biased predictions were found.

The inc_adj estimator represents the increase of accuracy when genomic models were used. ADG is the trait that was most favored from the introduction of genomic data, with a value of 45% and DP is the one with lower benefits (26.7%). WssGBLUP_1 presents similar inc_adj value than ssGBLUP, while in WssGBLUP_2 value rises to 4 percentage points in ADG as well as to 5 and 7 percentage points in EUROP and DP.

Table 4. Accuracy, bias, dispersion (Disp.) and reliability (Rel.) and adjusted increased of accuracy (Incr_adj) of estimated breeding values under different models: pedigree BLUP (PBLUP), single step genomic BLUP (ssGBLUP) and weight single step with bias value closet to optimal value (WssGBLUP_1) and weight single step with highest accuracy; for average daily gain (ADG), EUROP and dressing percentage (DP).

Trait	Model	Accuracy	Bias	Disp.	Rel.	Incr_adj
ADG	PBLUP	0.366	-0.040	1.140	0.060	-
	ssGBLUP	0.472	0.010	1.045	0.117	45.10%
	WssGBLUP_1	0.551	0.003	1.182	0.127	45.10%
	WssGBLUP_2	0.693	0.020	1.562	0.206	49.21%
EUROP	PBLUP	0.509	-0.009	0.902	0.081	-
	ssGBLUP	0.596	0.009	1.100	0.124	39.98%
	WssGBLUP_1	0.653	0.004	0.958	0.135	39.98%
	WssGBLUP_2	0.749	0.014	1.165	0.192	45.17%
DP	PBLUP	0.464	-0.021	1.114	0.114	-
	ssGBLUP	0.528	0.021	1.056	0.158	26.70%
	WssGBLUP_1	0.600	0.017	1.156	0.184	27.40%
	WssGBLUP_2	0.727	0.025	1.468	0.277	33.90%

DISCUSSION

In this study we evaluated how the use of genomic data can improve the estimates of breeding values in the local dual purpose Rendena cattle. We used data relative to beef traits collected in performance test of young bulls both because these traits are accounted for in the selection index, and also because of the smaller number of genotyped individuals needed with respect to traits such as milk production. In addition, this was the first approach to apply genomic selection in a small local cattle breed.

The three performance test phenotypes i.e., ADG, EUROP and DP, presented medium to high heritability, ranging from 0.30 to 0.40. We recorded little difference from the study of [4] even if our dataset was greater by an amount of about 40%. The heritabilities of these traits were similar to the ones observed in other dual-purpose (i.e., Alpine Grey; [36] or beef

specialized breeds [37–39]). All traits appeared highly genetically correlated, especially EUROP and DP, as expected and widely reported in literature [38,40,41]. Interestingly, after we introduced genomic data, we did not observe many discrepancies in terms of genetic (co)variance(s) with respect to those estimated with PBLUP. This is in agreement with what was reported in [42], i.e., that even for non-random genotyping strategies, the population variances in ssGBLUP are not influenced by the selective genotyping strategies as much as they are with GBLUP [43]. In fact, thanks to the contribution of non-genotyped animals present in the pedigree, the bias due to the preselection of genotyped animals in ssGBLUP is reduced. Furthermore, the genotypes resulted homogeneously distributed over years and this factor may have undoubtedly contributed to reducing discrepancy in terms of variance estimates.

The usefulness of genomic selection was assessed using LR as a cross-validation method, which provided accuracies, bias, and dispersion of the genetic evaluations. LR presents several advantages [33]: the robustness of genetic evaluations is inferred on a target group of animals, i.e., accuracy can be evaluated at the level of the preferred sub-group of the population. In our study our focus was on young bulls and close relatives, the sub-group in which phenotypic data were collected. In addition, another advantage consists in the fact that LR does not require the precorrection of phenotypes, thus avoiding potentially biased prediction due the heterogeneity of contemporary group (number of animals range from 4 to 20 animals per group [33]).

Results confirmed that when genomic data were integrated with pedigree there was a substantial increase in the accuracy of (G)EBVs prediction. Accuracy increases of about 30% on average when switching from BLUP to ssGBLUP. Moreover, an additional increase of accuracy was observed when weighting strategies were applied, i.e., from 0.366 to 0.472 for ADG, from 0.509 to 0.569 for EUROP, and from 0.464 to 0.528 for DP, respectively. These outcomes suggest that the genomic information can potentially capture variation in Mendelian sampling and thus leading to a greater accuracy of prediction when only kinship information is used [44]. A similar impact of ssGBLUP on the accuracy of performance test traits has been observed in Hanwoo beef cattle [45], in which the same number of phenotypes and genotypes were used; however, results cannot be compared numerically due to different cross-validation strategies implemented.

The findings of previous studies report that the ssGBLUP led to more accurate predictions than the BLUP. Other research conducted on a different type of beef-related traits presented a substantial increase in breeding value prediction when ssGBLUP was used. However, those investigations were conducted with breeds with much larger population sizes and results were expressed in terms of reliability [46–48]. Interestingly, Cesarani et al. [49] an analogous number of animals was used and results in terms of bias and dispersion agree with results obtained in this manuscript, with a similar influence of weighting strategies, although the number of animals with genotype in their study is much lower than ours. While generally different weighting strategies have led to different increases in the accuracy of the breeding value predictions [21], extreme shrinkage strategies (i.e., quadratic weight) can lead rapidly to a decline in accuracy as the interactions increase and generally present greater biased prediction [23]. These weighting strategies have thus been discarded from this study due to the excessive shrinkage caused by the influence of major QTLs (supplementary material). In addition, an extreme shrinkage can lead to an incompatibility between G and A matrix, consequently losing some properties of the single step such as unbiasedness of selection.

To that reason, nonlinearA methods were chosen over the other weighting strategies. The consistency of nonlinearA methods in a single step framework has been reported by [30]. The augment of accuracy of weighting strategies is particularly relevant when small datasets are used, such as in the present study [50]; moreover, the use of heterogeneous SNP weighting is useful when the number of SNPs exceed the number of animals [50]. This point could be relevant for our study since redundant information can be produced also by the genomic imputation [51]. In fact, according with [52] when the trait is controlled by a few QLTs and few genotyped animals are present, the information relative to the trait in the genome is usually divided in few blocks and consequently most of the SNPs information is considered redundant. Assigning different values to SNPs or to chromosome segments can remove redundant SNPs information [53]. The presence of major QTLs has a positive impact on WssGBLUP because the relationships between animals are focused on SNPs which are clearly linked to the QTLs [54].

Despite this, LR cross validation methods pointed out that major under-dispersion and bias is observed by applying WssGBLUP. In our study population proven and/or young animals are evaluated with the rest of performance test animals. The higher bias present in some of the weighting models led to an inaccurate estimation of genetic trends than in turn

lead a potentially biased selection decision, i.e., selecting only young animals as respect the older ones [35]. Because of that, the bias and dispersion parameters must be considered alongside the accuracy of selection [55]. For this reason, models over 2nd iteration can be discarded from the choice of model with “best” properties, due to the lack of mean’s exact estimation in selected animals [33]. Interestingly, decline of biases was observed in 1st iteration for all phenotypes. Conversely, PBLUP and ssGBLUP confirmed their unbiasedness prediction and ability to account implicitly for selection [17]. In addition, ssGBLUP presented dispersion parameters closest to the optimal value of one and it demonstrated the consistency of this estimator of this type of model. Indeed, dispersion represents regression of EBV from whole to partial data, thus making the model less affected by the addition/subtraction of information, and therefore the best model to be applied.

Our finding supports the use of genomic data, and in particular the use of ssGBLUP, as the new model for the routinary genetic evaluation on selected bulls of a local breed, the Rendena cattle. In local breeds genomic information has mainly been used to assess genetic variability or to study specific biological pathways underscoring peculiar traits such as [56]. As mentioned above, this is a first study investigating the impact of genomic information on selection in indigenous breeds [57]. We focused on performance test traits because of the antagonistic relation to milk traits, but genomic selection can be successfully applied also for other traits, depending on the amount of phenotypic and genomic information available. Notably, the increase in the accuracy of selection can impact the economic value of the breed [58], which is a pragmatic and effective strategy to guarantee the conservation of local breeds.

The present study shows that genomic imputation and the combination of genotyped and non-genotyped data through ssGBLUP could be a cost-efficiency strategy, compensating the limited genotype information available on local breeds. This could make genomic selection for limited populations an appealing strategy, as it already is in more cosmopolite breeds like Holstein [59] However, the LR cross-validation demonstrated that accuracy increased only in young bulls with genotypes, while the accuracy of non-genotyped animals was only marginally higher than that obtained with the PBLUP, in the subgroups of individuals with a genotyped close. For this reason, we would recommend, to keep increasing selection accuracy, that a majority of animals for each performance test cycle should still continue to be genotyped.

CONCLUSIONS

All models that included genomic data presented higher accuracy and reliability than the ones using only kinship information. These two estimators were particularly higher in models in which high heterogeneous variances among SNPs had been assumed; however, the same models presented under-dispersion and higher bias, and for that reason they can be discarded as models to be used in the selection. Models with “best properties” can be identified in the ssGBLUP or in the WssGBLUP, in which weighting strategies presented less shrinkage. Although these two models presented similar proprieties, ssGBLUP could be chosen as “best” model because it was neither under nor over-dispersed, presenting appropriate properties for long-term selection. In conclusion, the present study demonstrated how the use of genomic data in addition to ssGBLUP can lead to a better prediction of genetic effects even with a modest amount of molecular data, as typically happens in local populations. Therefore, we demonstrated how genomic data can be a suitable tool for breeding selection scenarios in local cattle breeds, as Rendena, guaranteeing the competitiveness and thus the conservation of the breed through its improvement of selection's accuracy**.**

Supplementary Materials: Supplementary material 1. Value of Accuracy, Dispersion and Bias divided by the genetic standard deviations (bias_std) for Average Daily Gain. (1) Models presented are Pedigree BLUP (PBLUP), single step genomic BLUP (ssGBLUP) and different weighting single step described as follow: non_linear is referred to the nonlinear weighting strategies presented in the manuscript with the respective CT value, limit_5 is referred when variance was set up to a maximum of 5. While quadratic referred to the quadratic weight applied to the SNP solutions, sliding stands for the quadratic weight applied to a window of sliding SNPs. iter stands for the number of iterations, NA values means that it was not possible to obtain the solution due to blending problem between between $\mathbf{A}^{-1}$ and $\mathbf{G}^{-}$

Supplementary material 2. Value of Accuracy, Dispersion and Bias divided by the genetic standard deviations (bias_std) for EUROP (1) Models presented are Pedigree BLUP (PBLUP), single step genomic BLUP (ssGBLUP) and different weighting single step described as follow: non_linear is referred to the nonlinear weighting strategies presented in the manuscript with the respective CT value, limit_5 is referred when variance was set up to a maximum of 5. While quadratic referred to the quadratic weight applied to the SNP solutions,

sliding stands for the quadratic weight applied to a window of sliding SNPs. iter stands for the number of iterations, NA values means that it was not possible to obtain the solution due to blending problem between between $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$

Supplementary material 3. Value of Accuracy, Dispersion and Bias divided by the genetic standard deviations (bias_std) for Dressing Percentage (DP). Models presented are Pedigree BLUP (PBLUP), single step genomic BLUP (ssGBLUP) and different weighting single step described as follow: non_linear is referred to the nonlinear weighting strategies presented in the manuscript with the respective CT value, limit_5 is referred when variance was set up to a maximum of 5. While quadratic referred to the quadratic weight applied to the SNP solutions, sliding stands for the quadratic weight applied to a window of sliding SNPs. iter stands for the number of iterations, NA values means that it was not possible to obtain the solution due to blending problem between $\mathbf{A}^{-1}$ and $\mathbf{G}^{-}$

Acknowledgment: Authors are grateful to National Breeders Association of Rendena cattle breed (ANARE) for data support and Renzo Bonifazi for helping in the discussion. The study was funded by the DUALBREEDING project (CUP J61J18000030005) and by BIRD183281

Author Contributions: Conceptualization, E.M.; methodology, B.T, E.M.; formal analysis, E.M.; investigation, B.T. C. S. R.M.and E.M.; resources, R.M.; data curation E.M., B.T. and R.M.; writing—original draft preparation, E.M.; B.T. writing—review and editing R.M., C.S., B.T. All authors have read and agreed to the published version of the manuscript

Reference

1. Schöpke, K.; Swalve, H.H. Review: Opportunities and challenges for small populations of dairy cattle in the era of genomics. *animal* **2016**, *10*, 1050–1060, doi:10.1017/S1751731116000410.

2. Mazza, S.; Guzzo, N.; Sartori, C.; Berry, D.P.; Mantovani, R. Genetic parameters for linear type traits in the Rendena dual-purpose breed. **2014**, *131*, 27–35, doi:10.1111/jbg.12049.

3. Sartori, C.; Guzzo, N.; Mazza, S.; Mantovani, R. Genetic correlations among milk yield, morphology, performance test traits and somatic cells in dual-purpose Rendena breed. *Animal* **2018**, *12*, 906–914, doi:10.1017/S1751731117002543.

4. Guzzo, N.; Sartori, C.; Mantovani, R. Analysis of genetic correlations between beef traits in young bulls and primiparous cows belonging to the dual-purpose Rendena breed. *animal* **2019**, *13*, 694–701, doi:10.1017/S1751731118001969.

5. Zhang, H.; Yin, L.; Wang, M.; Yuan, X.; Liu, X. Factors Affecting the Accuracy of Genomic Selection for Agricultural Economic Traits in Maize, Cattle, and Pig Populations. *Front. Genet.* **2019**, *10*, 189, doi:10.3389/fgene.2019.00189.

6. Blasco, A.; Toro, M.A. A short critical history of the application of genomics to animal breeding. *Livest. Sci.* **2014**, *166*, 4–9, doi:10.1016/j.livsci.2014.03.015.

7. Ibanez-Escriche, N.; Simianer, H. Animal breeding in the genomics era. *Anim. Front.* **2016**, *6*, 4, doi:10.2527/af.2016-0001.

8. VanRaden, P.M.; Wiggans, G.R. Derivation, Calculation, and Use of National Animal Model Information. *J. Dairy Sci.* **1991**, *74*, 2737–2746, doi:10.3168/jds.S0022-0302(91)78453-1.

9. VanRaden, P.M.; Van Tassell, C.P.; Wiggans, G.R.; Sonstegard, T.S.; Schnabel, R.D.; Taylor, J.F.; Schenkel, F.S. Invited review: reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* **2009**, *92*, 16–24, doi:10.3168/jds.2008-1514.

10. Masuda, Y.; VanRaden, P.M.; Misztal, I.; Lawlor, T.J. Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection. *J. Dairy Sci.* **2018**, *101*, 5194–5206, doi:10.3168/jds.2017-13310.

11. Ma, P.; Lund, M.S.; Nielsen, U.S.; Aamand, G.P.; Su, G. Single-step genomic model improved reliability and reduced the bias of genomic predictions in Danish Jersey. *J. Dairy Sci.* **2015**, *98*, 9026–9034, doi:https://doi.org/10.3168/jds.2015-9703.

12. Aguilar, I.; Misztal, I.; Johnson, D.L.; Legarra, A.; Tsuruta, S.; Lawlor, T.J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* **2010**, *93*, 743–752, doi:10.3168/jds.2009-2730.

13. Legarra, A.; Aguilar, I.; Misztal, I. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* **2009**, *92*, 4656–4663, doi:10.3168/jds.2009-2061.

14. Misztal, I.; Lourenco, D.; Legarra, A. Current status of genomic evaluation. *J. Anim. Sci.* **2020**, *98*, doi:10.1093/jas/skaa101.

15. Legarra, A.; Christensen, O.F.; Aguilar, I.; Misztal, I. Single Step, a general approach for genomic selection. *Livest. Sci.* **2014**, *166*, 54–65, doi:10.1016/j.livsci.2014.04.029.

16. Ricard, A.; Danvy, S.; Legarra, A. Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in French show-jumping horses1. *J. Anim. Sci.* **2013**, *91*, 1076–1085, doi:10.2527/jas.2012-5256.

17. Misztal, I.; Aggrey, S.E.; Muir, W.M. Experiences with a single-step genome evaluation. *Poult. Sci.* **2013**, *92*, 2530–2534, doi:10.3382/ps.2012-02739.

8. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423, doi:10.3168/jds.2007-0980.

19. Bedhane, M.; van der Werf, J.; Gondro, C.; Duijvesteijn, N.; Lim, D.; Park, B.; Park, M.N.; Hee, R.S.; Clark, S. Genome-Wide Association Study of Meat Quality Traits in Hanwoo Beef Cattle Using Imputed Whole-Genome Sequence Data. *Front. Genet.* **2019**, *10*, 1235, doi:10.3389/fgene.2019.01235.

20. Fu, L.; Jiang, Y.; Wang, C.; Mei, M.; Zhou, Z.; Jiang, Y.; Song, H.; Ding, X. A Genome-Wide Association Study on Feed Efficiency Related Traits in Landrace Pigs. *Front. Genet.* **2020**, *11*, 692, doi:10.3389/fgene.2020.00692.

21. Mehrban, H.; Naserkheil, M.; Lee, D.H.; Cho, C.; Choi, T.; Park, M.; Ibáñez-Escriche, N. Genomic Prediction Using Alternative Strategies of Weighted Single-Step Genomic BLUP for Yearling Weight and Carcass Traits in Hanwoo Beef Cattle. *Genes (Basel).* **2021**, *12*, doi:10.3390/genes12020266.

22. Zhang, X.; Lourenco, D.; Aguilar, I.; Legarra, A.; Misztal, I. Weighting Strategies for Single-Step Genomic BLUP: An Iterative Approach for Accurate Calculation of GEBV and GWAS. *Front. Genet.* **2016**, *7*, 151, doi:10.3389/fgene.2016.00151.

23. Mehrban, H.; Naserkheil, M.; Lee, D.H.; Cho, C.; Choi, T.; Park, M.; Ibáñez-Escriche, N. Genomic Prediction Using Alternative Strategies of Weighted Single-Step Genomic BLUP for Yearling Weight and Carcass Traits in Hanwoo Beef Cattle. *Genes (Basel).* **2021**, *12*, 266, doi:10.3390/genes12020266.

24. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575, doi:10.1086/519795.

25. Aguilar, I; Tsuruta, S.; Masuda, Y.; Lourenco, D.A.L.; Legarra, A.; Misztal, I. BLUPF90 suite of programs for animal breeding. *11th World Congr. Genet. Appl. to Livest. Prod.* **2018**, 11.751.

26. Whalen, A.; Hickey, J.M. AlphaImpute2: Fast and accurate pedigree and population based imputation for hundreds of thousands of individuals in livestock populations. *bioRxiv* **2020**, doi:10.1101/2020.09.16.299677.

27. Misztal, I.; Tsuruta, S.; Lourenco, D.; Aguilar, I.; Legarra, A.; Vitezica, Z. Manual for BLUPF90 family of programs. *Univ. Georg. Athens, USA* **2018**, 125.

28. Garcia-Baccino, C.A.; Legarra, A.; Christensen, O.F.; Misztal, I.; Pocrnic, I.; Vitezica, Z.G.; Cantet, R.J.C. Metafounders are related to Fst fixation indices and reduce bias in single-step genomic evaluations. *Genet. Sel. Evol.* **2017**, *49*, 34, doi:10.1186/s12711-017-0309-2.

29. Strandén, I.; Garrick, D.J. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.* **2009**, *92*, 2971–2975, doi:10.3168/jds.2008-1929.

30. Fragomeni, B.O.; Lourenco, D.A.L.; Legarra, A.; VanRaden, P.M.; Misztal, I. Alternative SNP weighting for single-step genomic best linear unbiased predictor evaluation of stature in US Holsteins in the presence of selected sequence variants. *J. Dairy Sci.* **2019**, *102*, 10012–10019, doi:10.3168/jds.2019-16262.

31. Lopez, B.I.; Lee, S.H.; Shin, D.H.; Oh, J.D.; Chai, H.H.; Park, W.; Park, J.E.; Lim, D. Accuracy of genomic evaluation using imputed high-density genotypes for carcass traits in commercial Hanwoo population. *Livest. Sci.* **2020**, *241*, 104256, doi:10.1016/j.livsci.2020.104256.

32. Tiezzi, F.; Maltecca, C. Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. *Genet. Sel. Evol.* **2015**, *47*, 24, doi:10.1186/s12711-015-0100-1.

33. Legarra, A.; Reverter, A. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method 01 Mathematical Sciences 0104 Statistics. *Genet. Sel. Evol.* **2018**, *50*, 1–18, doi:10.1186/s12711-018-0426-6.

34. Bermann, M.; Legarra, A.; Hollifield, M.K.; Masuda, Y.; Lourenco, D.; Misztal, I. Validation of single-step GBLUP genomic predictions from threshold models using the linear regression method: An application in chicken mortality. *J. Anim. Breed. Genet.* **2021**, *138*, 4–13, doi:10.1111/jbg.12507.

35. Macedo, F.L.; Christensen, O.F.; Astruc, J.M.; Aguilar, I.; Masuda, Y.; Legarra, A. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genet. Sel. Evol.* **2020**, *52*, 1–10, doi:10.1186/s12711-020-00567-1.

36. Mancin, E.; Sartori, C.; Guzzo, N.; Tuliozi, B.; Mantovani, R. Selection Response Due to Different Combination of Antagonistic Milk, Beef, and Morphological Traits in the Alpine Grey Cattle Breed. *Animals* **2021**, *11*, doi:10.3390/ani11051340.

37. Sbarra, F.; Mantovani, R.; Quaglia, A.; Bittante, G. Genetics of slaughter precocity, carcass weight, and carcass weight gain in Chianina, Marchigiana, and Romagnola young bulls under protected geographical indication. *J. Anim. Sci.* **2013**, *91*, 2596–2604, doi:10.2527/jas.2013-6235.

38. Bonfatti, V.; Albera, A.; Carnier, P. Genetic associations between daily BW gain and live fleshiness of station-tested young bulls and carcass and meat quality traits of commercial intact males in Piemontese cattle. *J. Anim. Sci.* **2013**, *91*, 2057–2066, doi:10.2527/jas.2012-5386.

39. Aass, L. Variation in carcass and meat quality traits and their relations to growth in dual purpose cattle. *Livest. Prod. Sci.* **1996**, *46*, 1–12, doi:10.1016/0301-6226(96)00005-X.

40. De Haas, Y.; Janss, L.L.G.; Kadarmideen, H.N. Genetic and phenotypic parameters for conformation and yield traits in three Swiss dairy cattle breeds. *J. Anim. Breed. Genet.* **2007**, *124*, 12–19, doi:10.1111/j.1439-0388.2007.00630.x.

41. Jensen, J.; Mao, I.L.; Andersen, B.B.; Madsen, P. Genetic parameters of growth, feed intake, feed conversion and carcass composition of dual-purpose bulls in performance testing. *J. Anim. Sci.* **1991**, *69*, 931–939, doi:10.2527/1991.693931x.

42. Cesarani, A.; Pocrnic, I.; Macciotta, N.P.P.; Fragomeni, B.O.; Misztal, I.; Lourenco, D.A.L. Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies. *J. Anim. Breed. Genet.* **2019**, *136*, 40–50, doi:10.1111/jbg.12367.

43. Gowane, G.R.; Lee, S.H.; Clark, S.; Moghaddar, N.; Al-Mamun, H.A.; van der Werf, J.H.J. Effect of selection and selective genotyping for creation of reference on bias and accuracy of genomic prediction. *J. Anim. Breed. Genet.* **2019**, *136*, 390–407, doi:10.1111/jbg.12420.

44. Habier, D.; Fernando, R.L.; Dekkers, J.C.M. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **2007**, *177*, 2389–2397, doi:10.1534/genetics.107.081190.

45. Mehrban, H.; Lee, D.H.; Naserkheil, M.; Moradi, M.H.; Ibáñez-Escriche, N. Comparison of conventional BLUP and single-step genomic BLUP evaluations for yearling weight and carcass traits in Hanwoo beef cattle using single trait and multi-trait models. *PLoS One* **2019**, *14*, e0223352.

46. Onogi, A.; Ogino, A.; Komatsu, T.; Shoji, N.; Simizu, K.; Kurogi, K.; Yasumori, T.; Togashi, K.; Iwata, H. Genomic prediction in Japanese Black cattle: Application of a single-step approach to beef cattle. *J. Anim. Sci.* **2014**, *92*, 1931–1938, doi:10.2527/jas.2014-7168.

47. Gordo, D.G.M.; Espigolan, R.; Tonussi, R.L.; Júnior, G.A.F.; Bresolin, T.; Magalhães, A.F.B.; Feitosa, F.L.; Baldi, F.; Carvalheiro, R.; Tonhati, H.; et al. Genetic parameter estimates for carcass traits and visual scores including or not genomic information1. *J. Anim. Sci.* **2016**, *94*, 1821–1826, doi:10.2527/jas.2015-0134.

48. Lee, J.; Cheng, H.; Garrick, D.; Golden, B.; Dekkers, J.; Park, K.; Lee, D.; Fernando, R. Comparison of alternative approaches to single-trait genomic prediction using genotyped and non-genotyped Hanwoo beef cattle. *Genet. Sel. Evol.* **2017**, *49*, 2, doi:10.1186/s12711-016-0279-9.

49. Cesarani, A.; Biffani, S.; Garcia, A.; Lourenco, D.; Bertolini, G.; Neglia, G.; Misztal, I.; Macciotta, N.P.P. Genomic investigation of milk production in Italian buffalo. *Ital. J. Anim. Sci.* **2021**, *20*, 539–547, doi:10.1080/1828051X.2021.1902404.

50. Lourenco, D.A.L.; Fragomeni, B.O.; Bradford, H.L.; Menezes, I.R.; Ferraz, J.B.S.; Aguilar, I.; Tsuruta, S.; Misztal, I. Implications of SNP weighting on single-step genomic predictions for different reference population sizes. *J. Anim. Breed. Genet.* **2017**, *134*, 463–471, doi:10.1111/jbg.12288.

51. Mancin, E.; Sosa-Madrid, B.S.; Blasco, A.; Ibáñez-Escriche, N. Genotype imputation to improve the cost-efficiency of genomic selection in rabbits. *Animals* **2021**, *11*, 1–16, doi:10.3390/ani11030803.

52. Pocrnic, I.; Lourenco, D.A.L.; Masuda, Y.; Legarra, A.; Misztal, I. The dimensionality of genomic information and its effect on genomic prediction. *Genetics* **2016**, *203*, 573–581, doi:10.1534/genetics.116.187013.

53. Hassani, S.; Saatchi, M.; Fernando, R.L.; Garrick, D.J. Accuracy of prediction of simulated polygenic phenotypes and their underlying quantitative trait loci genotypes using real or imputed whole-genome markers in cattle. *Genet. Sel. Evol.* **2015**, *47*, 1–11, doi:10.1186/s12711-015-0179-4.

54. Habier, D.; Fernando, R.L.; Garrick, D.J. Genomic BLUP decoded: A look into the black box of genomic prediction. *Genetics* **2013**, *194*, 597–607, doi:10.1534/genetics.113.152207.

55. Legarra, A. Reverter, A. Can We Frame and Understand Cross-Validation Results in Animal Breeding? *Proc. Assoc. Advmt. Anim. Breed. Genet.* **2017**, *22*, 73–80.

56 Senczuk, G., Mastrangelo, S., Ciani, E. Elena Ciani, L. Battaglini, F. Cendron, R. Ciampolini, P. Crepaldi, R. Mantovani, G. Bongioni, G.Pagnacco, B. Portolano, A. Rossoni, F. Pilla, M. Cassandro. The genetic heritage of Alpine local cattle breeds using genomic SNP data. Genet Sel Evol 52, 40 (2020). https://doi.org/10.1186/s12711-020-00559-1

57. Mrode, R.; Ojango, J.M.K.; Okeyo, A.M.; Mwacharo, J.M. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects. *Front. Genet.* **2019**, *10*, doi:10.3389/fgene.2018.00694.

58. Biscarini, F.; Nicolazzi, E.; Alessandra, S.; Boettcher, P.; Gandini, G. Challenges and opportunities in genetic improvement of local livestock breeds. *Front. Genet.* **2015**, *5*, 1–16, doi:10.3389/fgene.2015.00033.

59. García-ruiz, A.; Cole, J.B.; Paul, M.; Wiggans, G.R.; Ruiz-lópez, F.J.; Curtis, P. Erratum: Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection (Proceedings of the National Academy of Sciences of the United States of America (2016) 113 (E3995-E4004) DOI:10.1073. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E4928, doi:10.1073/pnas.1611570113.

9. IMPROVEMENT OF GENOMIC PREDICTIONS IN SMALL BREEDS BY CONSTRUCTION OF GENOMIC RELATIONSHIP MATRIX THROUGH VARIABLE SELECTION

STATUS: UNDER REVISION ON FRONTIERS IN GENETICS

Improvement of genomic predictions in small breeds by construction of genomic relationship matrix through variable selection

Enrico Mancin, Lucio Flavio Macedo Mota, Beniamino Tuliozi, Roberto Mantovani, Cristina Sartori

ABSTRACT

Genomic selection has been increasingly implemented in the animal breeding industry, and it is starting to become a routinary method in many livestock breeding contexts. However, its use is still limited in several small-population local breeds, which are nonetheless an important source of genetic variability of great economic value. A major roadblock for their genomic selection is the accuracy when population size is limited: for this reason, to improve the accuracy of breeding values, variable selections models that assume heterogeneous variance have been proposed over the last few years. However, while these models might outperform traditional and genomic predictions in terms of accuracy, they also carry a proportional increase of breeding values bias and dispersion. These mutual increases are especially striking when genomic selection is performed with a low number of phenotypes and high shrinkage value – which is precisely the type of situation that happens with small local breeds. In our study, we investigate several alternative methods to improve the accuracy of genomic selection in a small population. We investigated the impact of using only a subset of informative markers regarding accuracy of prediction, bias, and dispersion. We tested different machine learning variable selection algorithms to select them as recursive feature eliminations, penalized regression and XGboost. We compared our results with the predictions of pedigree based BLUP, single step genomic BLUP and weighted single step genomic BLUP in different simulated populations obtained by combining different parameters in terms of number of QTL and effective population size. We also investigated these approaches on a dataset belonging to the small local Rendena breed. Our results show that the accuracy of GBLUP in small sized populations increased when performed with SNPs selected via variable selection methods both in simulated and actual datasets. In addition, the use of variable selection models – especially those using XGboost – in our actual dataset did not impact bias and the dispersion of estimated breeding values. We discuss possible explanations for our results and how our study can help estimate breeding values for future genomic selection in small breeds.

INTRODUCTION

Genomic information has been successfully implemented in animal breeding due to its effectiveness in bringing significant improvements in accuracy (Blasco and Toro, 2014). These improvements in accuracy can lead to an increase in the rate of genetic gains and have reduced the cost of progeny testing by allowing to pre-select animals with great genetic merit early (Meuwissen et al., 2001). Combining these advances with the progressively reduced cost of genotyping makes Single Nucleotide Polymorphism (SNP) panels a promising tool to be used even in the selection of small local breeds (Biscarini et al., 2015).

SNP markers information allow for better modelling of the Mendelian Sampling compared to the traditional pedigree-based approach Best Linear Unbiased Prediction (PBLUP) (VanRaden, 2008a), which used only pedigree information. The genomic BLUP (GBLUP) method was developed to replace the pedigree-based relationships for genomic relationships estimated from SNP markers, which captured the genomic similarity between animals but are limited to the use of only genotyped animals (Habier et al., 2013). In addition, Legarra et al. (2009) proposed a naive method, single-step GBLUP (ssGBLUP), in which genotyped and non-genotyped animals are jointly combined under the assumption that the genomic and pedigree relationship matrixes are multivariate-normally distributed. Due to its straightforward computational approach (Misztal et al., 2013) and to its unbiased breeding values predictions compared to the GBLUP with its multi-step approach (Masuda et al., 2018), the ssGBLUP has become a routinary method for genomic evaluations in many livestock breeds and species (Aguilar et al., 2010; Christensen and Lund, 2010).

However, one major challenge in using (ss)GBLUP remains the accuracy of estimation when phenotyped animals are limited in number, such as in local breeds (Meuwissen et al., 2001). For example, Karaman et al. (2016) reported that GBLUP showed lower performance than models using only SNPs selected through Bayesian hierarchical model as Bayes B and C, but only when phenotyped animals were few. Indeed, when presented with a small number of animals and many SNP markers ($n < p$), models that select a number of priority SNPs (variable selection models) and models that assume heterogeneous variance can lead to improvements in EBVs' accuracy. These models can accomplish this by reducing the number of variables to estimate and by preventing the over-fitting linked to high-dimensional data (Gianola 2013). Frouin et al. (2020) went as far as deriving prediction accuracy of GBLUP as a

function of the ratio n/p, while Pocrnic et al. (2019) regarded the accuracy of GBLUP as not only dependent on the number of SNPs but also on the number of independent chromosome segments.

Several studies thus focused on relaxing the assumption of ssGBLUP that all SNPs must show a common variance, by applying different weights to the SNPs when the **G** matrix is calculated. Methods such as weighted ssGBLUP (WssGBLUP; Wang et al., 2014) were widely reported to outperform ssGBLUP accuracy of prediction (Gualdrón Duarte et al., 2020; Mehrban et al., 2021; Ren et al., 2021), but their use led to a proportional increase of breeding values bias and dispersion (Botelho et al., 2021; Cesarani et al., 2021; Mancin et al., 2021; Mehrban et al., 2021).

Moreover, it is not clear how models considering heterogeneous variances account for selection, since usually only k-folds cross validation is applied (Zhu et al., 2021). In real-life breeding scenarios time-cross validation should be taken into account (Liu, 2010) as this validation method mimics the true accumulation of information across time. The estimated breeding values (EBVs) are in fact used to select young bulls and after 3-5 years the bulls will receive daughter information and it is thus desirable that EBVs would highly correlate to the final EBVs. However, the few studies that evaluated the impact of WssGBLUP using time cross-validation with small samples of individuals (e.g., Cesarani et al., 2021) found higher bias and overdispersion. These mutual increases are relevant when a low number of phenotypes and high shrinkage values are present, and the reasons behind the loss of these unbiased properties in heterogenous SNP regression or GBLUP are still not entirely clear.

This issue is not trivial, as the bias and the slope of the regression (dispersion) need to be taken into account especially when proven and young animals are mixed in the population, as young candidates will have unfairly high EBVs (Legarra, Reverter, 2017).

Thus, the abovementioned issues of lack of accuracy of ssGBLUP when used in contexts with a low number of animals have not yet been conclusively resolved. For this reason, in the present study, we intend to explore alternative methods to improve accuracy in small populations within a single step framework. A possible solution could come from implementing a naïve approach, where instead of giving to each SNP a specific weight we removed the non-informative ones, or variable selection models.

We thus aim to investigate the impact, in terms of accuracy of predictions, but also of dispersion and bias, of reducing the dimensionality of **G** through the construction of this matrix using only a subset of informative markers.

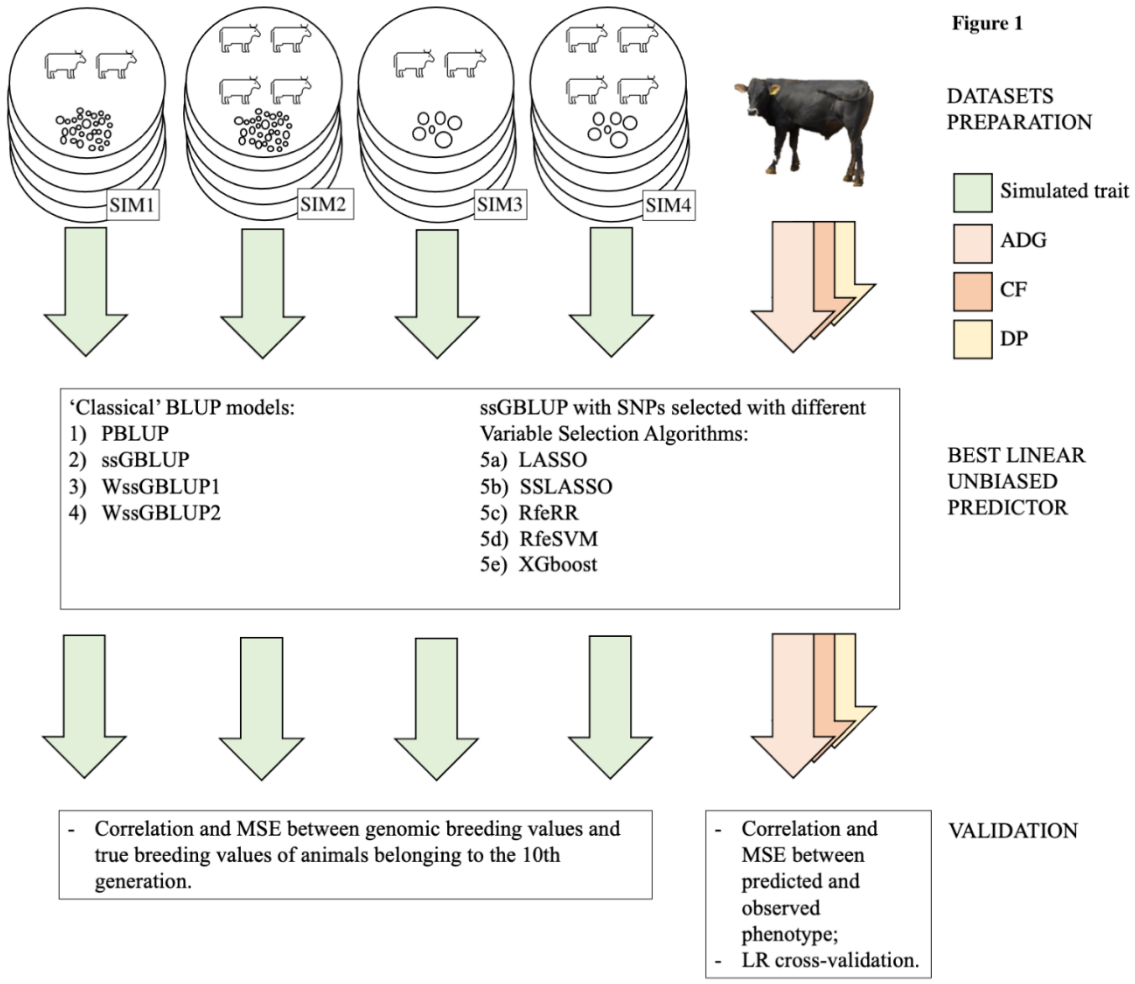
Because of this, we tried different machine learning and variables selection algorithms with the aim to identify the most informative SNPs by indirect prediction. These algorithms are: least absolute shrinkage and selection operator (LASSO), Spike and Slab Lasso (SSLASSO), Recursive Feature Elimination using Ridge Regression (RfeRR), Recursive Feature Elimination using Support Vector Machine regression (RfeSVM) and Extreme Gradient Boost (XGboost).

Our aim was to test suitable procedures for genomic estimation by considering both the abovementioned variable selection models ssGBLUP with the predictions of BLUP, classical ssGBLUP, and WssGBLUP. In order to do that we created different simulated populations and also considered a local population, the Rendena cattle. We then employed different cross-validation methods to assess our results.

MATERIALS AND METHODS

Graphical representation of our methodology for testing BLUP models see Figure 1.

Figure 1 Graphical representation of our methodology



Datasets

Simulated datasets

Simulations were performed with QMSim simulation program (Sargolzaei and Schenkel, 2009). Four different populations were simulated based on different combinations of quantitative trait loci (QTL) number and effective population size (Ne). Each simulation was replicated five times.

All simulations were generated starting from the historical population using a similar structure to Pocrnic et al. (2019): an initial bottleneck was generated contracting the historical population size from 5,000 to 1,000 animals in 1,250 generations, then expanded to 25,000. In the first generation, 10 bovine autosomes were simulated, placing evenly spaced 80,000 ca. biallelic SNPs with equal allele frequencies and a recurrent mutation rate of $2.5e^{-5}$ per generation. The number of SNPs per chromosome was set to 8000, while QTL number changed according to different simulation strategies. In two of the four simulations, one biallelic and randomly distributed QTL per chromosome was sampled from a gamma distribution with a shape parameter equal to 0.4 (oligogenic scenarios). In the other two simulations, 100 QTL per chromosome were generated using the same parameter (polygenic scenarios). In all these simulations, 10 discrete generations were created randomly mating 750 females and different number of sires according to the simulation strategies. In two scenarios, one oligogenic and one polygenic, we assumed a large Ne, with 100 males per generation used as sires, while in simulations with a low Ne only 10 males per generation were used as sires. The following four populations were thus created by mixing the different numbers of QTL and different Ne values, and five replicates for each population were generated:

SIM1 polygenic population with small Ne

SIM2 polygenic population with large Ne

SIM3 oligogenic population with small Ne

SIM4 oligogenic population with large Ne

The effective population size and number of QTL in the four different simulated populations are reported in Table 1 and numbers of genotyped animals are reported in Table 2 (2,250 animals). As a phenotype, a single trait with heritability of 0.3 was simulated, thus obtaining a single phenotype record per animal by adding an overall mean of 1.0 to the sum of the QTL effects plus a residual effect. As in Pocrnic et al. (2019), only phenotypes from generations 8 to 9 were retrieved, and genomic information of animals belonging to generations 8 to 10 was used for further analysis. The structure of simulated populations is reported in Table 2. Before proceeding with genomic prediction, SNPs with a minor allele

frequency (MAF < 0.01) and with high linkage disequilibrium (LD > 80) were removed. SNPrune was used to this purpose (Calus and Vandenplas, 2018).

Table 1 Effective population size and number of QTL in the four different simulated populations

	QTN	Ne
SIM1	1000	40
SIM2	10	350
SIM3	1000	40
SIM4	10	350

Actual dataset

An actual dataset containing information from the performance test evaluations of young bulls belonging to the Rendena cattle breed, was provided by the National Breeding Association (ANARE). ANARE also provided Herd Book information of the whole population traced back to the 1950s, whereas genomic data of bulls were in part provided by ANARE (PSRN DualBreeding, www.dualbreeding.it) and in part obtained under academic funding (SID Project, BIRD183281). Rendena is a small local population (6,384 heads for 249 breeding males and 6,135 breeding females belonging to 202 herds censed at 31.12.2020; fao/dad.is.org) bred for the dual-purpose attitude of milk and meat. Rendena is native of the North-East Alps in Italy and spread in the adjacent territory (Guzzo et al., 2018), and is thus now present in two characteristic locations, i.e., the origin mountain area and the plain dairy area located on the right side of the Brenta river in the Veneto Region (Po Valley).

Phenotypes considered in this study included single records per individual collected in the years 1985-2020 and are: Average Daily Gain (ADG), in vivo estimates of Carcass fleshiness (CF) and Dressing percentage (DP). These traits have been extensively described in Guzzo et al. (2019) and Mancin et al. (2021b). The Illumina Bovine LD GGP v3, comprising 26,497 SNP markers and Bovine 150K Array GGPv3 Bead Chip, including 138,974 SNPs (Illumina Inc., San Diego, CA, USA) was used for genotyping the Rendena cattle.

The LD panels belonging to 1,416 individuals were imputed on HD panels belonging to 554 bulls. The overlap between the two panels was about 60%. Information about data quality control and imputation are reported in greater detail in Mancin et al. (2022). In addition to the previous study, further quality control was performed by removing SNPs with high linkage disequilibrium > 80, using SNPrune (Calus and Vandenplas, 2018), that removed a total of 28,049 SNPs. A final amount of 85,331 SNPs was finally retained for analysis. Overall, the study considered 1,691 young bulls with only phenotypic information, 1,739 animals with only genotypic information, and 687 animals with both phenotypic and genotypic information. The data structure of the actual dataset used for genomic prediction is reported in Table 2.

Table 2 Population structure of simulated and actual data set

	SIMULATED		ACTUAL
	SIM1-SIM3[1]	SIM2-SIM4[1]	
Number of records	1500	1500	1691
Number of animals in the pedigree	3413	3794	6926
Number of genotyped animals	2250	2250	1739
Number of genotyped animals with records	1500	1500	687
Inbreeding from Pedigree	0.0126	0.0009	0.0316

[1] Since population structure is the same for SIM1 and SIM3, and for SIM2 and SIM4, populations were grouped together in pairs in the table

Prediction models

The breeding values for the single trait of the four simulated populations and the three performance test traits of the actual Rendena dataset were estimated using several BLUP models. Firstly, we used four 'classical' BLUP models:

1) standard Pedigree Best Linear Unbiased Prediction (PBLUP, described in section 2.2.1);
2) single step Genomic BLUP (ssGBLUP, described in section 2.2.2);
3) small shrinkage Weighted single-step Genomic BLUP (WssGBLUP1, described in section 2.2.3);

4) high shrinkage Weighted single-step Genomic BLUP (WssGBLUP2, described in section 2.2.3).

Then, we performed five ssGBLUP with preselected SNPs (described in 2.3.4). SNP selection was performed through different algorithms:

5a) Least absolute shrinkage and selection operator (LASSO, described in section 2.4.1)
5b) Spike-and-Slab LASSO (SSLASSO, described in section 2.4.2);
5c) Recursive Feature Elimination using Ridge Regression (RfeRR, described in section 2.4.3);
5d) Recursive Feature Elimination using Support Vector Machine (RfeSVM, described in section 2.4.4);
5e) Extreme Gradient Boosting (XGboost, described in section 2.4.5).

Pedigree Best Linear Unbiased Predictor

The PBLUP was the first method that has been applied, described by the following equation (Henderson, 1975):

$$\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

Where **y** is the vector of phenotypic observations, while **X** is the matrix of incidence of fixed effect and, **b** is the vector of these effects. In the actual dataset, fixed effects are represented by contemporary group (young bulls tested at the same period in the same pen; 142 levels) and parity group of dams in four classes (Guzzo et al., 2019). Whereas for simulated dataset **X** was substituted by a vector of **1**'s, consequently **b** stands for the mean of the models. Matrix **Z** represents the incidence matrix that relates the random genetic additive effect, included in vector $\boldsymbol{a}$, to the phenotype. The random residual error was included in a vector **e** showing a normal distribution $N(0, I\sigma_e^2)$**,** where σ_e^2 is the residual variance. The vector of additive genetic effects is distributed as $N(0, A\sigma_a^2)$, where σ_a^2 is the genetic variances and **A** is the Identical by Descent (IBD) relationship matrix constructed from pedigree data.

Single Step Genomic Best Linear Unbiased Predictor

The ssGBLUP was considered as a benchmark to evaluate the impact of the other models (see further, WssGBLUP and ssGBLUP with selected SNPs). The ssGBLUP presents the same structure of equation in 2.2.1., except for the co-variance matrix of random genetic effect, which is substituted by **H**, as described in Aguilar et al. (2010):

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G^{-1}} - \mathbf{A_{22}^{-1}} \end{bmatrix}$$

where $\mathbf{A}$ and $\mathbf{A_{22}^{-1}}$ are the reverse of IBD matrix for all animals and for only genotyped animals, respectively, and **G** was the genomic matrix including the genomic relationships among animals.

The **G** matrix was built using the first methods proposed by (VanRaden, 2008b):

$$\mathbf{G_0} = \frac{\mathbf{MM'}}{\mathbf{2}\sum \boldsymbol{p_i}\,(\mathbf{1} - \boldsymbol{p_i})}$$

Where *p* is the allele frequency of i[th] locus, **M** is a matrix of SNP content centered by twice the current allele frequencies. Since the frequencies of the current genotyped population are used to center **G**, pedigree and genomic matrices have different bases, thus **G** was adjusted so the average diagonal and off-diagonal matched the averages of diagonal and off-diagonal in $\mathbf{A^{22}}$ as described in Vitezica et al. (2011).

Weighted Single Step Genomic Best Linear Unbiased Predictor

The WssGBLUP is the third method we employed (two models, each with a different CT value, as explained below). This approach is equal to model 2.2.2, except for matrix **G** which has been built following the second method of VanRaden (2008), as:

$$\mathbf{G_0} = \frac{\mathbf{MDM'}}{2\sum p_i\,(1 - p_i)}$$

Where *p* is the allele frequency of i[th] locus, **M** is a matrix of SNP content centered by twice the current allele frequencies, and **D** is the diagonal matrix in which SNP specific weights have been contained. The iterative algorithm reported in Zhang et al. (2016) has been used as weighting strategy. The SNPs weights presented in **D** were obtained as a function of the estimated SNP effect ($\widehat{\boldsymbol{u}}$). The weighting function used in this study was the one called

NonlinearA, as reported in Fragomeni et al. (2019). This method was preferred over other weighting strategies due to its stability among the iterations. The iterative algorithm applied followed the steps reported below:

1. Set the initial parameter $\mathrm{t}=1, \mathbf{D}_{(\mathrm{t})}=\mathbf{I}$, $\mathrm{G}_{(\mathrm{t})}=\frac{\mathbf{MD}_{(\mathrm{t})}\mathbf{M}'}{2\sum \mathrm{p_i}(1-\mathrm{p_i})}$, where **I** is an identity matrix;
2. Obtain GEBV ($\hat{\boldsymbol{a}}$) **,** where $\hat{\boldsymbol{a}}$ is the vector of solutions of additive genomic breeding value, using ssGBLUP algorithm;
3. Obtain SNP effect ($\hat{\boldsymbol{u}}$) as in (Gualdrón Duarte et al., 2014):

$$\hat{u}=\frac{1}{2\sum p(1-p)}\mathbf{DM}'[\mathbf{MDM}']^{-1}\hat{a}$$

4. Transform $d_{i(t+1)}$ as in Fragomeni 2019, in $\mathrm{CT}^{\frac{|\hat{\boldsymbol{u}}_i|}{sd(\hat{\boldsymbol{u}})}-2}$, where CT is a shrinkage factor determining how much the SNP effects distribution deviates from normality;
5. Standardize the weight of SNPs by maintaining a constant genetic variance among iteration:

$$\mathrm{D}_{(\mathrm{t}+1)}=\frac{\mathrm{tr}(\mathbf{D}_{(1)})}{\mathrm{tr}(\mathbf{D}_{(\mathrm{t}+1)})}\mathrm{tr}(\mathbf{D}_{(\mathrm{t}+1)});$$

6. Matrix **G** is then recreated including the new weights: $G_{(t+1)}=\frac{\mathbf{MD}_{(\mathrm{t}+1)}\mathbf{M}'}{2\sum \mathrm{p_i}(1-\mathrm{p_i})}$;
7. Set $t=t+1$, and go to point 2 for a new iteration.

We created two different WssGBLUP models, with two different CT values: WssGBLUP1 had CT value of 1.105 while WssGBLUP2 had CT value of 1.250. This was done to grant WssGBLUP1 the lowest possible shrinkage effect and WssGBLUP2 the highest possible shrinkage effect. For both models the maximum number of iterations was set to 5. For a matter of simplicity, we report only two WssGBLUP predictions instead of the ten analysed in this study (combination of two CT values and five iterations). Thus, we retained two opposite WssGBLUP scenarios: WssGBLUP1, which presents the lowest SNPs shrinkage effect, and WssGBLUP2, which presents the highest shrinkage effect

Single Step Linear Unbiased Predictor with only informative SNPs

The last group of models (five models) consisted in ssGBLUP in which the **G** matrix of 2.2.2 was constructed using SNPs obtained after the application of the different variable selection algorithms (described below, section 2.4). The number of columns of **Z** are thus different for each trait and each dataset.

Models' computations

In all models, **A** was built with the pedigree information tracking back up to 3 generations. In addition, according to what was reported in Cesarani et al. (2019), the variance components of each dataset were estimated under PBLUP models by tracing back of all animals in the pedigree. Variance components were estimated using the AIReml algorithm (Gilmour et al., 1995). All genetic and genomic prediction analyses were performed using the BLUPF90 family of programs (Aguilar et al., 2018b). The consistency of all this information is reported in Table 2. Preliminary analysis such as LD calculations was conducted using preGSf90 (Aguilar et al., 2018b, belonging to BLUPF90 family of programs).

Featured selection algorithms

The EBVs of the target trait were used to map the major SNP markers associated with the trait, using five different statistical approaches. The genome content was considered as a covariance matrix, while EBVs of genotyped animals $(\hat{a})$ (estimated using models in 2.2.2). were considered as the observed variable. The genome content was scaled in advance. Hyperparameters search and the choice of best models was performed by dividing the dataset in two parts: a training group and a test group. In the actual dataset young animals born after 2015 belong to the test group while older animals belong to the training group. In the simulation, animals of 8-9th generations were part of training group while animals of 10th generation belonged to test group

Least absolute shrinkage and selection operator (LASSO)

In high-dimensional information literature a large number of penalized likelihood approach was proposed. Given the baseline $y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + e_i$; a variant of penalized likelihood approach can be described as:

$$\hat{\beta} = argmax - \frac{1}{2}\left\|\sum_{i=1}^{N}\{y_i - (\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j)\right\|_2^2 + pen_\lambda(\beta)$$

Where: N is the number of animals for each trait, β_0 is model mean, $\beta_j$is the SNPs contribution p are the number of columns in x; N number of data and λ is the regularization parameters; and $pen_\lambda(\beta)$ is a penalty function. In LASSO (Tibshirani, 1996), penalty is:

$$pen_\lambda(\beta) = -\lambda\sum_{j=1}^{p}|\beta_j|$$

A grid search was performed to find the optimal values for the was obtained testing values from 0 to 20 in increments of 0.1. These values were used to maximize the LASSO model performance, based on the highest coefficient of determination and the lowest Mean Squared Error (MSE) in the training set. To do this we used glmet R package (Friedman et al., 2010)

Spike-and-Slab LASSO

The Spike-and-Slab LASSO (SSLASSO) was proposed by Ročková and George (2018). It was based on the idea that every penalized likelihood has a Bayesian interpretations (Bai et al., 2021). For instance, the LASSO penalization is equivalent to a Laplace distribution regulated by hyperparameter λ, where posterior mode of β are:

$$p(\beta|\lambda) = \prod_{j=1}^{p}\frac{\lambda}{2}e^{-\lambda|\beta_j|}$$

The SSLASSO is the equivalent to a two-point mixtures of Laplace distributions defined as:

$$p(\beta|\lambda) = \prod_{j=1}^{p}\left[(1-\gamma_j)\left(\frac{\lambda}{2}e^{-\lambda_0|\beta_j|}\right) + \gamma_j\left(\frac{\lambda}{2}e^{-\lambda_1|\beta_j|}\right)\right]$$

Where:

$$p(\gamma|\theta) = \prod_{j=1}^{p}[\theta^{\gamma_j}(1-\theta)^{1-\gamma_j}] \text{ and } p(\theta) \sim Beta[a,b]$$

The Bayesian prior can be re-arranged in an penalized likelihood context by took his marginal logarithm prior (Bai et al., 2021); after some derivation is possible to obtain:

$$\lambda_\theta(\beta_j) = \lambda_1 p_\theta(\beta_j) + \lambda_0[1 - p_\theta(\beta_j)]$$

Where:

$$p_\theta(\beta_j) = \frac{1}{1 + \frac{(1-\hat{\theta})}{\hat{\theta}}\frac{\lambda_0}{\lambda_1}\exp[-|\beta_j|(\lambda_0 - \lambda_1)]}$$

SSLASSO was computed using SSLASSO R packages (Ročková and George, 2018), error variances was assumed to be unknown and self-adaptivity penalty was set. This means that θ was assumed to be random and applied different shrink to each β_j.

Recursive Feature Elimination using Ridge Regression (RfeRR)

Similar to LASSO, the Ridge Regression is based on a principle of penalized likelihood, with penalty equal to $\lambda\sum_{j=1}^{p}{\beta_j}^2$. Before we proceeded with recursive feature elimination, the optimal values of λ were obtained as in LASSO section. glmet R package was used (Friedman et al., 2010). After that recursive feature elimination using penalized Ridge Regression was performed as follow. In each of iteration, SNP effect $\beta_j$were estimated on training data. Then 10% of variable with lowest was removed form next iterations. Variable (SNP) present in the iteration with lowest $|\beta|$ mean squared error (MSE) was considered for following prediction. MSE was calculate as where y_{test} is the EBV belong to test database and $y_{test}$is the predicted ones.

Recursive Feature Elimination using Support Vector Machine (RfeSVM)

The SVM is a kernel-based supervised learning technique, often used for regression analysis SVM can map linear or nonlinear relationships between phenotypes and SNP markers depending on the kernel function considered. The best kernel function to map genotype to phenotype was determined in different training subsets: a 5-folds split was used to determine the kernel function, which adjusted better to the data, either linear, polynomial, or

radial basis. We found that performing the SVM with a linear basis function outperformed the polynomial and radial basis function of about 12.5% in predictive ability.

The general model for SVM (Evgeniou and Pontil, 2005; Hastie et al., 2009) can be described as:

$$\boldsymbol{y_i^* = b + h(m) * w + e}$$

where $h(m)$ represents the linear kernel basis function ($h(m) = m'm$) used to transform the original predictor variables (i.e., SNP markers information (m)), b denotes the model bias, and w represents the unknown weight vector. In the SVM model, the learn function $h(m)$ was given by minimizing the loss function as follows: $C\sum_{i=1}^{N} L\left(y_i^* - \hat{y}_i^*\right) + \frac{1}{2}\|w\|^2$. The C represents a regularization parameter which controls the trade-off between predictor error and model complexity, and $\|.\|^2$ denotes the squared norm under a Hilbert space. The SVM model was fitted using a epsilon-support vector regression that ignores residuals absolute value ($|y_i^* - \hat{y}_i^*|$) smaller than some constant (ε) and penalize larger residuals (Vapnik, 2000). The parameters C and ϵ were defined using the training data set as proposed by Cherkasky & Ma (Cherkassky and Ma, 2004): $C = \max\left(\left|\overline{y^*} + 3\sigma_{y^*}\right|, \left|\overline{y^*} - 3\sigma_{y^*}\right|\right)$ and $\epsilon = 3\sigma_{y^*}\left(\sqrt{\ln(n)/n}\right)$, in which the $\overline{y^*}$ and σ_{y^*} are the mean and the standard deviation of the target EBV for the traits on the training population and *n* represents the number of animals in the training set. The SVM was performed with the e1071 R package (Meyer et al., 2020).

After that, recursive feature elimination using SVM was performed using same procedure described for RfeRR in Sanz et al. (2018).

Boosting ensemble

The Boosting approach (xGBoost) is an ensemble technique that combines gradient descent error minimization with boosting, aiming to convert weak regression tree models into strong learners (Hastie et al., 2009; Natekin and Knoll, 2013). This ensemble process combines different predictor variables sequentially in the regression tree model, using regularization via selection and shrinkage of the predictors to control the residual from the previous model (Friedman, 2002). The xGBoost can employ parallel computation to use more

regularized models to control the model over-fitting. The xGBoost approach can be described as follows:

$$\boldsymbol{y} = \sum_{\boldsymbol{w=1}}^{\boldsymbol{W}} \boldsymbol{\beta_w h(x, \gamma_w)} + \boldsymbol{e}$$

where y is the vector of the target EBV, W is the numbers of iterations (expansion coefficients), β_w is the shrinkage factor, also known as the "boost", and $h(x, \gamma_w)$ is the base learner, a function of the multivariate argument x with a set of parameters $\gamma_w = \{\gamma_1, \gamma_2, \ldots, \gamma_w\}$, and e is the vector of the residuals. Expansions of the coefficients $\{\beta_w\}_1^W$ and parameters $\{\gamma_w\}_1^W$ are used to map the predictor variables (x), i.e., SNP markers, to the target EBV (y) considering the joint distribution of all values (y, x) minimizing the loss function $L\{y_i, F(x)\}$ given $[y, F_{w-1}(x_i) + h(y_i; x_i, p_w)]$, where p_w is the predictor to minimize $\sum_{i=1}^{n} L\,[y, F_{m-1}(x_i) + h(y_i; x_i, p_m)]$. The xGBoost follows the algorithm specified by Chen and Guestri (2016). In the xGBoost method, a regularization term is added in the loss function, representing the weight vectors learned in the loss function and penalizes ponderation of large weights. This regularization term is represented as follows: $\sum_{i=1}^{n} L\,[y, F_{m-1}(x_i) + h(y_i; x_i, p_m)] + \sum_n \Omega(f_n)$, where the L is the error between the true value of the target trait and the predicted value, and $\Omega(f_n)$ is the regularization function used to prevent overfitting: $\Omega(f_n) = \gamma T + 0.5\lambda\|\omega\|^2$, where T is the number of leaves in the regression tree f_n and ω represents the weight for the leave in each tree (i.e. the predicted values stored at the leaf nodes). Including $\Omega(f_n)$ in the objective function makes the tree less complex, which minimizes the loss function and helps reduce overfitting; γT is a constant penalty for each additional tree leaf and $\lambda\|\omega\|^2$ penalizes extreme weights. The γ and λ are the regularization terms L1 and L2 (Mitchell and Frank, 2017). The random search for xGBoost was performed considering the four most important parameters able to increase prediction accuracy and minimize the prediction error. These hyperparameters were Ntree (total number of trees in the sequence used in the model), learning rate (determines the contribution of each tree to the final model and performs shrinkage to avoid variable overfitting), maximum tree depth (controls the depth of the individual trees to be considered in the model), and minimum samples per leaf (controls the complexity of each tree). The Ntree values ranged from 600 to 5,000 in intervals of 200; the learning rate was in the range of 0.05 to 1 in intervals of 0.05; maximum tree depth was determined with a value ranging from 5 to 80 in intervals of 5; minimum samples per leaf was

determined from 5 to 100 in intervals of 5 and considering lambda and alpha regularization values ranging from 0 to 1 in intervals of 0.05. The random grid search xGBoost was performed using the *h2o.grid* function of the *h2o* R package (https://cran.r-project.org/web/packages/h2o), considering as fixed parameters a maximum of 150 models with random combinations of the hyperparameters over 60 min.

1.5 Effective population size calculations

In order to have a proper comparison between actual and simulated data, the effective population size (N_e) has been computed from the individual increase in inbreeding (ΔF) (Falconer and Mackay, 1996). Individual ΔF has been computed as:

$$\Delta F = \frac{F_n - F_{n-1}}{1 - F_{n-1}}$$

$$Ne = \frac{1}{2\Delta F}$$

where F_n is the inbreeding in the n^{th} generation. Ne was calculated using purge R package (https://cran.r-project.org/web/packages/purgeR).

Validation

Simulated dataset

Quality of prediction was measured as the correlation and MSE between the genomic breeding values estimated under different models and the true breeding values for animals belonging to the 10th generation, that is the last generation of animals, including individuals without phenotypes but with genotype.

Actual dataset

In the actual dataset, two different cross-validation methods were applied. The first method that we used to cross-validate predictive ability was to calculate both the correlation and the MSE between predicted and observed phenotype. In this case five-fold cross validation with 10 iterations were performed. Since not all animals were genotyped in each

iteration 1/5 of non-genotype and 1/5 genotype animal were masked. In the current manuscript we reported predicted ability metrics only for genotyped animals; result about non-genotype animals was reported on Supplementary materials Figure S1.

Linear regression (LR) (Legarra and Reverter, 2018) was used as the second cross-validation method. It compares the prediction performances of different models on groups of focal individuals born after a given date, in this case the young bulls. LR is particularly suited to the specific needs of Rendena population since predicting the future performance of young bulls without phenotype is one of the main objectives of the breeding plans for performance tests (Mancin 2021b).

The LR method evaluates the goodness of a model by comparing its performance in a complete dataset and a partial dataset. With complete dataset we refer to the dataset containing the whole amount of information, or the dataset used for prediction. Partial dataset is referred to the complete dataset with some animals with phenotype removed, usually young animals known as candidates to selection. According to Macedo et al. (2020) we built partial datasets by excluding phenotypes since a target recent birth year of young bulls (since 2012 to 2020; since 2014 to 2020; … since 2017 to 2020) to describe possible variations and random deviations of the estimator, consistencies are reported on Table 3. LR considered three parameters: bias, dispersion, and accuracy. Bias is the difference between the expected breeding values estimated under the complete vs. the partial datasets. The dispersion was estimated as the regression coefficient considering the breeding values from the complete dataset on the ones estimated from partial data and the accuracy as correlations between the two breeding values.

Table 3: Description of the different validation set used in cross-validation, first and last years of born and number of animals used in the validation cohort are reported.

Since	Last	Number
2012	2020	178
2013	2020	154
2014	2020	130
2015	2020	109
2016	2020	106
2017	2020	72
2018	2020	45

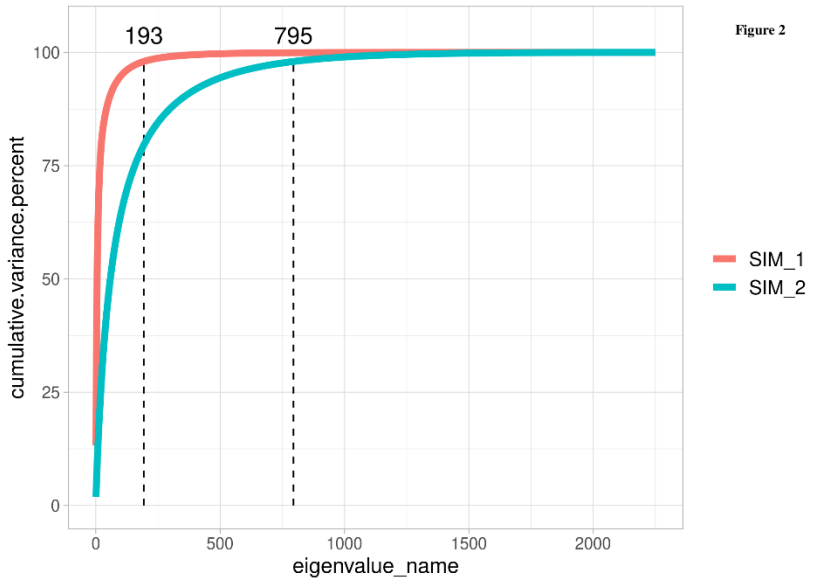
RESULTS

Genomic structure

Simulated datasets

Figure 1 highlights the different genomic assets of small Ne populations (SIM1 and SIM3; 10 sires per generation) and large Ne populations (SIM2 and SIM4; 200 sires per generation). Since the different number of QTL assumed for the populations with the same Ne (that is, 10 vs. 1000 QTL) did not have an impact on **G** matrix dimensionality, only SIM1 and SIM2 were plotted for a matter of simplicity. In SIM1, 193 eigenvalues were necessary to explain 98% of **G** matrix variance, while in SIM2 795 eigenvalues were necessary to explain 98% of **G** matrix variance. When only ten sires per generation were used, it was possible to observe different sub-populations (Figure 1A); however, no population structure was found when plotting the first two eigenvalues (Supplementary Material, Figure S2. On the other hand, SIM2 appeared homogenous, and individuals appeared almost unrelated to each other. In addition, when LD per chromosome was also calculated, a greater value was observed in SIM1 (0.161 ± 0.076) than in SIM2 (0.067 ± 0.054; data not shown). A Ne value of respectively 81.18 ± 4 and 1869± 546 was also determined for SIM1 and SIM2.

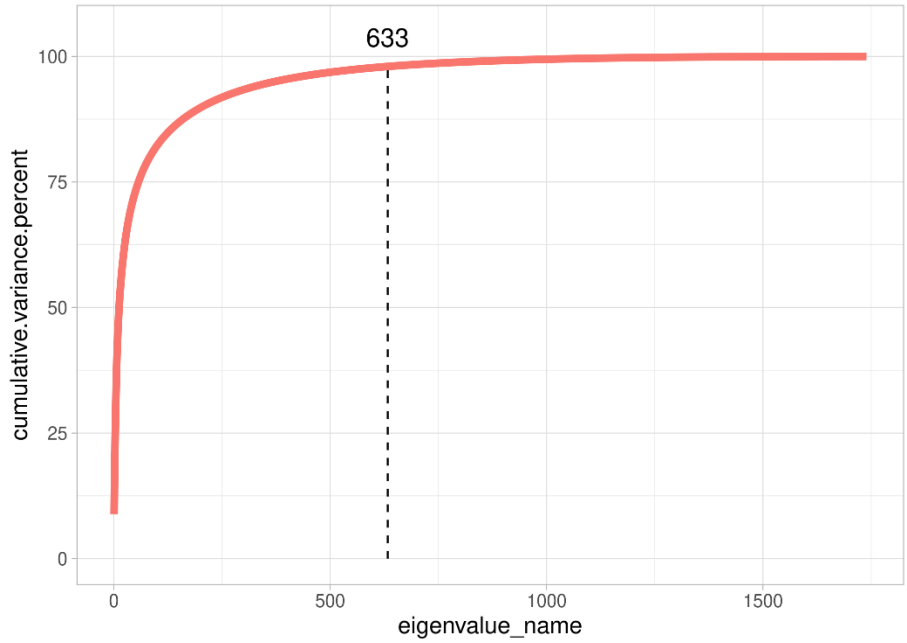
Figure 2 Cumulative explained variance of all eigenvalues of genomic relationship matrix of two simulated populations



Actual dataset

We also investigated **G**'s dimensionality on the actual dataset of Rendena cattle population (Figure 2). The actual dataset presented a situation closer to SIM2 than to SIM1. In fact, it presented an average Ne value of 108.2 ± 0.74 calculated from pedigree data. It is possible to observe a few clusters in the genomic relationship matrix (Figure 2), however, they are not as clear as in SIM1; we therefore can observe that no population structure is present in Rendena breed, which is in line with previous research (Mancin et al., 2022). The 98% of **G** variance was explained by only 633 eigenvalues, thus the scenario was closer to SIM1 than SIM2. In addition, we observed for LD an average value of 0.187 ± 0.107 per chromosome (Mancin et al., 2022).

Figure 3 Cumulative explained variance of all eigenvalues of genomic relationship matrix of Rendena populations



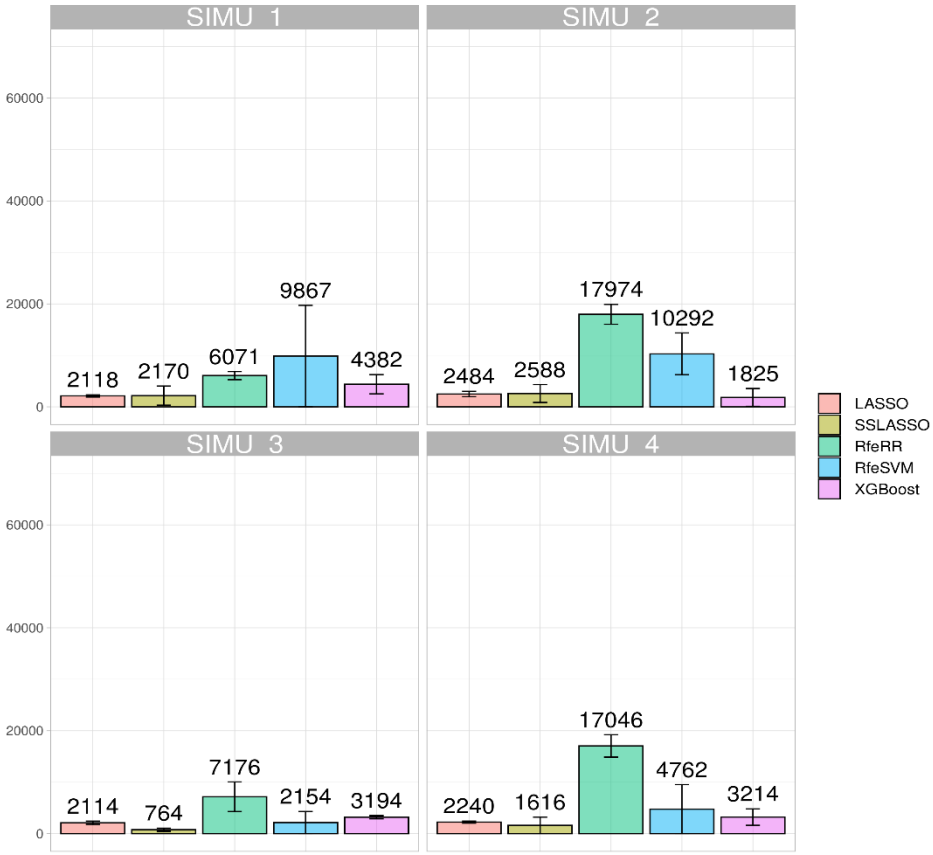
SNPs retained by variable selection models

SNPs retained in simulated datasets

The impact of the different algorithms was appraised in terms of the number of informative markers retained, as reported in Figure 3. Specifically, we were interested in identifying the impact that different G matrixes' dimensionality and QTL had on the number of SNPs considered informative. In all simulations, LASSO and SSLASSO retained the lowest

number of SNPs (roughly 2,000 SNPs averaged across simulations) and they presented lower intra and between scenarios variability. On the contrary, RfeSVM and RfeRR algorithms retained higher numbers of SNPs on average 12,000 for RfeRR and 7,000 for RfeSVM . RfeSVM presented also an extreme variability across scenarios (Figure 3). XGboost retained an intermediate number of SNPs, with an average of 3,000 SNPs retained across simulations. As we show in Figure 3, different numbers of QTL did not affect the number of SNPs retained by each algorithm. In fact, no difference was observed between respectively SIM1 and SIM3, and SIM2 and SIM4; only LASSO and SSLASSO algorithms seem to be slightly affected by number of QTL. Interestingly, dimensionality of G matrix seems to be more influential, as scenarios with higher Ne presented a higher number of SNPs (SIM1-SIM2). The XGboost is the only algorithm where this trend has not been seen. In addition it was also interesting to observe that the negative gap in models accuracy present in simulations with lower QTL (SIM3-4) fades when variable selection models is introduced.

Figure 3 Bar plot representing representing number of SNPs retained by each algorithms on the four simulated population, error bar rappresent the standard deviation



SNPs retained in real Dataset

We show the impact of variable selection methods in terms of the number of informative markers retained in the Rendena population in Figure 4. Although the number of initial SNPs was similar to the simulated populations, in general in the actual dataset a higher number of SNPs were retained by the algorithms. Similarly, to what was reported in the simulated data, LASSO and SSLASSO were the most restrictive algorithms of SNP selection with an average of 2,000 SNPs retained across the simulations. XGboost was the second most restrictive algorithm in terms of SNPs retained by the models about 3,000 on average. RfeSVM and RfeRR algorithms retained about on half of the SNPs presented in the panels. No clear patterns were identified across different phenotypes: some algorithms found greater number of SNPs in certain traits and some in others. For example, the lowest number of informative markers retained by RFE algorithms was identified on DP trait, but the opposite situation occurred for XGboost where the algorithm presented almost twice the number of informative SNPs for DP.

Figure 5 Bar plot representing number of SNPs retained by each algorithm on the three phenotype of the Rendena population

Breeding values prediction

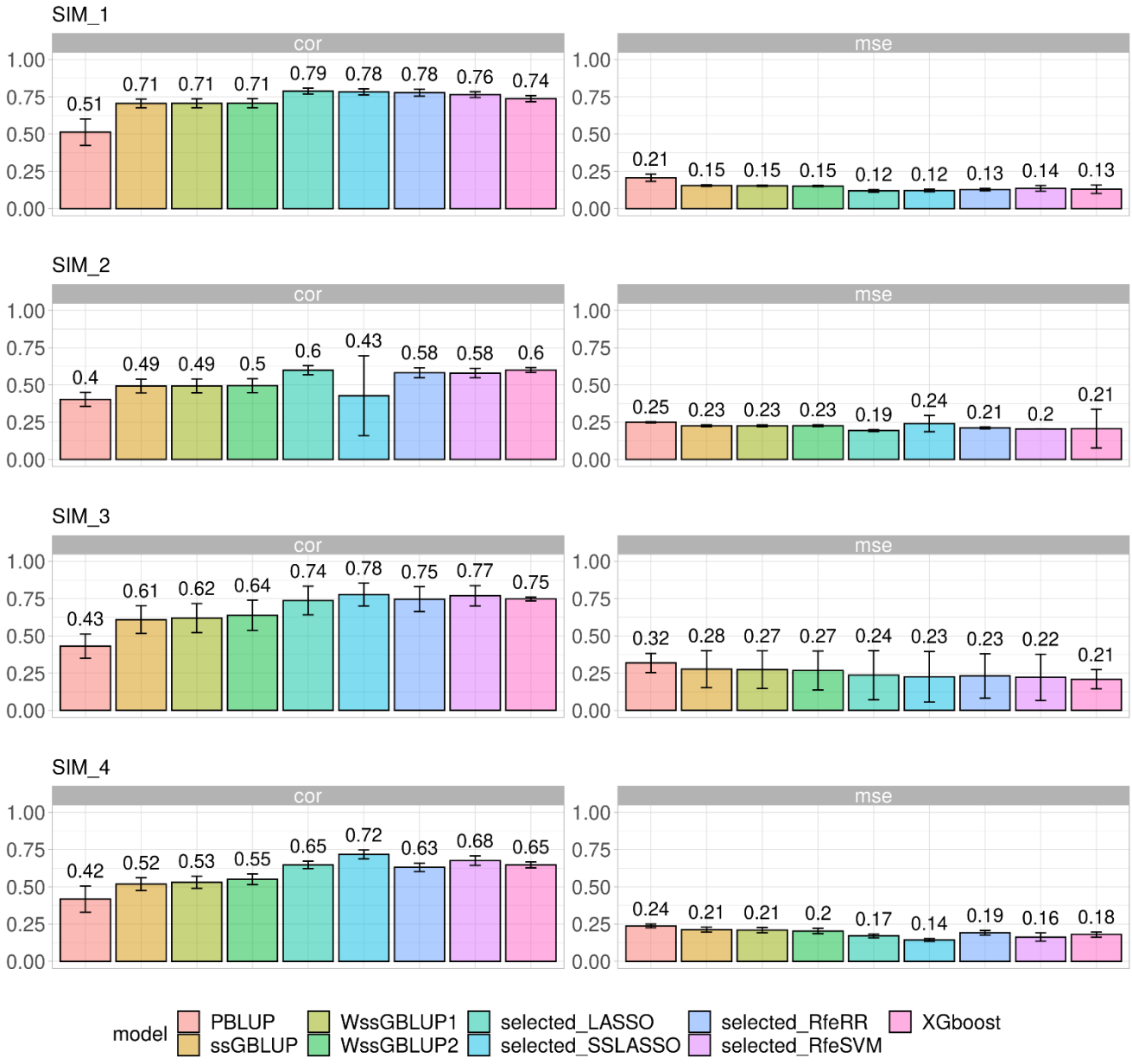
We compared the prediction accuracy of four 'classical' models for BLUP and ssGBLUP with five different SNP pre-selection strategies. The models were 1) PBLUP; 2) single ssGBLUP; 3) WssGBLUP1; 4) WssGBLUP2; 5a) ssGBLUP with SNPs preselected via LASSO; 5b) ssGBLUP with SNPs preselected via SSLASSO; 5c) ssGBLUP with SNPs preselected via RfeRR; 5d) ssGBLUP with SNPs preselected via RfeSVM; 5e) and ssGBLUP with SNPs preselected via XGboost.

Breeding values prediction in simulated datasets

Results of different prediction models' accuracy are reported in figure 5, with correlation and MSE as metrics of comparison. MSE values were comparable to those obtained for correlations. Standard BLUP models achieved lowest accuracy. A substantial increase in accuracy was observed in ssGBLUP models (Figure 5), i.e., when genomic data were integrated: this increase of accuracy was more relevant for populations with small Ne (SIM1, SIM3).

A slight increase of accuracy with respect to ssGBLUP was observed when a heterogeneous distribution of SNPs was considered within the matrix G (WssGBLUP). The gap in accuracy was greater in the populations with few QTL (SIM3, SIM4), especially for WssGBLUP2: on the other hand, the increase in accuracy for SIM1-2 under WssGBLUP was almost close to zero. A substantial variation in accuracy values was observed when ssGBLUP was performed with **G** matrixes constructed with selected SNPs; however, the accuracy of the prediction performance of each variable selection model changed according to the simulation structure. Generally, SSLASSO presented the highest increase in accuracy among the genetic models in all simulations, with the exception of SIM2 where we observed a dramatic drop of accuracy. LASSO on the other hand, achieved greater accuracy on both SIM1 and 2. Other algorithms presented an intermediate increase in accuracy among the genetic models in all simulations, namely (RfeRR, RfeSVM and XGboost), with different ranking the different scenarios.

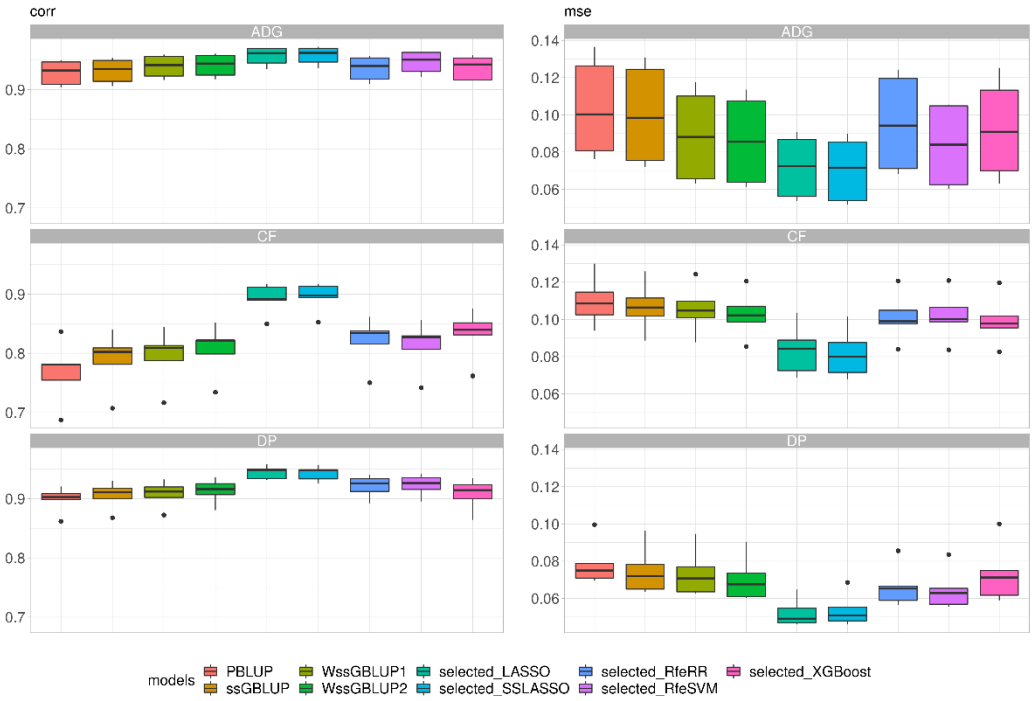
Figure 6: Bar plot representing correlation (corr)and mean squared error (mse) between predicted and true breeding values on the four different simulation, error bar represent the standard deviation.



Breeding values prediction in actual dataset

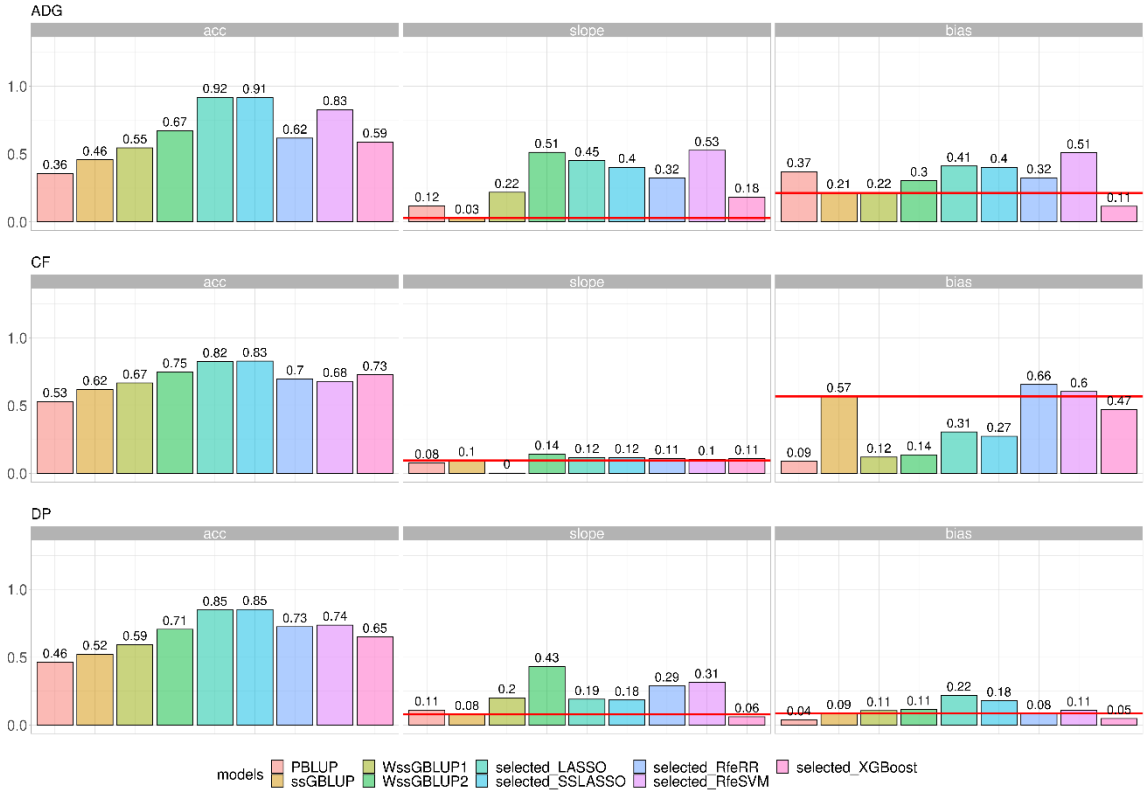
With our real datasets we were interested at first in evaluating the performance of these models in terms of prediction; and then we wanted to evaluate the feasibility of introducing them in real breeding plans scenario. This point was achieved by using LR cross validation methods (Legarra and Reverter, 2018). Figure 6 represents the results of repeated five folds cross validation. The integrations of genomic data led again to a substantial increase in accuracy: the PBLUP presented the overall lowest correlation values (r from 0.36 to 0.53). The ssGBLUP presented the lowest correlation values among genomic models (r from 0.46 to 0.62), while a slight increment observed for WssGBLUP1 (from 0.55 to 0.67) and in WssGBLUP2 (from 0.67 to 0.75). As with simulated data, variables selection models improved models' accuracy substantially. Again, highest correlation was found for LASSO and SSLASSO, with values of r ranging from 0.83 to 0.92, while other algorithms presented intermediate values (r around 0.70). This pattern was observed across all traits. MSE reflected the results obtained with correlations.

Figure 7 Box plot representing correlation (corr)and mean squared error (mse) between predicted and true breeding values phenotype of Rendena performance test. Target phenotype are ADG: Average Daily Gain; CF: in vivo Carcass Fleshiness; DP: in vivo Dressing Percentage



LR methods considered in addition to accuracy, also dispersion and bias. Figure 7 represents the different results obtained through LR cross-validation methods in the various validation sets of 2015-2020. This set of years was chosen as representative of all seven validation cohorts. Figure 8 reports the summary statistic of all seven validation cohorts.

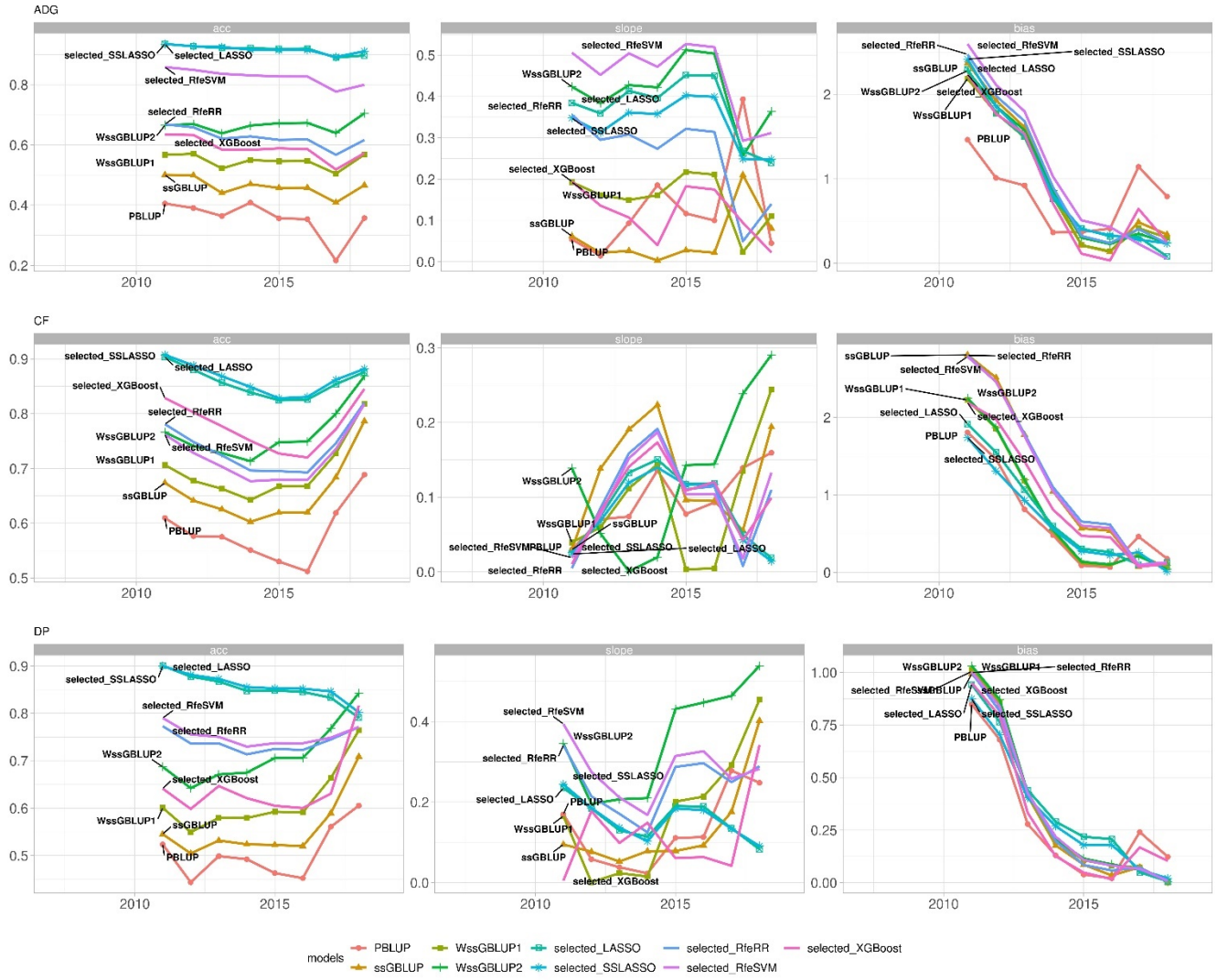
Figure 7 Bar plot representing accuracy, dispersion and bias of Rendena Dataset estimated using LR cross validation, in the validation cohort of 2015-2020. Dispersion was represented as 1-absolute values of dispersion while bias as absolute values of bias. This allowed a better models ranking while horizontal lines represented the values of ssGBLUP this allowed better comparison among models, only genotyped animals were considered in the validation.



Accuracy trends of the actual dataset measured with LR method were similar to accuracy obtained with five-fold cross validation. However, looking at the other statistics (slope and bias) we can observe that LASSO, SSLASSO RfeRR and RfeSVM cannot be considered suitable variable selection approaches in real breeding plans, due to their higher bias and dispersion values, especially if compared with ssGBLUP. XGboost was the only model that presented similar or even lower bias and dispersion values than ssGBLUP but

greater accuracy. As seen in Figure 8, we demonstrate that these trend is consistent over the different validation cohort.

Figure 9: line plot representing accuracy, dispersion and bias of Rendena Dataset estimated using LR cross validation, in all validation cohort. Dispersion was represented as 1-absolute values of dispersion while bias as absolute values of bias.



DISCUSSION

The present study had two objectives: testing if reducing the number of SNPs used to construct **G** could lead to an increase in the accuracy of (ss)GBLUP, and whether this method could be introduced in genomic evaluations of the Rendena breed.

In our study, using both simulated and actual datasets, we demonstrated that the accuracy of (ss)GBLUP increases when it is performed with SNPs selected via variable

selection methods. This is in agreement with the extensive literature supporting the increased accuracy of Bayesian variable selection models in many different species (Lourenco et al., 2014; Mehrban et al., 2017; Yoshida et al., 2018; Zhu et al., 2021). However, few studies until now had investigated the impact of SNP preselection on **G** matrix using *ad hoc* algorithms. Akbarzadeh et al. (2021) integrated in a GBLUP framework only a subset of chosen SNPs based on classical GWAS analysis (i.e., 1%, 5%, 10%, 50% of significant SNPs). A slight increase in accuracy with respect to the canonical GBLUP was observed when **G** was constructed using only the best 10% and 50% SNPs; contrariwise, models using the 1% and 5% of the SNPs prediction underperformed. Furthermore, Akbarzadeh et al. (2021) reported a dramatic decline in performance when the same percentage of SNPs was randomly chosen. Preliminary tests of a similar approach – construction of the **G** matrix using the top 500, 1000, 50000 SNPs ranked by their absolute SNP effect values calculated through back solutions – have been tried in Rendena. However, we immediately discarded this approach because of the extreme bias and inflated breeding values predictions (these findings are reported in Mancin et al 2022 in press). In addition, choosing so few and unrepresentative SNPs greatly reduced the compatibility between the two matrices, and thus ssGBLUP properties were affected (Misztal et al., 2017).

Li et al. (2018) and then Piles et al. (2021) showed how the use of different methods to select the most informative SNPs could significantly improve the performance of the variable selection models. Li et al. (2018) constructed the **G** matrix by using the best 400, 1,000, and 3,000 SNPs, ranking SNPs effects by three different machine learning models. As in the previous case, an increase in accuracy was obtained only with a certain number of selected SNPs (1,000 SNPs), while a decrease in accuracy with respect to canonical GBLUP was observed with a lower number of SNPs. Additionally, Piles et al. (2021) and Azodi et al. (2019) showed how by combining different variable selection algorithms with several different parametric and non-parametric prediction models (i.e., ensemble predictions), it is possible to obtain a consistent increase in accuracy compared to models without variable selection. However, our study has not explored these scenarios since prediction methods other than ssGBLUP or ssSNP-BLUP (Fernando et al., 2014) do not seem to bring any concrete improvement for livestock traits (Abdollahi-Arpanahi et al., 2020). Furthermore, ssGBLUP and ssSNP-BLUP are the only methods that allow combining straightforwardly non-genotyped animals with genotyped ones – a crucial feature for a real-life routine selection plan and something that the other algorithms cannot do.

Our result that reducing the number of parameters has a positive impact on accuracy is also supported by Frouin et al. (2020). In that study, it was demonstrated that the error of the prediction tends to increase linearly when n > p until to the "irreducible" error $(1 - h^2)$ that occur when n $\gg$ p. In addition, Pocrnic et al. (2019), demonstrated that accuracy of (ss)GBLUP is connected by the distribution of eigenvalues of **G**, thus "n" becomes the number of Me captured by SNPs (Pocrnic et al., 2016). In highly related populations (small Ne) higher values of accuracy can be achieved than in populations with larger Ne, because fewer eigenvalues and thus "n" are necessary to explain **G**: in large Ne populations more data are needed to increase accuracy. This is also intuitive since prediction error accuracy (Henderson, 1988) is directly proportional to C^{aa}, thus in highly related populations tends C^{aa} to have lower values. C^{aa} is the inversion of coefficient matrix of the mixed model equation where aa is the block referring to the genetic effect of animals. What was reported on Pocrnic et al. (2019) could explain the lower performance identified in Akbarzadeh et al. (2021) when 1% and 5% were considered (Akbarzadeh et al., 2021). Indeed, discarding too many SNPs from the construction of **G** may omit the inclusion of important eigenvalues. From another perspective, Fragomeni et al. (2017) demonstrated the positive impact of removing non informative SNPs on GBLUP. The authors demonstrated in a simulated dataset that better accuracy was found when the **G** was built by removing all SNPs outside the window where the QTL was situated or using only QTL information. However, a practical limit to this method is that knowing all the QTL within a genome is nearly impossible, especially when the population is small (Mancin et al., 2021a).

Our simulated results support the abovementioned theory, as simulations with lower Ne presented higher accuracy of ssGBLUP (SIM1, SIM3). Furthermore, differences between scenarios emerge when comparing simulations differing for their number of QTL. ssGBLUP showed lower performance in the SIM3-4 (QTL10) than in the SIM1-2 (QTL1000); however, this discrepancy in accuracy decreases when variable selection is applied. This is in agreement with what is reported in Daetwyler et al. (2010) that demonstrated that selection of SNPs via BayesB presents concrete advantages when number of QTL is small compared to the number of independent chromosome segments (Me).

As mentioned above Bayesian SNPs regression, or (ss)GBLUP using a weighted realized relationship matrix (Tiezzi and Maltecca, 2015; Zhang et al., 2016), always improve prediction accuracy with respect to models that assume homogenous variance among SNPs

(GBLUP or SNP-BLUP). However this increase of accuracy is often connected with increases of bias especially when time-cross validation is used (Mehrban et al., 2017) as opposed to five folds or leave-and-out cross validation (Zhu et al., 2021). However, when the goal is to achieve the “best predictor”, namely a value close as possible to real one, models assuming heterogeneous variances and models with variable selection can be identified as the best models, as they have highest MSE, intended as bias-variance trade off (Gianola et al., 2018). In this regard, LASSO and SSLASSO thus appeared as “best models”, for both simulated and real data. We showed that (SS)LASSO regression performs automatic feature selection especially in the presence of features that are linearly correlated, such as SIM1 and SIM3, since their simultaneous presence will increase the value of the cost function. Thus, Lasso regression will try to shrink the coefficient of the less important SNPs to 0, in order to select the best features.

However, in real-life breeding scenarios time-cross validation must be taken into account (Liu, 2010; Legarra, A. Reverter, 2017) as this procedure simulates the natural accumulation of information across time. Only few studies evaluated the impact of heterogeneous or variable selection models using time cross-validation with small samples of individuals. Cesarani et al. (2021) and Mancin et al. (2021b) found higher bias and overdispersion values in WssGBLUP with respect to ssGBLUP.

When we performed LR cross validation methods the same pattern emerged (Cesarani et al., 2021; Mancin et al., 2021b), namely that higher shrinkages or selected SNPs have high accuracy but carried higher bias and dispersion values. Specifically, (SS)LASSO models were identified as the models with best accuracy in all three traits when measured with LR. Other feature selection models and WssGBLUP presented lower accuracy. Among the variable selection models we found slightly lower values of accuracy in the XGboost; however, we suggest that XGboost could be regarded as the best variable selection model among those tested, as it is the only model that presented higher accuracy than ssGBLUP, at net of better bias and dispersion.

Several questions still persist about the use of these models in routine evaluation. One of these concerns the implementation of pre-selected SNPs in multitraits models. However, this is a recurring problem not only when the **G** matrix is built with pre-selected SNPs, but more in general whenever models take into account the specific genomic architecture of traits,

as for example WssGBLUP does. A possible solution to bypass this issue mighet be using multiple **G** matrix prediction models, one for each trait: yet, this is not computationally straightforward. A more concrete approach for future studies could represented by a preliminary selection of SNPs by multi-objective optimization framework algorithms as in Garcia (2019). Another possible concern about a large-scale use of variable selection ssGBLUP is the fluctuations of SNPs across generations. Similarly to the issue with multitrait models, this regards all genomic selections (Hidalgo et al., 2020): however, it is true that with respect to other methods, such as Bayesian SNPs regression, generation-by-generation recalibration of SNPs preselection algorithms can be extremely computationally demanding, especially when algorithms such as XGboost are chosen. Finally, SNP preselection could be influence by variability in SNPs frequency across animals, or more in general in the presence of population structure. In our study nonetheless, the PCA plots referring to SIM1 (Supplementary Materials S2), where some clusters are present, show that variable selection models overcome this issue quite handily. In future studies it would be interesting to choose one or more variable selection models and evaluate their impact on more stratified populations.

Besides increasing the EBV's accuracies, developing an optimal strategy for SNPs variable selection in high-density panels will be particularly useful in local breeds. It would in fact allow the use of informative but lower density and cheaper panels. Furthermore, given that small breeds cannot attract the same level of technological investment as their cosmopolitan counterparts (e.g., Holstein), decreasing the costs of genomic selection could be critical to help guarantee their selection, and thus their survival.

Aside from the economic factors, the importance of developing *ad hoc* selection methods for small-population cattle, especially for local breeds, is of primary importance for their conservation. Maintaining genetic progress for the productive characters and at the same time keeping intact the genetic variability and the distinct characteristics of the breeds can be guaranteed through breeding plans implementing careful selection (Biscarini et al., 2015). These plans are needed to preserve genetic variability within livestock local populations, a goal which, in the medium term, is critical for the animal husbandry industry to ensure the conservation of native breeds, their productive and reproductive efficiency, health, survival, and overall resilience to future changing environmental pressures (Mastrangelo et al., 2014).

CONCLUSIONS

Genomic information, especially the single-step GBLUP technique, has brought great improvements to selection and breeding decisions in livestock. However, these methods still present methodological issues when applied to populations with a small size, such as local and endemic cattle breeds. Our rigorous testing of different algorithms for variable selection of informative SNPs has highlighted that prediction accuracy of variable selection ssGBLUP (especially that of XGboost) was greater than that of other ssGBLUP methods, without the inflated bias and dispersion that accompany the Weighted ssGBLUP. Our use of machine learning models could thus represent a solution to the issue of genomic selection in small populations. Local cattle breeds are an often-untapped resource of genetic diversity and have great potential to adapt to varying environmental conditions; the methods presented here might thus be employed in their conservation, study, and increase their economic competitiveness.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

Idea E.M.; Conceptualization, E.M., C.S., B.T., L.M. and E.M.; methodology, E.M. L.M; formal analysis, E.M. and L.M.; support to analysis B.T., investigation, B.T., C.S., R.M., and E.M.; resources, R.M.; data curation, E.M., R.M.; writing original draft preparation, E.M. and B.T. writing—review and editing L.M., C.S., and R.M. All authors have read and agreed to the published version of the manuscript.

Funding

The study was funded by the DUALBREEDING project (CUP J61J18000030005) and by BIRD183281.

ACKNOWLEDGMENTS

Authors are grateful to National Breeders Association of Rendena cattle breed (ANARE) for data support.

REFERENCES

Abdollahi-Arpanahi, R., Gianola, D., and Peñagaricano, F. (2020). Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genet. Sel. Evol.* 52, 1–15. doi:10.1186/s12711-020-00531-z.

Aguilar, I, Tsuruta, S., Masuda, Y., Lourenco, D. A. L., Legarra, A., and Misztal, I. (2018). BLUPF90 suite of programs for animal breeding. in *The 11th World Congress of Genetics Applied to Livestock Production* (Auckland, New Zealand), 11.751.

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93, 743–752. doi:10.3168/jds.2009-2730.

Akbarzadeh, M., Dehkordi, S. R., Roudbar, M. A., Sargolzaei, M., Guity, K., Sedaghati-khayat, B., et al. (2021). GWAS findings improved genomic prediction accuracy of lipid profile traits: Tehran Cardiometabolic Genetic Study. *Sci. Rep.* 11, 1–9. doi:10.1038/s41598-021-85203-8.

Alvarenga, A. B., Veroneze, R., Oliveira, H. R., Marques, D. B. D., Lopes, P. S., Silva, F. F., et al. (2020). Comparing Alternative Single-Step GBLUP Approaches and Training Population Designs for Genomic Evaluation of Crossbred Animals. *Front. Genet.* 11, 263. doi:10.3389/fgene.2020.00263.

Azodi, C. B., Bolger, E., McCarren, A., Roantree, M., de los Campos, G., and Shiu, S. H. (2019). Benchmarking parametric and machine learning models for genomic prediction of complex traits. *G3 Genes, Genomes, Genet.* 9, 3691–3702. doi:10.1534/g3.119.400498.

Bai, R., Ročková, V., and George, E. I. (2021). Spike-and-Slab Meets LASSO: A Review of the Spike-and-Slab LASSO. *Handb. Bayesian Var. Sel.*, 81–108. doi:10.1201/9781003089018-4.

Biscarini, F., Nicolazzi, E. L., Stella, A., Boettcher, P. J., and Gandini, G. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. *Front. Genet.* 6, 33. doi:10.3389/fgene.2015.00033.

Blasco, A., and Toro, M. A. (2014). A short critical history of the application of genomics to animal breeding. *Livest. Sci.* 166, 4–9. doi:10.1016/j.livsci.2014.03.015.

Botelho, M. E., Lopes, M. S., Mathur, P. K., Knol, E. F., Guimarães, S. E. F., Marques, D. B. D., et al. (2021). Applying an association weight matrix in weighted genomic prediction of boar taint compounds. *J. Anim. Breed. Genet. = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie* 138, 442–453. doi:10.1111/jbg.12528.

Calus, M. P. L., and Vandenplas, J. (2018). SNPrune: An efficient algorithm to prune large SNP array and sequence datasets based on high linkage disequilibrium. *Genet. Sel. Evol.* 50, 1–11. doi:10.1186/s12711-018-0404-z.

Cesarani, A., Biffani, S., Garcia, A., Lourenco, D., Bertolini, G., Neglia, G., et al. (2021). Genomic investigation of milk production in Italian buffalo. *Ital. J. Anim. Sci.* 20, 539–547. doi:10.1080/1828051X.2021.1902404.

Cesarani, A., Pocrnic, I., Macciotta, N. P. P., Fragomeni, B. O., Misztal, I., and Lourenco, D. A. L. (2019). Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies. *J. Anim. Breed. Genet.* 136, 40–50. doi:10.1111/jbg.12367.

Chen, T., and Guestrin, C. (2016). XGBoost. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA: ACM), 785–794. doi:10.1145/2939672.2939785.

Cherkassky, V., and Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* 17, 113–126. doi:10.1016/S0893-6080(03)00169-2.

Christensen, O., and Lund, M. (2010). Genomic relationship matrix when some animals are not genotyped. *Genet. Sel. Evol.* 42, 1–8.

Evgeniou, T., and Pontil, M. (2005). *Support Vector Machines: Theory and Applications*. In Machine. , ed. L. Wang Berlin, Heidelberg: Springer Berlin Heidelberg doi:10.1007/b95439.

Falconer, D. S., and Mackay, T. F. C. (1996). Introduction to Quantitative Genetics. *Longmans Green, Harlow, Essex, UK* Ed 4., 464.

Fragomeni, B. O., Lourenco, D. A. L., Legarra, A., VanRaden, P. M., and Misztal, I. (2019). Alternative SNP weighting for single-step genomic best linear unbiased predictor evaluation of stature in US Holsteins in the presence of selected sequence variants. *J. Dairy Sci.* 102, 10012–10019. doi:10.3168/jds.2019-16262.

Fragomeni, B. O., Lourenco, D. A. L., Masuda, Y., Legarra, A., and Misztal, I. (2017). Incorporation of causative quantitative trait nucleotides in single-step GBLUP. *Genet. Sel. Evol.* 49, 1–11. doi:10.1186/s12711-017-0335-0.

Friedman, J. H. (2002). Stochastic gradient boosting. *Comput. Stat. Data Anal.* 38, 367–378. doi:10.1016/S0167-9473(01)00065-2.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33, 1–22.

Frouin, A., Dandine-Roulland, C., Pierre-Jean, M., Deleuze, J. F., Ambroise, C., and Le Floch, E. (2020). Exploring the Link Between Additive Heritability and Prediction Accuracy From a Ridge Regression Perspective. *Front. Genet.* 11, 1–15. doi:10.3389/fgene.2020.581594.

Gianola, D. (2013). Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* 194, 573–596. doi:10.1534/genetics.113.151753.

Gilmour, A. R., Thompson, R., and Cullis, B. R. (1995). Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics* 51, 1440–1450. doi:10.2307/2533274.

Gualdrón Duarte, J. L., Cantet, R. J. C., Bates, R. O., Ernst, C. W., Raney, N. E., and Steibel, J. P. (2014). Rapid screening for phenotype-genotype associations by linear

transformations of genomic evaluations. *BMC Bioinformatics* 15, 1–11. doi:10.1186/1471-2105-15-246.

Gualdrón Duarte, J. L., Gori, A. S., Hubin, X., Lourenco, D., Charlier, C., Misztal, I., et al. (2020). Performances of Adaptive MultiBLUP, Bayesian regressions, and weighted-GBLUP approaches for genomic predictions in Belgian Blue beef cattle. *BMC Genomics* 21, 1–18. doi:10.1186/s12864-020-06921-3.

Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genomic BLUP decoded: A look into the black box of genomic prediction. *Genetics* 194, 597–607. doi:10.1534/genetics.113.152207.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. New York, NY: Springer New York.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model published by : international biometric society stable. *Biometrics* 31, 423–447.

Karaman, E., Cheng, H., Firat, M. Z., Garrick, D. J., and Fernando, R. L. (2016). An upper bound for accuracy of prediction using GBLUP. *PLoS One* 11, 1–18. doi:10.1371/journal.pone.0161054.

Legarra, A. Reverter, A. (2017). Can We Frame and Understand Cross-Validation Results in Animal Breeding? *Proc. Assoc. Advmt. Anim. Breed. Genet.* 22, 73–80.

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. doi:10.3168/jds.2009-2061.

Legarra, A., and Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method 01 Mathematical Sciences 0104 Statistics. *Genet. Sel. Evol.* 50, 1–18. doi:10.1186/s12711-018-0426-6.

Li, B., Zhang, N., Wang, Y. G., George, A. W., Reverter, A., and Li, Y. (2018). Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Front. Genet.* 9, 1–20. doi:10.3389/fgene.2018.00237.

Liu, Z. (2010). Interbull validation test for genomic evaluations. *Interbull Bull.*, 17.

Macedo, F. L., Christensen, O. F., Astruc, J. M., Aguilar, I., Masuda, Y., and Legarra, A. (2020). Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genet. Sel. Evol.* 52, 1–10. doi:10.1186/s12711-020-00567-1.

Mancin, E., Lourenco, D., Bermann, M., and Mantovani, R. (2021a). Accounting for Population Structure and Phenotypes From Relatives in Association Mapping for Farm Animals : A Simulation Study. doi:10.3389/fgene.2021.642065.

Mancin, E., Tuliozi, B., Pegolo, S., Sartori, C., and Mantovani, R. (2022). Genome Wide Association Study of Beef Traits in Local Alpine Breed Reveals the Diversity of the Pathways Involved and the Role of Time Stratification. 12, 1–22. doi:10.3389/fgene.2021.746665.

Mancin, E., Tuliozi, B., Sartori, C., Guzzo, N., and Mantovani, R. (2021b). Genomic prediction in local breeds: The rendena cattle as a case study. *Animals* 11, 1–19. doi:10.3390/ani11061815.

Mastrangelo, S., Saura, M., Tolone, M., Salces-Ortiz, J., Di Gerlando, R., Bertolini, F., et al. (2014). The genome-wide structure of two economically important indigenous sicilian cattle breeds. *J. Anim. Sci.* 92, 4833–4842. doi:10.2527/jas.2014-7898.

Masuda, Y., VanRaden, P. M., Misztal, I., and Lawlor, T. J. (2018). Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection. *J. Dairy Sci.* 101, 5194–5206. doi:10.3168/jds.2017-13310.

Mehrban, H., Naserkheil, M., Lee, D. H., Cho, C., Choi, T., Park, M., et al. (2021). Genomic Prediction Using Alternative Strategies of Weighted Single-Step Genomic BLUP for Yearling Weight and Carcass Traits in Hanwoo Beef Cattle. *Genes (Basel).* 12, 266. doi:10.3390/genes12020266.

Meuleman, W., and Shaker, A. J. (2012). “A rough R Impementation of the Bagplot , Data Peeling , Skyline Plots , and Graphical Summaries,” in.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi:11290733.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., et al. (2020). “e1071: Misc Functions of the Department of Statistics, Probability Theory Group,” in (TU Wien), 1–63.

Misztal, I., Aggrey, S. E., and Muir, W. M. (2013). Experiences with a single-step genome evaluation. *Poult. Sci.* 92, 2530–2534. doi:10.3382/ps.2012-02739.

Mitchell, R., and Frank, E. (2017). Accelerating the XGBoost algorithm using GPU computing. *PeerJ Comput. Sci.* 3, e127. doi:10.7717/peerj-cs.127.

Natekin, A., and Knoll, A. (2013). Gradient boosting machines, a tutorial. *Front. Neurorobot.* 7, 1–21. doi:10.3389/fnbot.2013.00021.

Piles, M., Bergsma, R., Gianola, D., Gilbert, H., and Tusell, L. (2021). Feature Selection Stability and Accuracy of Prediction Models for Genomic Prediction of Residual Feed Intake in Pigs Using Machine Learning. *Front. Genet.* 12. doi:10.3389/fgene.2021.611506.

Pocrnic, I., Lourenco, D. A. L., Masuda, Y., and Misztal, I. (2019). Accuracy of genomic BLUP when considering a genomic relationship matrix based on the number of the largest eigenvalues: A simulation study. *Genet. Sel. Evol.* 51, 1–10. doi:10.1186/s12711-019-0516-0.

Ren, D., An, L., Li, B., Qiao, L., and Liu, W. (2021). Efficient weighting methods for genomic best linear-unbiased prediction (BLUP) adapted to the genetic architectures of quantitative traits. *Heredity (Edinb).* 126, 320–334. doi:10.1038/s41437-020-00372-y.

Ročková, V., and George, E. I. (2018). The Spike-and-Slab LASSO. *J. Am. Stat. Assoc.* 113, 431–444. doi:10.1080/01621459.2016.1260469.

Sanz, H., Valim, C., Vegas, E., Oller, J. M., and Reverter, F. (2018). SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinformatics* 19, 432. doi:10.1186/s12859-018-2451-4.

Sargolzaei, M., and Schenkel, F. S. (2009). QMSim: a large-scale genome simulator for livestock. *Bioinformatics* 25, 680–681. doi:10.1093/bioinformatics/btp045.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. Available at: http://www.jstor.org/stable/2346178.

VanRaden, P. M. (2008a). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980.

VanRaden, P. M. (2008b). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980.

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. New York, NY: Springer New York.

Vitezica, Z. G., Aguilar, I., Misztal, I., and Legarra, A. (2011). Bias in genomic predictions for populations under selection. *Genet. Res. (Camb).* 93, 357–366. doi:10.1017/S001667231100022X.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., Fernando, R. L., Vitezica, Z., et al. (2014). Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front. Genet.* 5, 134. doi:10.3389/fgene.2014.00134.

Zhang, X., Lourenco, D., Aguilar, I., Legarra, A., and Misztal, I. (2016). Weighting Strategies for Single-Step Genomic BLUP: An Iterative Approach for Accurate Calculation of GEBV and GWAS. *Front. Genet.* 7, 151. doi:10.3389/fgene.2016.00151.

Zhu, S., Guo, T., Yuan, C., Liu, J., Li, J., Han, M., et al. (2021). Evaluation of Bayesian alphabet and GBLUP based on different marker density for genomic prediction in Alpine Merino sheep. *G3 Genes, Genomes, Genet.* 11. doi:10.1093/g3journal/jkab206.

10. ACCOUNTING FOR POPULATION STRUCTURE AND PHENOTYPES FROM RELATIVES IN ASSOCIATION MAPPING FOR FARM ANIMALS: A SIMULATION STUDY

STATUS: PUBLICATED ON FORNTIERS IN GENETICS
https://doi.org/10.3389/fgene.2021.642065

Accounting for population structure and phenotypes from relatives in association mapping for farm animals:a simulation study

Enrico Mancin*, Daniela Lourenco, Matias Bermann, Roberto Mantovani, Ignacy Misztal

ABSTRACT

Population structure or genetic relatedness should be considered in genome association studies to avoid spurious association. The most used methods for genome-wide association studies (GWAS) account for population structure but are limited to genotyped individuals with phenotypes. Single- step GWAS (ssGWAS) can use phenotypes from non-genotyped relatives; however, its ability to account for population structure has not been explored. Here we investigate the equivalence among ssGWAS, efficient mixed-model association expedited (EMMAX), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS), and how they differ from the single-SNP analysis without correction for population structure (SSA-NoCor). We used simulated, structured populations that mimicked fish, beef cattle, and dairy cattle populations with 1040, 5525, and 1400 genotyped individuals, respectively. Larger populations were also simulated that had up to 10-fold more genotyped animals. The genomes were composed by 29 chromosomes, each harboring one QTN, and the number of simulated SNPs was 35,000 for the fish and 65,000 for the beef and dairy cattle populations. Males and females were genotyped in the fish and beef cattle populations, whereas only males had genotypes in the dairy population. Phenotypes for a trait with heritability varying from 0.25 to 0.35 were available in both sexes for the fish population, but only for females in the beef and dairy cattle populations. In the latter, phenotypes of daughters were projected into genotyped sires (i.e., deregressed proofs) before applying EMMAX and SSA-NoCor. Although SSA-NoCor had the largest number of true positive SNPs among the four methods, the number of false negatives was two- to five-fold that of true positives. GBLUP-GWAS and EMMAX had a similar number of true positives, which was slightly smaller than in ssGWAS, although the difference was not significant. Additionally, no significant differences were observed when deregressed proofs were used as pseudo-phenotypes in EMMAX compared to daughter phenotypes in ssGWAS for the dairy cattle population. Single-step GWAS accounts for population structure and is a straightforward method for association analysis when only a fraction of the population is genotyped and/or when phenotypes are available on non-genotyped relatives.

INTRODUCTION

Genome-wide association (GWA) aims to identify regions in the genome that are related to diseases or traits of interest (Begum *et al.,* 2012). The method is most often based on statistical tests to determine if a single nucleotide polymorphism (SNP) is statistically associated with the trait, at a given probability value (p-value). If the association is significant, the interrogated SNP may be in high linkage disequilibrium (LD) with a causative variant, or the SNP itself may be a common variant that has a large effect on the trait, although having one or a few causative variants and validating them can be difficult (Kennedy *et al.,* 1992). In fact, results from GWA study (GWAS) have confirmed that most of the complex traits in humans (Yang *et al.,* 2011), animals (Oliveira Silva *et al.,* 2017) and plants (Bian and Holland*,* 2017) are polygenic. Even in such a case, the GWAS still fulfills the primary goal of helping to better understand the biology of a trait.

The first GWAS was developed to understand the biology of human diseases aiming the prevention (Bruton *et al.,* 2007). Although a couple of studies were published a few years before, the study from 2007 is considered the landmark of GWAS because it resulted from a well-designed, large- scale study (Visscher *et al.,* 2012). After that, GWAS was also adopted in livestock and plants. The very first studies were based on single-SNP analysis where each SNP is tested independently (Baling, 2006). However, this approach assumes SNPs are identically and independently distributed, which is only true when a population is comprised of unrelated individuals (Risch and Merikangas, 1996). As populations contain related individuals, not considering population structure or genetic relatedness in GWAS can result in spurious associations (Sul *et al.,* 2018) To resolve the problem with population structure, the use of principal components (PC) to model relationships have been suggested (Price *et al.,* 2006). Still, the level of confounding in GWAS was considerable when 100 PC were fit into the model or when highly related individuals were removed from a human population (Sul *et al.,* 2018).

A well-known approach among animal breeders, the mixed linear models (Henderson, 1975), was then adopted for human GWAS showing to be a reasonable approach to take population structure into account (Kang *et al.,* 2008; Kang *et al.,* 2010). In this method, known as efficient mixed-model association expedited (EMMAX), one SNP is fit in the model as a

fixed covariate and, at the same time, a relationship matrix corrects for population structure. However, EMMAX-based methods consider only genotyped individuals with phenotypes. However, only a fraction of individuals in a population are genotyped, particularly in livestock and aquaculture. Because of that, the original mixed linear models were extended to account for genotyped and non-genotyped individuals in prediction analysis (Aguilar *et al.,* 2010; Christensen and Lund 2010). This method is called single- step genomic best linear unbiased prediction (ssGBLUP) and is widely adopted for genomic predictions in livestock (Legarra *et al.,* 2014; Misztal *et al.,* 2020) and plants (Cappa *et al.,* 2019), and was recently applied to predict polygenic risk score in humans (Truong *et al.,* 2020). The popularity of ssGBLUP is due to the added value of phenotypes for relatives that are not genotyped, and the simplicity when combining information from genotyped and non-genotyped individuals (Legarra *et* al., 2014).

The usefulness of ssGBLUP to GWAS in a procedure called single-step GWAS (ssGWAS) was subsequently extended (Wang *et al*., 2012). In this method, SNP effects and variance explained by SNPs are computed simultaneously for all SNPs while accounting pedigree and genomic relationships in addition to all phenotypes available. However, no statistical significance test was available under the ssGWAS framework. Later it has been shown that the statistical test used in EMMAX has a mathematical equivalent that can be used in GBLUP-based methods (Gualdrón Duarte *et al.,* 2014; Bernal Rubio *et al.;* 2016) even though SNPs are considered fixed in the former and random in the latter. This equivalent statistical test was then implemented in ssGWAS (Aguilar *et al.,* 2019) so that p-values are computed based on prediction error variance of SNP effects. Because of the mathematical equivalence, results from ssGWAS are expected to be similar to the ones from EMMAX. Here we use different simulated, structured populations (i.e., beef cattle, dairy cattle, and fish) to investigate the equivalence among EMMAX, ssGWAS, and GBLUP-GWAS, and how they differ from single-SNP analysis. We also evaluate whether the population structure is fully considered by the mixed linear models, and when ssGWAS should be the method of choice forassociation studies in related populations. We demonstrate ssGWAS performs similarly to EMMAX and GBLUP-GWAS when genotyped animals have their own phenotypes or when only progeny phenotypes are available, so deregressed proofs must be used for EMMAX and GBLUP-GWAS.

MATERIALS AND METHODS

Although the main objective here was to demonstrate the equivalence of EMMAX and ssGWAS, we also compared those two methods against the single-SNP analysis without correction for population structure (SSA-NoCor) and the GBLUP-based GWAS (GBLUP-GWAS) that uses only genotyped individuals with phenotypes.

Methods and Computations

Single-SNP Analysis without Correction for Population Structure (SSA-NoCor):

To estimate allele substitution effect of the *i*th SNP with SSA-NoCor, the following model was used:

$$y = 1\mu + \mathrm{x}_i g_i + e \quad (1)$$

where y is the vector of phenotypes, μ is the mean, x_i is a vector that contains the genotype for the *i*th SNP for each animal, g_i is the *i*th allele substitution effect, and $e \sim \mathrm{N}(0, \mathrm{I}\sigma_e^2)$ is the residual. The estimate of g_i and its variance was obtained by least squares.

Single-SNP Analysis with Correction for Population Structure Using a Genomic Relationship Matrix (EMMAX):

For the EMMAX method, the estimated allele substitution effect and its variance were obtained from the BLUE of the following linear mixed model:

$$y = 1\mu + \mathrm{x}_i g_i + \mathrm{Z}a + e \quad (2)$$

where Z is a design matrix, $a \sim \mathrm{N}(0, \mathrm{G}\sigma_a^2)$ is the vector of breeding values (i.e., animal effect), G is the genomic relationship matrix, and the rest of the components were previously defined. The G matrix was calculated as first method following literature (Zhou and Stephens, 2012):

$$G = \frac{1}{p} \sum_{i=1}^{p} (x_i - 1_n \bar{\mathrm{x}}_i)(x_i - 1_n \bar{\mathrm{x}}_i)^T \quad (3)$$

where x_i stands for ith SNP locus column and $\bar{\mathrm{x}}_i$ represented the mean of ith locus, *n* is the number of samples.

GBLUP Association (GBLUP-GWAS):

For the GBLUP-GWAS, the vector of estimated allele substitution effects $\hat{g}$ was obtained from a linear transformation of the BLUP of a under a GBLUP model:

$$y = 1\mu + \mathrm{Z}a + e \qquad (4)$$

of which the mixed model equations can be represented by:

$$\begin{bmatrix} 1'1 & 1'\mathrm{Z} \\ \mathrm{Z}'1 & \mathrm{Z}'\mathrm{Z}+\mathrm{G}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} 1'y \\ \mathrm{Z}'y \end{bmatrix} \qquad (5)$$

In this method, G_0 was estimated by the second method based on [26]:

$$\mathrm{G}_0 = \frac{\mathrm{MM}'}{2\sum p_i(1-p_i)} \qquad (6)$$

where M is a matrix of SNP content centered by twice the current allele frequencies, and p_i is the allele frequency for the i^{th} SNP (VanRaden, 2008). Additionally, to avoid singularity problems, the final G was computed as

$$\mathrm{G} = \lambda\mathrm{G}_0 + \beta\mathrm{I} \quad (7)$$

with λ=0.95 and β=0.05.

Afterwards, the vector of allele substitution effects ($\hat{g}$) was calculated for all SNPs simultaneously (Wang *et al*. 2012):

$$\hat{g} = \lambda\frac{1}{2\sum pq}\mathrm{M}'\mathrm{G}^{-1}\hat{a} \quad (8)$$

with $q = 1 - p$.

The variance of SNP effects, which is needed to compute *p-values* when SNPs are considered random was calculated following as (Gualdrón Duarte *et al.,* 2014):

$$Var(\hat{g}) = \lambda\frac{1}{2\sum pq}\mathrm{Z}'\mathrm{G}^{-1}(\mathrm{G}\hat{\sigma}_a^2 - \mathrm{C}^{22})\mathrm{G}^{-1}\mathrm{Z}\,\lambda\frac{1}{2\sum pq} \quad (9)$$

where C^{22} is the block of the inverse of the MME corresponding to the animal effect. The *p-value* for each SNP effect was then computed with the formula (Gualdrón Duarte *et al*., 2014):

$$p - value_i = 2\left(1 - \Phi\left(\left|\frac{\hat{g}_i}{sd(\hat{g}_i)}\right|\right)\right) \quad (10)$$

where $sd(\hat{g}_i)$ is the standard error of the SNP effect or simply $sd(\hat{g}_i) = \sqrt{Var(\hat{g}_i)}$; $\Phi(\cdot)$ is the cumulative density function (CDF) of the standard normal distribution. For a justification of using $sd(\hat{g}_i) = \sqrt{Var(\hat{g}_i)}$ in the denominator instead of $\sqrt{Var(g_i) - Var(\hat{g}_i)}$ (Gualdrón Duarte *et al*., 2014):

Single-Step GBLUP association (ssGWAS):

This method differs from GBLUP-GWAS in the sense that all animals in the pedigree can be used, not only genotyped animals with phenotypes. Therefore, G^{-1} is replaced by H^{-1} in (5), and the latter combines pedigree and genomic relationships (Aguilar *et al.,* 2010):

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \quad (11)$$

where A^{-1} and A_{22}^{-1} are the inverses of the pedigree relationship matrix for all animals and only genotyped animals, respectively. Pedigree and genomic relationships have different genetic base because allele frequencies from the current genotyped population are used to center G. Therefore, G in ssGWAS is adjusted so the average diagonal and off-diagonal matches the averages of A_{22}. Because of this adjustment, (8) and (9) were modified to:

$$\hat{g} = \lambda\delta\frac{1}{2\sum pq}M'G^{-1}\hat{a}_{22} \quad (12)$$

and

$$Var(\hat{g}) = \lambda\delta\frac{1}{2\sum pq}Z'G^{-1}(G\hat{\sigma}_a^2 - C^{22})G^{-1}Z\,\lambda\delta\frac{1}{2\sum pq} \quad (13)$$

where $\hat{a}_{22}$ is a vector of genomic estimated breeding values (GEBV) for genotyped animals; δ accounts for the difference in genetic base between the pedigree and genomic relationship matrices, and was calculated as (Vitezica *et al.,* 2011):

$$\delta = 1 - \frac{0.5}{n^2}(\sum_i \sum_j \mathrm{A}_{22(\mathrm{i,j})} - \sum_i \sum_j \mathrm{G}_{\mathrm{i,j}}) \qquad (14)$$

with *n* the number of genotyped animals. After the modification, P-values in ssGWAS were obtained as in (10) and previously suggested (Aguilar *et al.,* 2019).

Note that the dimension of ssGWAS system of equations is greater than the dimension of GBLUP-GWAS because of the inclusion of non-genotyped animals.

Computations:

The allele substitution effects were estimated with different software: (*i*) SSA-NoCor solutions were computed with GASTON R-package (Dandine-Roulland and Perdry, 2018) , (*ii*) EMMAX solutions were obtained with GEMMA software (Zhou*,* 2016), (*iii*) GBLUP-GWAS and (*iv*) ssGWAS were computed using the BLUPF90 software suite (Misztal *et al.,* 2015).

Overview of Data Simulation

We used different simulated datasets to investigate the equivalence between EMMAX and ssGWAS, and to explore the usefulness of each method compared to SSA-NoCor and GBLUP-GWAS. A fish, a beef cattle, and two dairy cattle populations were simulated using QMSim (Sargolzaei and Schenkel, 2009), with five replicates for each. The general parameters for each population such as the number of genotyped animals, effective population size, type of trait, and heritability are reported in Table 1.

Table 1. Simulated population structure

	Simulated population		
	Fish	Beef	Dairy
Number of records	2040	5088	70,000
Number of animals in the pedigree	2040	10,000	140,000
Number of genotyped animals	1040	5525	1400
Number of genotyped animals with records	1040	3039	1400[1]
Type of trait	Both sex	Sex-limited	Sex-limited
Heritability	0.25	0.30	0.30

[1] Number of genotyped animals with projected phenotypes of progeny (i.e., deregressed proofs)

Fish Population

The historical population began with 5000 animals and decreased to 3100 after 200 non-overlapping generations, that were carried out to generate linkage disequilibrium (LD) and mutation-drift equilibrium. The proportion of males in the historical population was 31%. Aiming to mimic a real fish selection scheme (Garcia *et al.,* 2018), a recent population was created by randomly selecting 20 sires and 20 dams. The recent population was subject to random mating for five generations. In every generation, each female had 2 offspring. After the fifth generation, a new line was created by mating 20 males and 20 females randomly. For this line, the litter size was set to 100 offspring per dam. Sire and dam culling rates were set to 0.5 and 0.2, respectively. Phenotypes and pedigree of the animals in the new line, together with their parents, were considered for the association analysis.

The genome was composed of 29 chromosomes with a length of approximately 100 cM each, 35,000 evenly spaced SNPs, and one QTN per chromosome. Each QTN was placed in the middle of its respective chromosome. Although this number of QTN is not the reality of most of the traits of interest (i.e., complex traits), this assumption was made to facilitates the QTN discovery in the association analysis. Altogether, the QTNs accounted for the total of the genetic variation and their effects were assumed to follow a gamma distribution with shape parameter equal to 0.40. The allele frequencies for SNPs and QTNs in the first historical

generation were 0.5, and a recurrent mutation rate of 2.5e-5 per locus per generation was assumed. A single trait with heritability of 0.25 was simulated, and a single phenotype per animal was obtained by adding an overall mean of 1.0, the sum of the QTN effects, and a residual effect.

Beef Cattle Population

In this dataset, the historical population began with 1000 animals and steadily increased to 50,000 after 1000 generations of random mating. Then, a decrease in number of individuals followed for another 1000 generations. After 2000 generations, the historical population was composed by 23,000 animals, of which 3000 were males. The recent beef cattle population was created by randomly selecting 10000 dams and 200 sire and allowing them to mate randomly for five discrete generations. Afterwards, five groups of 10 sires and 500 dams each were selected based on TBV to create five different lines. With the aim of maximizing the difference between the five lines, selection based on TBV was used in each of them. Finally, 10 sires and 500 dams from each of the five lines were pooled in one single line and underwent random mating for five generations. This process was designed to create an intricate population structure. For the present population, a sex-limited trait was simulated so that only females had a phenotype for a trait with heritability of 0.30.

Genotypes were simulated for males and females from the last generation of the population and their parents (n_{beef} = 5525). The parameters to simulate the beef cattle genome were the same as in the fish population except for the number of SNPs, which for the beef cattle population was equal to 65,000.

Dairy Cattle Population

The parameters for the simulation of the dairy historical population were the same as those in beef cattle. A total of 1000 sires and 20,000 were chosen as founders of the recent population. This population was subject to selection based on estimated breeding values (EBV) for 10 generations and assortative mating based on inbreeding (Sonesson and Meuwissen, 2000). In this simulation, an Ne between 100-150 was maintained. The Ne was calculated as the change in inbreeding (ΔF) from one generation to the next using the following formula (Falconer and Mackay, 1996):

$$\Delta F = \frac{F_n - F_{n-1}}{1 - F_{n-1}} \quad (15)$$

$$Ne = \frac{1}{2\Delta F} (16)$$

where F_n is the inbreeding in the n^{th} generation.

All the parameters for the genome simulation were similar to the ones in the beef cattle population. The only difference was the genotyping strategy that included only sires of the seventh generation (n_{dairy} = 1400). Phenotypes for a trait of heritability 0.3 were available only on females (Table 1). On average, each genotyped sire in this population had 10 daughters with records, and dairy_10d will be used to refer to this dataset. A second dairy cattle population was generated with the same parameters, but with five daughters with records per sire. This population will be referred to as dairy_5d and was created to mimic a situation where deregressed proofs (DP) of sires have lower reliability.

Deregressed Proofs (DRP)

One requirement in association analysis is that individuals should have both genotypes and phenotypes. In some livestock populations, genotypes may be available for males and phenotypes for females (e.g., milk production in dairy cattle). In such a case, DRP are needed as an input for SSA-NoCor, EMMAX, or GBLUP-GWAS. The DRP are projections of female phenotypes into their relatives' genotypes. Because sex-limited traits were simulated for the beef and dairy cattle populations, DRP were computed for sires in both populations following (VanRaden and Sullivan, 2010; Wiggans *et al.,* 2011).

$$DRP = PA + \frac{EBV - PA}{DE_{prog}/(DE_{prog} + DE_{PA} + 1)} \qquad (17)$$

*w*here PA is parent average; $DE_{prog} = \left[\frac{EBV_{rel}}{(1 - EBV_{rel})}\right] - DE_{PA}$ and is the daughter equivalent from progeny information; and $DE_{PA} = \frac{PA_{rel}}{(1 - PA_{rel})}$ is the daughter equivalent from PA. The EBV_{rel} and PA_{rel} are reliabilities of parent average and EBV, respectively. All EBV, PA, and reliabilities used in the DRP formula were computed using the BLUPF90 software suite (Misztal *et al.,* 2015). The DRP were used for the association analysis of dairy cattle datasets under SSA-

NoCor, EMMAX, and GBLUP-GWAS. As ssGWAS uses all phenotypes, genotypes, and pedigree information available, it does not rely on DP.

Quality Control Prior to the Association Analysis

Quality control of genomic data removed monomorphic SNPs, SNPs with minor allele frequency (MAF) lower than 0.05, and with deviation between observed and expected allele frequencies greater than 0.15. After quality control, and average of 35,000, 58,000, and 58,00 SNPs were kept for the analysis in the fish, beef cattle, and dairy cattle population, respectively.

Significance and Concordance Tests

A single SNP was considered significantly associated with the considered trait when its *p*-value was smaller than a certain significance level, which was 0.05 with a Bonferroni correction for multiple testing, i.e., $0.05/(number\ of\ SNPs)$. Additionally, true positive (TP) and false positive (FP) rates were computed for each scenario using a window size of ± 2 cM, which is equivalent to 20-30 markers (Toosi *et al.,* 2018)

RESULT AND DISCUSION

Population Structure

Figures 1 and 2 show plots with the first (PC1) and second (PC2) principal components of **G** for small and large populations, respectively. PC1 and PC2 represent the two largest sources of variation in the data, and are often used to investigate population structure, which was deemed important in our study. The level of population stratification differed among the simulated populations because of the different selection and mating strategies. For the small populations, distinct family groups (full-sibs) were observed for the fish population, with variable size and impact on the model (extreme clusters farther from PC1=PC2=0); however, the PC scores were of small magnitude. No distinct clustering was detected in the beef cattle population, although a level of variability was observed. Overall, it was not possible to discriminate different groups of individuals but some of them appear more genetically different than others. In fact, animals belonging to five different lines were randomly mated for ten generations, and only genotypes for animals in the 7th generation were retained. Therefore, less genetic distance among animals was created. Still, the largest graphical distance was

observed in the dairy cattle population, with a large cluster centered in zero and a few animals genetically distant from the main cluster; however, the amount of variation explained by PC1 and PC2 was small. Possibly, this pattern resulted from non- overlapping generations that created extra genetic distance among some animals. These distant animals were sires that had EBV departing from the population mean. For the large populations, no clustering was observed for the fish and beef populations, whereas the same pattern was observed for dairy cattle in both populations.

Figure 1. Linkage disequilibrium (r2) decay for the small populations of fish (A), beef (B), and dairy cattle(C).

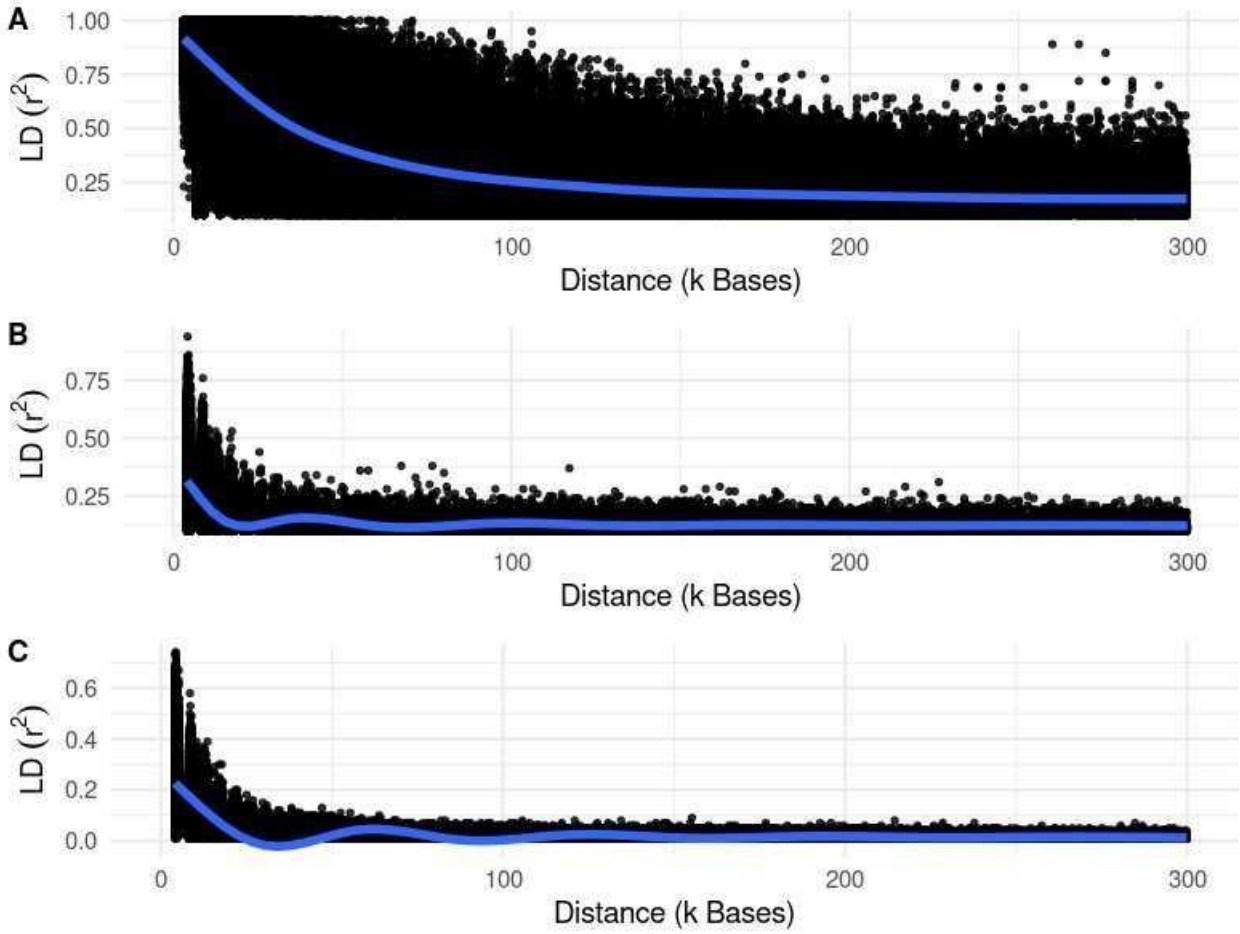


Figure 2. Linkage disequilibrium (r2) decay for the large populations of fish (A), beef (B), and dairy cattle(C).

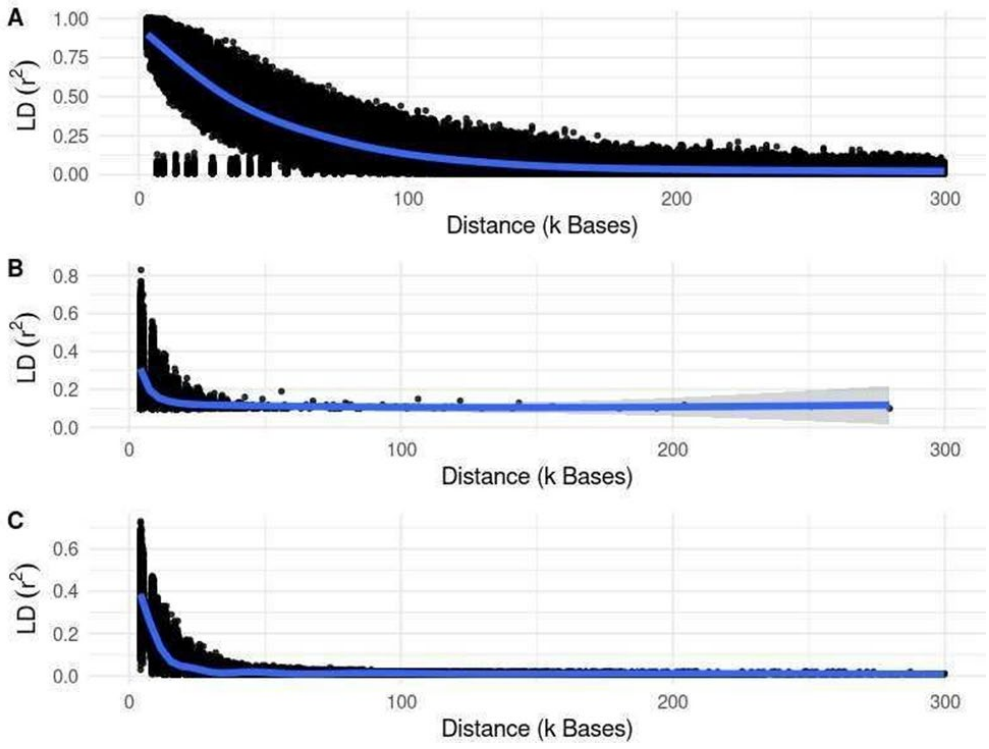


Figure 3. First and second principal components of the genomic relationship matrices for the small populations of fish, beef, and dairy cattle.

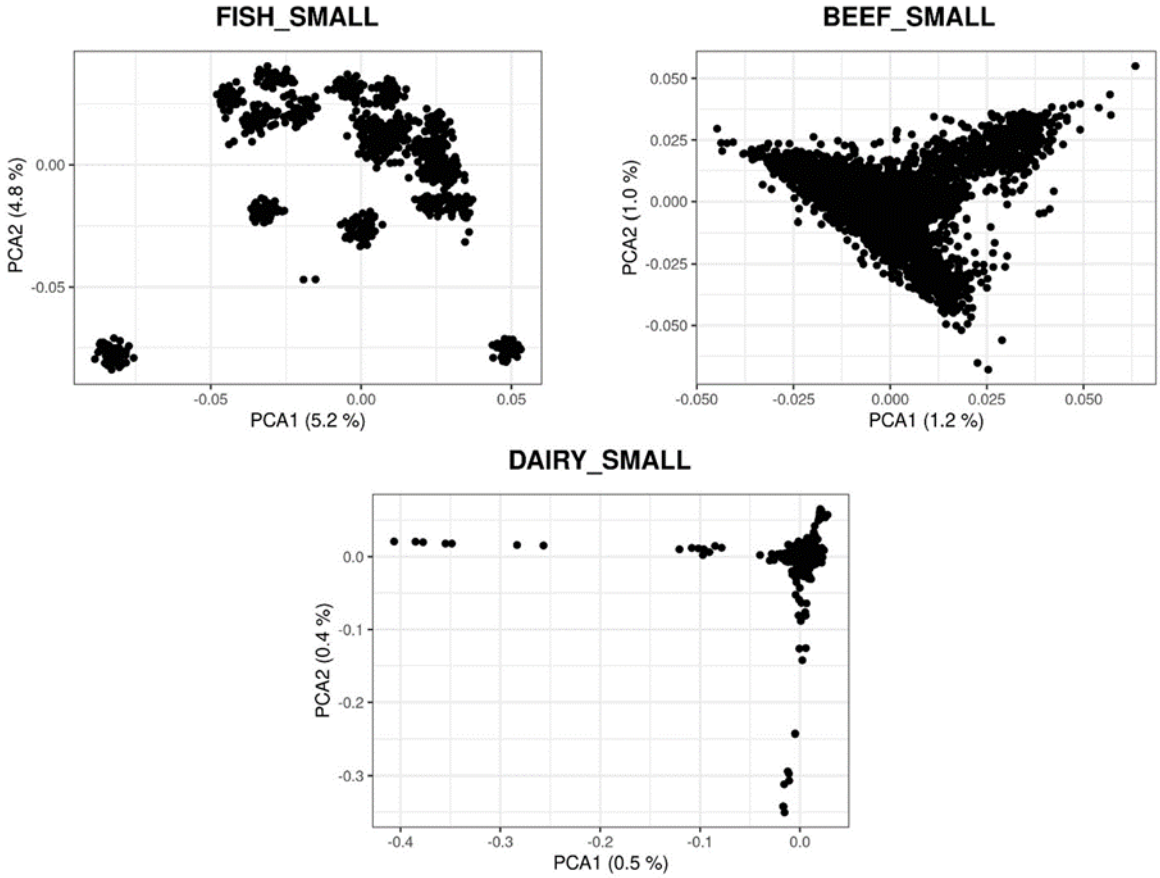
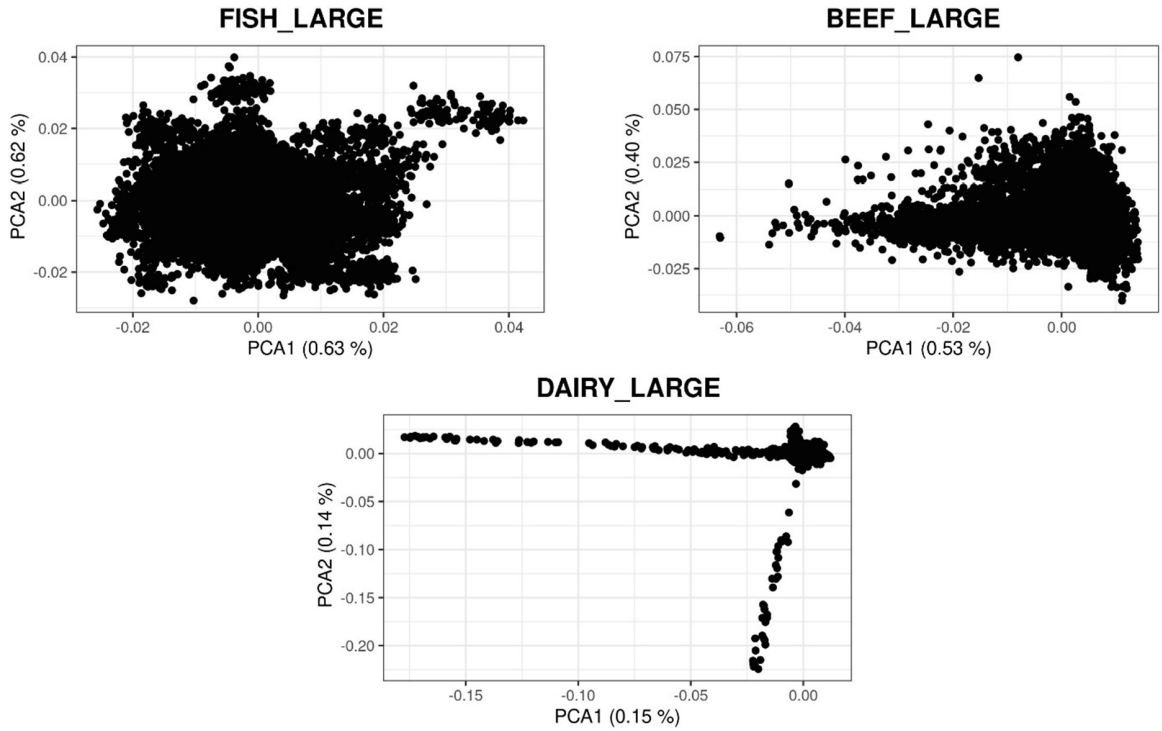


Figure 4. First and second principal components of the genomic relationship matrices for the large populations of fish, beef, and dairy cattle



Association Analysis

Manhattan plots with p-values for the fish, beef cattle, and dairy cattle populations using SSA- NoCor, EMMAX, and ssGWAS, and GBLUP-GWAS are in Figures 5-10. Although one QTN was simulated in each chromosome, the signals were not equally strong because of the assumption of a Gamma distribution, and the selection that the populations underwent. Overall, selection caused fixation for 5 to 6% of the QTN. To better access the information in the Manhattan plots, the average number of true and false positive SNPs were computed and placed in Table 2. Although thenumber of TP and FP differed among EMMAX, ssGWAS, and GBLUP-GWAS, the differences were not statistically significant (p-value > 0.05).

Table 2. Average number (± SE) of true positive (TP) and false positive (FP) SNPs for all the simulated populations

Population	Association	SSA_NoCor	EMMAX	ssGWAS	GBLUP-GWAS
Fish	TP	68.4 ± 7.50	16.2 ± 7.85	16.8 ± 11.70	13.6 ± 7.85
	FP	155 ± 99.90	0.4 ± 0.54	0.2 ± 0.45	0.0 ± 0.00
Beef	TP	16.2 ± 11.2	5.6 ± 4.10	7.6 ± 3.78	5.6 ± 4.10
	FP	64.8 ± 56.5	0.0 ± 0.00	0.2 ± 0.44	0 ± 0.00
Dairy_10d	TP	13.5 ± 3.11	8.75 ± 5.91	13.2 ± 0.50	8.75 ± 5.12
	FP	24.5 ± 8.8	0.0 ± 0.00	0.25 ± 0.50	0.0 ± 0.00
Dairy_5d	TP	18 ± 4.20	10 ± 2.45	13 ± 1.87	5.6 ± 0.89
	FP	51.6 ± 25.6	0.0 ± 0.00	0.4 ± 0.55	0 ± 0.00

For all the simulated populations, the greatest number of false positive SNPs was observed for SSA-NoCor. These results agree with those from previous studies (e.g. Yang et al., 2014), which showed that the number of false positives drastically decreased when correcting for population structure. For the small populations, the number of false positive SNPs in SSA-NoCor for the dairy dataset was the smallest one compared to the other simulated populations, whereas for the large populations, the smallest number of false positive SNPs occurred in the beef population. In both cases, the fish population had the greatest number of false positive signals. Since this population had a strong structure (e.g., several separate clusters), it can be concluded that the population structure is related to the

number of false positive signals captured by SSA-NoCor. False positive associations capture SNPs that relate to the genetic differences between sub-populations and also with the trait considered (Sul *et al.,* 2018). These spurious signals can also be interpreted as a wrong prior assumption of marker effects in the SSA-NoCor model. In such a model, markers are considered independently distributed, which implicitly means linkage disequilibrium among SNPs is neglected (Sul et al., 2018; Finno et al., 2014). The effect of population structure is even more evident when small sample size and high-density panels are used in association analyses (Finno et al., 2014). Furthermore, when traits are polygenic or have low heritability, signals deriving from population structure can completely override those deriving from true QTNs (Toosi *et al.,* 2018; Atwell *et al.,* 2010).

In terms of true positives, two situations were observed. First, SSA-NoCor detected significantly more true positives than the other methods in the fish and large dairy cattle simulated populations. However, the number of false positives was sometimes 3-fold greater than that of true positives. The identification of false and true positives is straightforward in simulated data but not in real data. Second, no significant differences were observed among methods for the beef and the small dairy population. Based on the results from the dairy population, it can be concluded that the use of DRP for bulls in EMMAX and GBLUP-GWAS, compared to the raw phenotypes in ssGWAS, did not promote a loss in GWA resolution. The loss in the ability to correctly detect QTNs when using DRP is expected in complex models when the estimation of fixed and random effects is not very accurate. According to Aguilar et al. (2019), information can be lost in the deregression process, which may result in spurious signals in GWA. Although DRP were used in the dairy population, no significant differences in TP and FP were observed between EMMAX and ssGWAS because the model was simple and included only a general mean as fixed effect and the additive genetic as random.

The methods in our study were used as binary classifiers when trying to identify true and false positives. The quality of a binary classifier can be evaluated from the degree of randomness of the decisions of that classifier. A perfect classifier is not random, whereas the worst classifier would determine whether a signal is true or false with a probability equal to 0.5 (Agresti, 2013). To compare the methods in our study as binary classifiers, Receiver Operating Characteristic (ROC) curves were provided for each simulated population in Figures 11 and 12. Among the plots, it can be observed that the curve corresponding to SSA-NoCor is the lowest one. Therefore, as a binary classifier, SSA-NoCor performs worse than

the rest of the methods. The classifier ability of the models improved in the large populations, but SSA-NoCor still had poorer performance compared to other methods. The fact large data improves the resolution of GWAS is well documented in the literature.

Overall, we observed the size of the populations (e.g., small and large) did not change the outcome of our study, and we confirmed, using simulated populations with intricate structure, that EMMAX, GBLUP-GWAS, and ssGBLUP account for population structure. The equivalence between p-values obtained in EMMAX and GBLUP-GWAS has been analytically demonstrated (Bernal Rubio et al., 2016), although the former considers SNPs as fixed effects and the latter as random. Lu et al. (2018) extended this idea to single-step and implemented it with the addition of p-values for ssGWAS in the BLUPF90 software suite (Aguilar *et al*., 2019; Misztal *et al.,* 2015). This methodology was successfully applied to a beef cattle population with almost 2 million animals in the pedigree, 1 million birth weight records, and a little over 1400 genotyped sires (Aguilar *et al.,* 2019). In our study, we confirmed that ssGWAS can account for population structure as EMMAX or GBLUP- GWAS.

Recently, single-step was applied for predicting polygenic risk score in humans using phenotypes from related individuals that were not genotyped (Truong *et al.,* 2020). In this study, authors observed an increase in prediction accuracy when raw phenotypes of non-genotyped relatives were included in the model, which is only possible with single-step method. As the number of genotyped individuals in (Truong *et al.,* 2020) was 288k, the authors complained about the computing cost of single step, which is mainly due to the inverse of **G**. An efficient algorithm to compute $\mathbf{G}^{-1}$ without having to directly invert **G** –the Algorithm for Proven and Young (APY)– is also available (Misztal *et al.,* 2014). With the APY algorithm, animals are designated as core or noncore, and recursion are done on core animals, whereas predictions for noncore animals are functions of the information for core animals. This is possible because of the assumption that core animals carry all the information about the independent chromosome segments segregating in the population (Misztal et al., 2016). In addition, it was found (Pocrnic *et al.,* 2016) that the number of largest eigenvalues explaining 98% of the variance in **G** approaches the number of independent chromosome segments (Stam, 1980) and can be used as the number of core animals in APY. This algorithm enables the computation of genomic predictions for millions of genotyped individuals with much less memory usage and computing time. Indeed, a successful computation of genomic predictions for 13.5 million animals in the pedigree, of which 2.3 million were

genotyped, using the BLUPF90 software suite has recently been shown to be feasible (Tsuruta *et al.,* 2020). Although the computation of genomic predictions (GEBV), SNP effects, and variance explained by SNPs can be done efficiently with APY in ssGBLUP, the same does not apply to the computation of p-values in ssGWAS. This is because the formula for p-values (10) relies on the standard error of SNP effects (i.e., square root of prediction error variance), which is currently obtained based on the prediction error variance of GEBV. The latter requires the inverse of the left-hand-side of the single-step mixed model equations, and the computation of inverses of large matrices is extremely expensive. Therefore, the use of ssGWAS may be limited to samples of about 20k genotyped individuals, given the number of total animals in the pedigree is less than 500k. Approximating the prediction error variance of GEBV or SNPs directly may be a way to overcome this limitation, and research on the issue is currently undergoing.

Figure 5. Manhattan plots for the small population of fish using single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

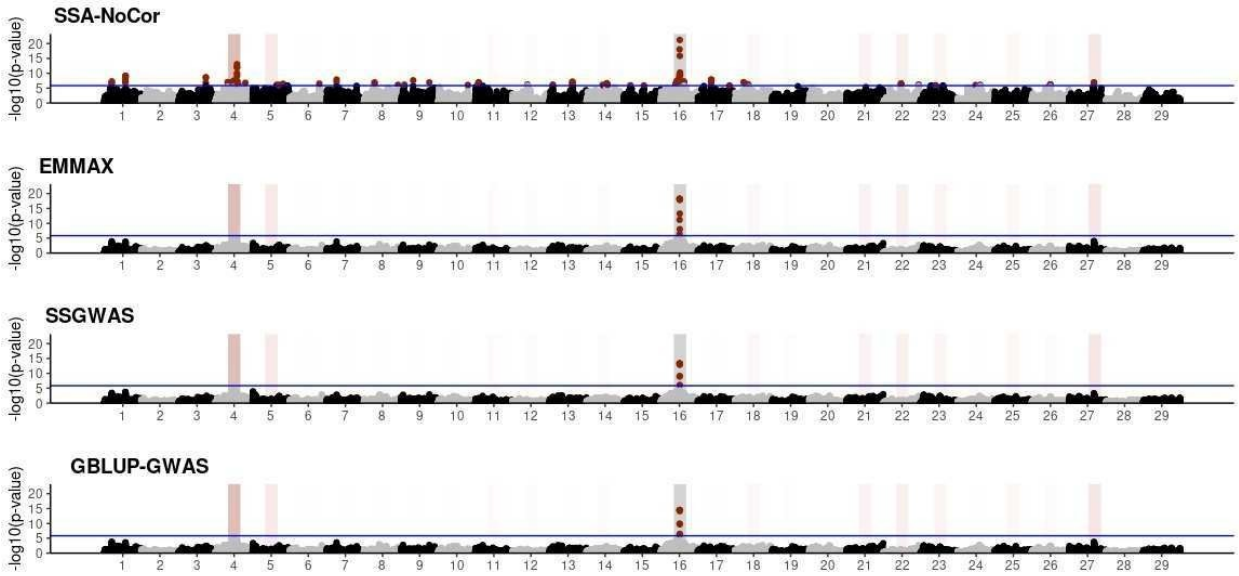


Figure 6. Manhattan plots for the small population of beef cattle using single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

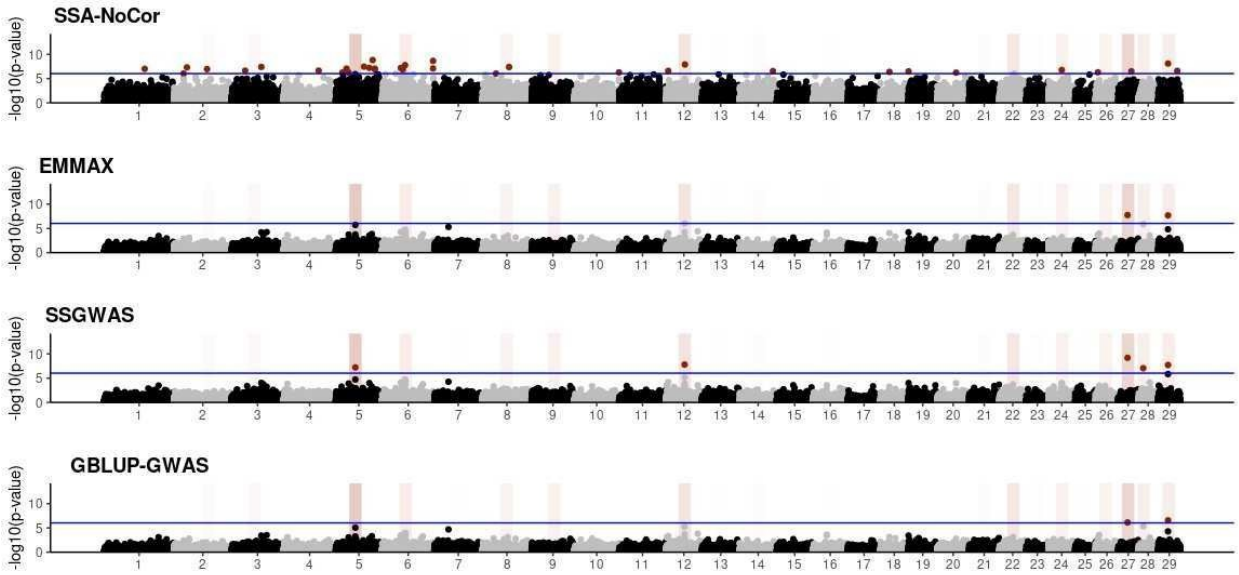


Figure 7. Manhattan plots for the small population of dairy cattle when sires had an average of ten daughters. The association methods used were single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

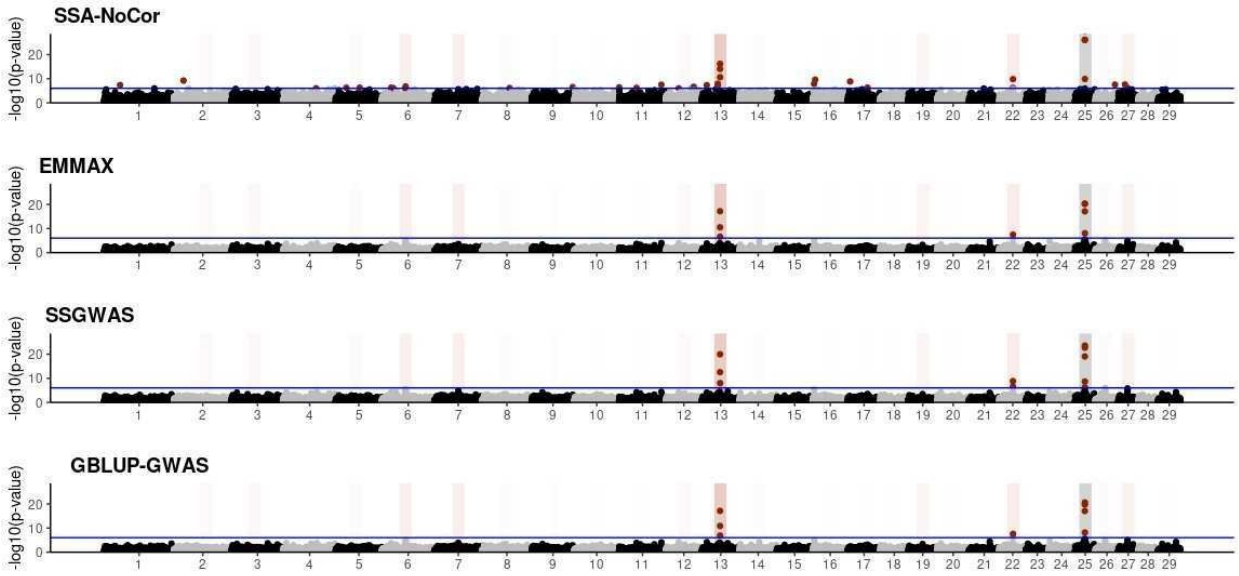


Figure 8. Manhattan plots for the large population of fish using single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

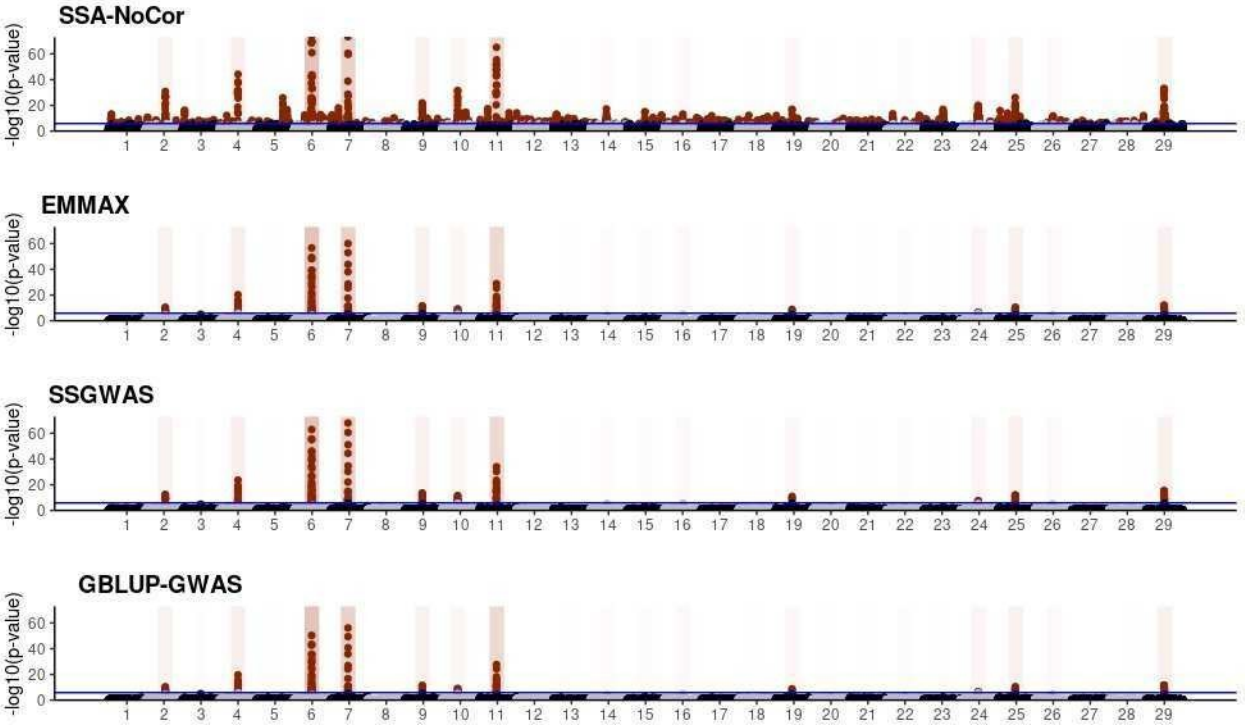


Figure 9. Manhattan plots for the large population of beef cattle using single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

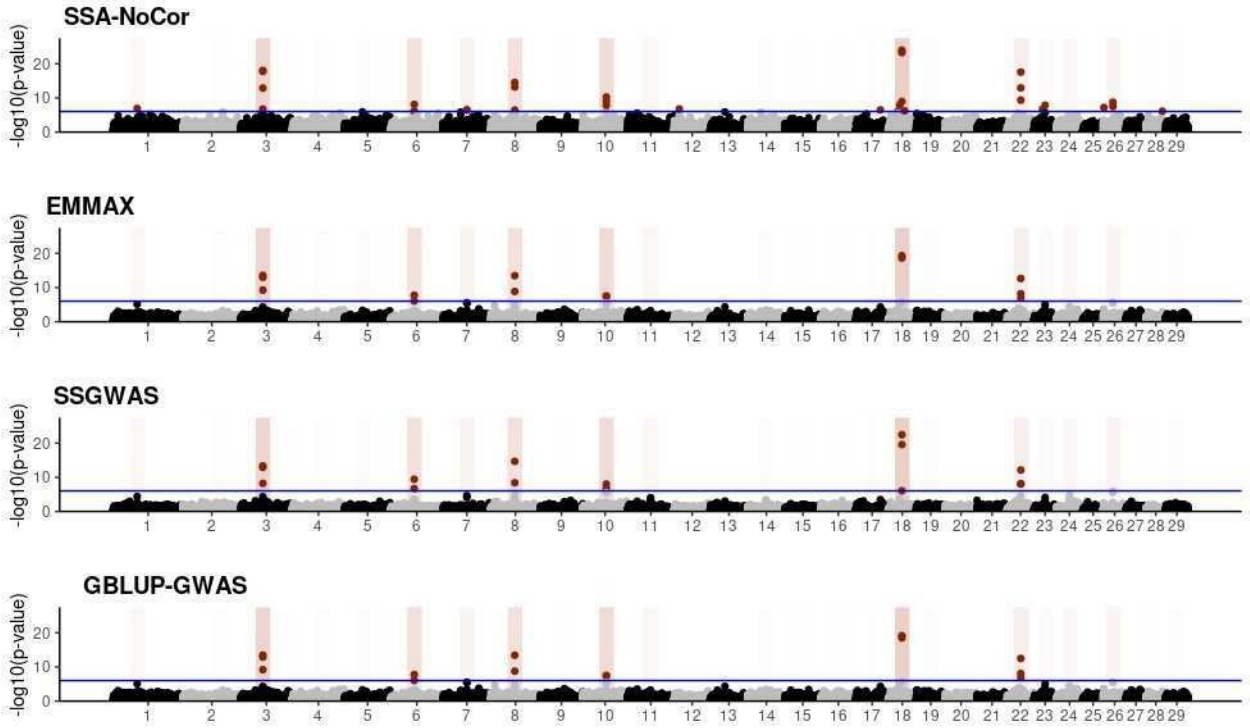


Figure 10. Manhattan plots for the large population of dairy cattle when sires had an average of ten daughters. The association methods used were single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing**.**

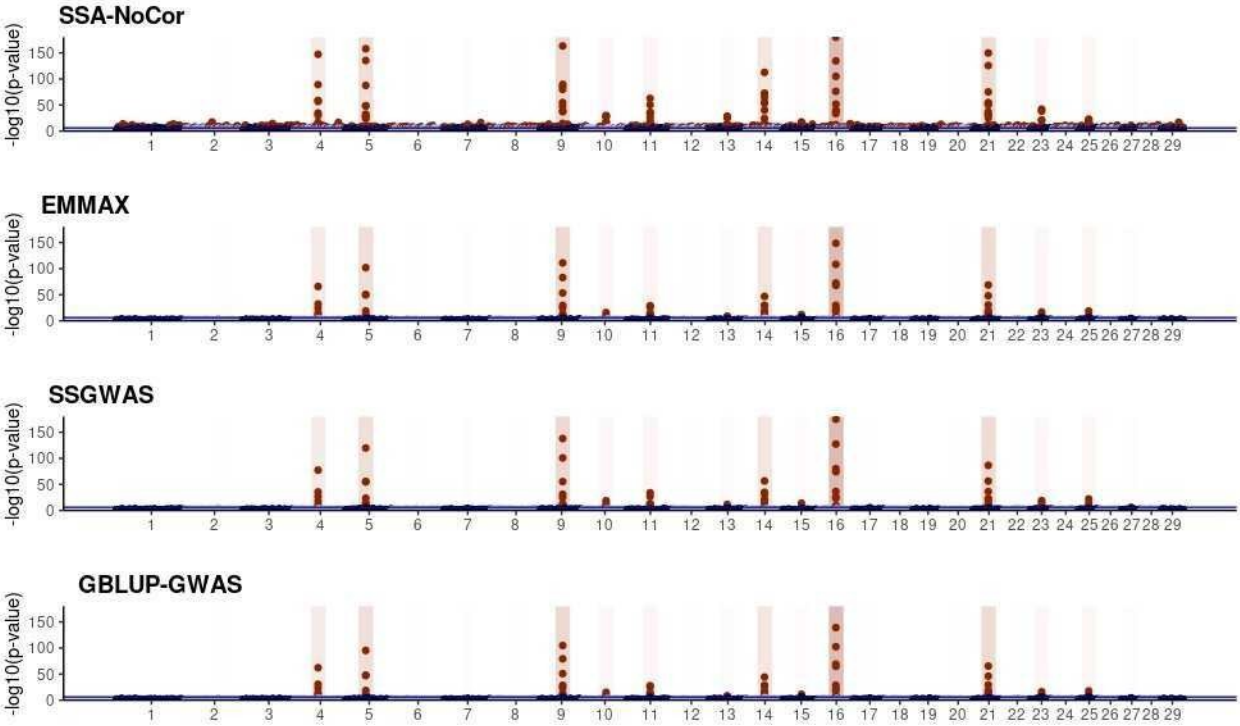


Figure 11. Receiver operating characteristic (ROC) curves for GWAS results for the small populations of fish (A), beef cattle (B), dairy cattle with ten daughters per sire. The association methods used were single- SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). The dashed line has slope equal to one and null intercept.

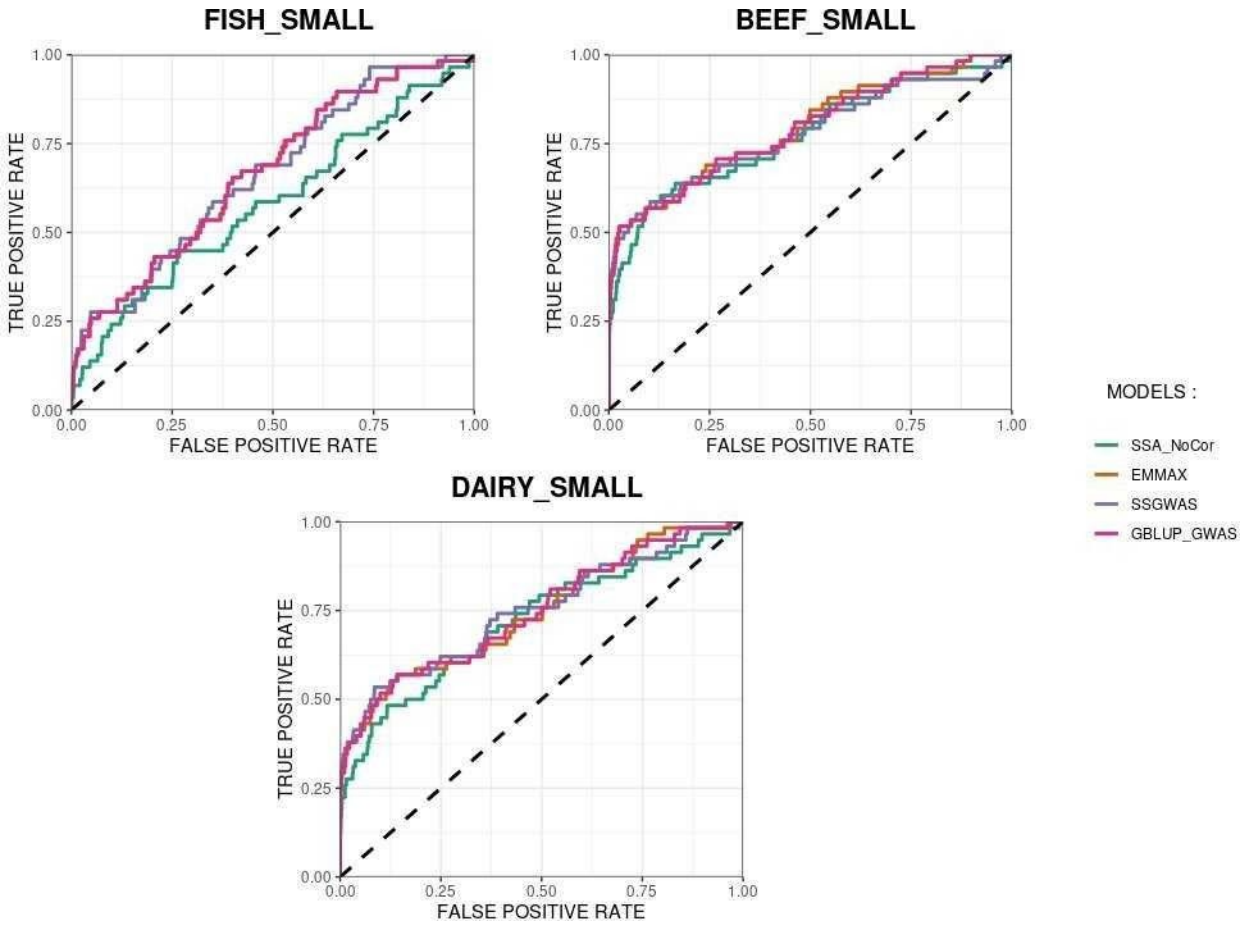
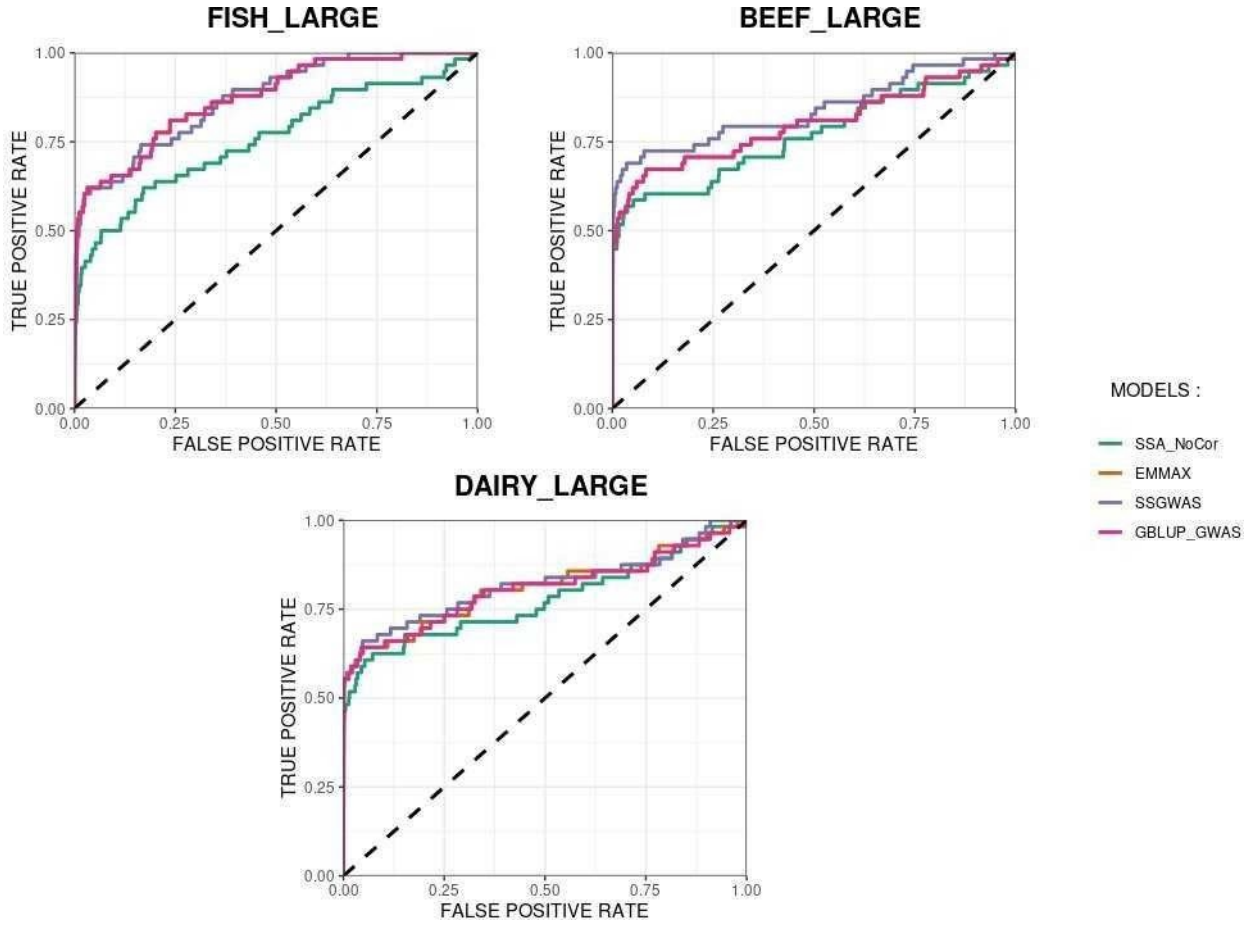


Figure 12. Receiver operating characteristic (ROC) curves for GWAS results for the large populations of fish, beef cattle, dairy cattle with ten daughters per sire. The association methods used were single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). The dashed line has slope equal to one and null intercept.



CONCLUSION

Genome-wide association studies in related populations require the correction for population structure to avoid false positive statistical associations between SNPs and trait phenotypes. Several classes of mixed linear models as EMMAX, GBLUP-GWAS, and ssGWAS can take care of this issue by fitting a random effect whose covariance matrix is proportional to a relationship matrix. We demonstrate the three methods did not significantly

differ across association studies in several simulated populations, regardless of if deregressed proofs or phenotypes from non-genotyped animals are used in the statistical analysis. Further studies are needed to investigate the repeatability of those results in real populations under complex models. Single-step GWAS accounts for population structure as EMMAX or GBLUP-GWAS and allows for the inclusion of phenotypes from non- genotyped relatives

ACKNOWLEDGEMENTS

This study was partially funded by Agriculture and Food Research Initiative Competitive Grant no. 2020-67015-31030 from the US Department of Agriculture's National Institute of Food and Agriculture. The authors would like to thank Andres Legarra for his helpful comments.

REFERENCES

Agresti, A. (2013) Categorical Data Analysis. 3rd Edition, John Wiley & Sons Inc., Hoboken.

Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., Misztal I. (2019). Frequentist p-values for large scale single step genome wide association, with an application to birth weight in American Angus cattle. *Genetics Selection Evolution* **51,** 28. doi: 10.1186/s12711-019-0469-3

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., Lawlor, T. J. (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. doi: 10.3168/jds.2009-2730

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y, et al., (2010). M. Genome- wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines Nature 465, 627—631. doi:10.21958/study:1

Balding, D. J. A tutorial on statistical methods for population association studies. (2006). Nature Reviews Genetics, 7, 781–791. doi: 10.1038/nrg1916

Begum, F., Ghosh, D., Tseng, G. C., and Feingold, E. (2012). Comprehensive literature review and statistical considerations for GWAS meta-analysis. Nucleic Acids Research 40, 3777–3784. doi: 10.1093/nar/gkr1255

Bernal Rubio, Y. L., Gualdrõn Duarte, J. L., Bates, R. O., Ernst, C. W., Nonneman, D., Rohrer, G. A. et al., (2016). Meta-analysis of genome-wide association from genomic prediction models. Animal Genetics 47, 36–48. doi: 10.1111/age.12378

Bian, Y., & Holland, J. B. Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. (2017). Heredity 118, 585–593. doi: 10.1038/hdy.2017.4

Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A. et al., (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–678. doi:10.1038/ng.2007.17.

Cappa, E. P., de Lima, B. M., da Silva-Junior, O. B., Garcia, C. C., Mansfield, S. D., & Grattapaglia, D. (2019) Improving genomic prediction of growth and wood traits in Eucalyptus using phenotypes from non-genotyped trees by single-step GBLUP. Plant Science 284, 9–15. doi: 10.1016/j.plantsci.2019.03.017

Christensen, O., & Lund, M. (2010). Genomic relationship matrix when some animals are not genotyped.

Genetics Selection Evolution 42, 1–8. doi: 10.1186/1297-9686-42-2

Cesarani, A., Pocrnic, I., Macciotta, N. P. P., Fragomeni, B. O., Misztal, I., & Lourenco, D. A. L. (2019). Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies. Journal of Animal Breeding and Genetics 136, 40–50. doi: 10.1111/jbg.12367

Dandine-Roulland, C., & Perdry, H. Manipulation of genetic data (SNPs). Computation of GRM and dominance matrix, LD, heritability with efficient algorithms for linear mixed model (AIREML). 46th European Mathematical Genetics Meeting (EMGM) 2018, Cagliari, Italy, April 18-20, 2018: Abstracts. Abstract retrieved from Human Heredity 83:1-29 (2018). doi: 10.1159/000488519 de Oliveira Silva, R., Bonvino Stafuzza, N., de Oliveira Fragomeni, B., de Camargo, G., Matos Ceacero T.,et al., (2017). Genome-Wide Association Study for Carcass Traits in an Experimental Nelore Cattle Population. PLoS One, 12, 1–14. doi: 10.1371/journal.pone.0169860

Falconer, D. S., and Trudy F. C. Mackay. Introduction to Quantitative Genetics (1996) Longman Essex, England

Finno, C. J., Aleman, M., Higgins, R. J., Madigan, J. E., and Bannasch, D. L. (2014). Risk of false positive genetic associations in complex traits with underlying population structure: A case study. Veterinary Journal 202, 543–549. doi: 10.1016/j.tvjl.2014.09.013

Garcia, A. L. S., Bosworth, B., Waldbieser, G., Misztal, I., Tsuruta, S., and Lourenco, D. A. L. (2018). Development of genomic predictions for harvest and carcass weight in channel catfish 06 Biological Sciences 0604 Genetics. Genetics Selection Evolution 50, 1–12. doi: 10.1186/s12711-018-0435-5

Gualdrón Duarte, J. L., Cantet, R. J. C., Bates, R. O., Ernst, C. W., Raney, N. E., and Steibel, J. P. (2014). Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. BMC Bioinformatics 15, 1–11 doi: 10.1186/1471-2105-15-246

Henderson, C. R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model.

Biometrics, 31, 423–447. doi: 10.2307/2529430

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nature Genetics 42, 348–354. doi: 10.1038/ng.548

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., and Eskin E. (2008). Efficient control of population structure in model organism association mapping. Genetics 178, 1709– 1723 doi: 10.1534/genetics.107.080101

Kennedy, B W, Quinton, M., and van Arendonk, J.A. Estimation of effects of single genes on quantitative traits. (1992). Journal of Animal Science 70, 2000–2012 doi: 10.2527/1992.7072000x

Kiser, J. N., Clancey, E., Moraes, J. G. N., Dalton, J., Burns, G. W., Spencer, T. E., & Neibergs, H. L. (2019). Identification of loci associated with conception rate in primiparous Holstein cows. BMC Genomics 20, 1–13. doi: 10.1186/s12864-019-6203-2

Legarra, A., Aguilar, I. and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. Journal of Dairy Science 92, 4656–4663. doi: 10.3168/jds.2009-2061.

Legarra, A., Christensen, O.F., Aguilar, I., and Misztal, I. (2014). Single Step, a general approach for genomic selection. Livestock Science 166, 54–65. doi: /10.1016/j.livsci.2014.04.029

Li, G., and Zhu, H. Genetic Studies: The Linear Mixed Models in Genome-wide Association Studies. (2013).

The Open Bioinformatics Journal 7, 27–33. doi: 10.2174/1875036201307010027

Lu, Y., Vandehaar, M. J., Spurlock, D. M., Weigel, K. A., Armentano, L. E., Connor, E. E., Coffey, et al. (2018). Genome-wide association analyses based on a multiple-trait approach for modeling feed efficiency. Journal of Dairy Science 101, 3140–3154. doi: 10.3168/jds.2017-13364

Misztal, I., Legarra, A., & Aguilar, I. (2014). Using recursion to compute the inverse of the genomic relationship matrix. Journal of Dairy Science 97, 3943–3952. doi: 10.3168/jds.2013-7752

Misztal, I., Lourenco, D., & Legarra, A. (2020). Current status of genomic evaluation. Journal of Animal Science 98, 1-14. doi: 10.1093/jas/skaa101

Misztal, I. (2016). Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. Genetics 202, 401-409. doi: 10.1534/genetics.115.182089

Misztal, I., Tsuruta, S., Lourenco, D.A.L., Aguilar, I., Legarra, A., and Vitezica, Z. 2014. Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch. php?media=blupf90_all1.pdf. (Accessed 1 December 2020.)

Pocrnic, I., Lourenco, A. L., Masuda, Y., Legarra, A., and Misztal, I. (2016). The dimensionality of genomic information and its effect on genomic prediction. Genetics 203, 573-581. doi: 10.1534/genetics.116.187013

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38, 904–909. doi: 10.1038/ng1847

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. Science 273, 1516-1517 doi: 10.1126/science.273.5281.1516

Sargolzaei, M., and Schenkel (2009). F. S. QMSim: a large-scale genome simulator for livestock.

Bioinformatics 25, 680–681. doi: 10.1093/bioinformatics/btp045

Sonesson, A. K., and Meuwissen, T.H. (2000). Mating schemes for optimum contribution selection with constrained rates of inbreeding. Genetics Selection Evolution 32, 231-248. doi: 10.1186/1297-9686-32- 3-231

Stam, P. The distribution of the fraction of the genome identical by descent in finite random mating populations. (1980). Genetics Research 35, 131-155. doi: 10.1017/S0016672300014002

Spencer, C. C. A., Su, Z., Donnelly, P., and Marchini, J. (2009). Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. PLoS Genetics 5, e1000477. doi: 10.1371/journal.pgen.1000477

Sul, J. H., Martin, L. S., and Eskin, E. Population structure in genetic studies: (2018). Confounding factors and mixed models. PLoS Genetics 14, 1–22. doi: 10.1371/journal.pgen.1007309

Tiezzi, F., Maltecca, C., Cecchinato, A., Penasa, M., & Bittante, G. (2012). Genetic parameters for fertility of dairy heifers and cows at different parities and relationships with production traits in first lactation. Journal of Dairy Science 95, 7355–7362. doi: 10.3168/jds.2012-5775

Toosi, A., Fernando, R. L., and Dekkers, J. C. M. (2018). Genome-wide mapping of quantitative trait loci in admixed populations using mixed linear model and Bayesian multiple regression analysis. Genetics Selection Evolution 50, 1–13. doi: 10.1186/s12711-018-0402-1

Truong, B., Zhou, X., Shin, J., Li, J., van der Werf, J. H. J., et al., (2020). Efficient polygenic risk scores for biobank scale data by exploiting phenotypes from inferred relatives. Nature Communications 11, 3074 doi: 10.1038/s41467-020-16829-x

Tsuruta, S., Lawlor, T. J., Lourenco, D. A. L., Misztal, I. (2020) Bias in genomic predictions by mating practices for linear type traits in a large-scale genomic evaluation. Journal of Dairy Science 104, in press. doi: 10.3168/jds.2020-18668.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. Journal of Dairy Science 91, 4414–4423. doi: 10.3168/jds.2007-0980

VanRaden, P. M., Sanders, A. H., Tooker, M. E., Miller, R. H., Norman, H. O., Kuhn, et al., (2004). Development of a national genetic evaluation for cow fertility. Journal of Dairy Science 87, 2285–2292.doi: 10.3168/jds.S0022-0302(04)70049-1

VanRaden, P. M., and Sullivan, P.G. (2010). International genomic evaluation methods for dairy cattle.

Genetics Selection Evolution 42, 1–9. doi: 10.1186/1297-9686-42-7

Vitezica, Z. G., Aguilar, I., Misztal, I., and Legarra, A. (2011). Bias in genomic predictions for populations under selection. Genetics Research 93, 357–366. doi: 10.1017/S001667231100022X

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery.

American Journal of Human Genetics 90, 7–24. doi: 10.1016/j.ajhg.2011.11.02

Wang, M., Jiang, N., Jia, T., Leach, L., Cockram, J., Waugh, L., et al., (2012). Genome-wide association mapping of agronomic and morphologic traits in highly structured populations

of barley cultivars. Theoretical and Applied Genetics 124, 233–246. doi: 10.1007/s00122-011-1697-2

Wiggans, G. R., T. a Cooper, P. M. VanRaden, and Cole, J. B. (2011). Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. Journal of Dairy Science 94:6188–6193. doi: 10.3168/jds.2011-4481

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J., et al., (2011). Genome partitioning of genetic variation for complex traits using common SNPs. Nature Genetics 43, 519–525. doi: 10.1038/ng.823

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. and Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nature Genetics 46, 100–106. doi: 10.1038/ng.2876

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al., (2006). A unified mixed- model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38, 203–208. doi: 10.1038/ng1702

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al., (2010). Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42, 355–360. doi:10.1038/ng.546

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies.

Nature Genetics 44, 821-824. doi: 10.1038/ng.2310 Zhou, X. (2016). GEMMA User Manual. 1–27

11. GWAS OF BEEF TRAITS IN A LOCAL ALPINE CATTLE BREED REVEALS DIVERSITY OF PATHWAYS INVOLVED AND VARIABILITY WITHIN TRAITS COLLECTED AT DIFFERENT TIMES

STATUS PUBBLICATED ON FORNTIERS IN GENETIC
https://doi.org/10.3389/fgene.2021.746665

GWAS of beef traits in a local alpine cattle breed reveals diversity of pathways involved and variability within traits collected at different times

Enrico Mancin[1], Beniamino Tuliozi[1], Sara Pegolo[1], Cristina Sartori[1], Roberto Mantovani[1]

ABSTRACT

Knowledge of the genetic architecture of key growth and beef traits in livestock species has greatly improved worldwide thanks to genome-wide association studies (GWAS), which allow to link target phenotypes to Single Nucleotide Polymorphisms (SNPs) across the genome. Local dual- purpose breeds have rarely been the focus of such studies; recently, however, their value as a possible alternative to intensively farmed breeds has become clear, especially for their greater adaptability to environmental change and potential for survival in less productive areas. We performed single-step GWAS and post-GWAS analysis for body weight (BW), average daily gain (ADG), carcass fleshiness (CF) and dressing percentage (DP) in 1690 individuals of local alpine cattle breed, Rendena. This breed is typical of alpine pastures, with a marked dual-purpose attitude and good genetic diversity. Moreover, we considered two of the target phenotypes (BW and ADG) at different times in the individuals' life, a potentially important aspect in the study of the traits' genetic architecture. We identified 8 significant and 47 suggestively associated SNPs, located in 14 autosomal chromosomes (BTA). Among the strongest signals, 3 significant and 16 suggestive SNPs were associated with ADG and were located on BTA10 (50-60 Mb), while the hotspot associated with CF and DP was on BTA18 (55-62 MB). Among the significant SNPs some were mapped within genes, such as *SLC12A1*, *CGNL1*, *PRTG* (ADG), *LOC513941* (CF), *NLRP2* (CF and DP),

CDC155 (DP). Pathway analysis showed great diversity in the biological pathways linked to the different traits; several were associated with neurogenesis and synaptic transmission, but actin- related and transmembrane transport pathways were also represented. Time-stratification highlighted how the genetic architectures of the same traits were markedly different between different ages. The results from our GWAS of beef traits in Rendena led to the detection of a variety of genes both well-known and novel. We argue that our results show that expanding genomic research to local breeds can reveal hitherto undetected genetic architectures in

livestock worldwide. This could greatly help efforts to map genomic complexity of the traits of interest and to makeappropriate breeding decisions.

INTRODUCTION

Genome-wide association is a powerful analysis that allows to identify genomic regions associated with phenotype variations in a target population to understand better the genetic architecture of thephenotype (Begum et al., 2012); such analysis has proved to be invaluable in the study of the genetic architecture of livestock species traits, especially cattle (Schmid and Bennewitz, 2017).

Most of the target traits in livestock are polygenic phenotypes (de Oliveira Silva et al., 2017), which are suitable for investigation with robust GWAS. However, the GWAS is only the start of the investigation of the target traits genetic architecture (Atwell et al., 2010). Weaker signals that would be missed by GWAS analysis can be identified and described via pathways enrichment analysis, under the assumption that these signals are related to genes involved in complex pathways and biological processes (Buitenhuis et al., 2014; Pegolo et al., 2020). In beef cattle, traits such as growth or carcass conformation are critical to the profitability of meat production since greater growth means a shorter fattening period, and more conformed animals have higher economic value (Samorè et al., 2016). GWAS analysis in different species highlighted the strongly polygenic nature of these traits (Mateescu et al., 2017; Huang et al., 2018; Falker-Gieske et al., 2019; Gershoni et al.,2021).

In recent years, many studies have proposed more advanced approaches to investigate these phenotypes, such as the inclusion of whole genome sequences (Mao et al., 2016) or the analysis of growth traits in a longitudinal perspective (Yin and König, 2019). This latter approach has been scarcely used in beef cattle breeding (Yin and König, 2019; Gershoni et al., 2021), but there are dramatic differences in the functional elements involved in determining morphological traits at different ages (Helgeland et al., 2019): these differences could be investigated by separate analyses of the same trait collected at various ages. Investigations on beef traits (Mudadu et al., 2016) have been extensively performed in cattle, but most studies have regarded few cosmopolitan, specialized breeds. Dual-purpose breeds, which consist of local populations apart from a few exceptions (such as Simmental cattle), have rarely been the target of GWAS.

Local breeds are genetically more diverse than the cosmopolitan ones and have generally better health parameters and fitness due to a much-reduced specialization (Biscarini et al., 2015). Also, the negative genetic correlations occurring between dairy and beef traits make the genetic improvement of both aptitudes in dual- purpose populations far from its optimum (Frigo et al., 2013; Mazza et al., 2016; Sartori et al., 2018). Moreover, such breeds often present unique characteristics that allow them to adapt to harsher conditions (Krupová et al., 2016; Sutera et al., 2021) and better respond to environmental shifts or challenges (Biscarini et al., 2015). Thus, these dual-purpose local breeds represent an unexploited source of diversity for the animal breeding sector and a rare opportunity to conduct GWAS on key economic traits that have not been under excessive specialization.

Rendena is an autochthonous breed from Alpine regions of North-East of Italy with a dual-purpose aptitude for meat and milk still maintained through the current selection scheme, assigning 65% of the economic weight to milk and 35% to meat (Guzzo et al., 2019; for further details on the selection scheme see Mantovani et al., 1997; and Supplementary Material, Figure S1).

The dual-purpose aptitude also allows to counteract inbreeding erosion and maintain good genetic variability despite the small population size (the current number of animals is around 7,000 of which 4,000 are cows). Rendena also presents good fertility and longevity parameters and excellent adaptability to local environments, ranging from plains to Alpine pastures (Ovaska and Soini, 2017; Guzzo et al., 2018). As in various other local breeds, genomic information of Rendena has started to be available just recently, after implementing a routine activity of genotyping. This information might allow identifying and describing genes and functional pathways involved in the genomic architecture of traits of economic or functional interest (Senczuk et al., 2020). Moreover, as genomic selection has just been implemented in Rendena (Mancin et al., 2021a), investigating these traits could also be helpful to increase the prediction accuracy (see Tiezzi and Maltecca, 2015).

In this study, we performed a single-step GWAS and pathway analysis in Rendena cattle to investigate the genetic architecture of growth and carcass conformation traits, i.e., body weight, average daily gain, in vivo dressing percentage, and in vivo fleshiness (SEUROP grade).

Additionally, body weight and average daily gain were analyzed using records taken at different ages, to study possible temporal variation in the genetic architecture of growth at the early stages.

MATERIALS AND METHODS

Animals and Phenotypes:

All phenotypic records were collected at the performance test (PT) station of the National Breeders Association of Rendena cattle - ANARE, Trento Italy (www.ANARE.it). All phenotypes belonged to young (on average of one month of age) candidate bulls. About 60 young bulls are tested every year at the PT station for a total period of 11 months, following the criteria reported in Mantovani et al. (1997). Records have been collected since 1985, when PT started, until present times. The phenotypes collected during the PT are body weight (BW), average daily gain (ADG), carcass fleshiness (CF) and dressing percentage (DP). Both CF and DP are evaluated in vivo by 3 skilled operators at the end of the PT period and averaged to obtain the final score. The CF evaluation applies the same scores of post-mortem carcass appraisal established by the European Union Council (SEUROP), where the middle class (R) is equal to 100 points and other classes (upper or lower classes) correspond to 10-points-variations. Furthermore, the evaluation also considers sub- classes (e.g., R+ and R- for the middle class) that are spaced 3.33 points from the class score. DP is a visual prediction of the post-mortem measure of DP: the operator makes a visual appraisal of the individual at the end of the performance test, offering an estimate of the expected DP – i.e., conformation – at slaughter (Mantovani et al., 1997). Average daily gain (ADG) is calculated as the linear regression of weight (BW) on age. For this study, ADG and BW were collected at different stages of PT. ADG has been divided into ADG_i and ADG_f: ADG_i covers the daily gain of the first half of the testing period (since entering the PT station until the 6th month), while ADG_f covers the daily gain of the second half (from the 6th month to the end of the period). ADG covering the entire PT test was labeled as ADG_tot. BW was split along the same timeline as ADG: body weight at the entrance to the station (BW_i), at six months (BW_m) and at the end of PT (BW_f). Data cleaning consisted of removing animals with a regression of weight on age showing a coefficient of determination below 0.9 (for further details, see Guzzo et al., 2019).

Genomic data and quality control

The biological material of the animals chosen for the genotyping resulted from salivary swab, hair (at least 30 bulbs), or ear tissue from biopsy brand, collected by ANARE on females and young candidate bulls at PT, as well as from semen of proven bulls, already subjected in the past to PT and progeny test for milk and to a large extent now eliminated. The Bovine 150K

Array GGPv3 Bead Chip (HD, 138,974 SNPs), and Illumina Bovine LD GGPv3 (LD, 26,497 SNPs), were used for genotyping (Illumina, Illumina Inc., San Diego, CA, USA). The overlapping between the two panels is about 60%. The HD platform was used for 554 young bulls, while 1,416 individuals (174 males and 1,242 females) were genotyped with LD chips. To achieve a reliable genomic imputation accuracy, the 174 males were animals with at least one parent and one half-sib genotyped with HD chips. The genotyped females were individuals with a kinship of at least 0.2 with phenotyped animals. Before proceeding with imputation, we performed a preliminary quality control removing SNPs with a minor allele frequency (MAF) < 0.01 and call rate lower than 0.90, using Plink program (Purcell et al., 2007). Only the 29 autosomal chromosomes (BTA) were used for association, and progeny conflicts were fixed using the seekparentsf90 program (Aguilar et al., 2018).

AlphaImpute2 was used for imputation (Whalen and Hickey, 2020), as it combines a population imputation algorithm (Positional Burrows Wheeler Transform) with pedigree-based imputation (iterative peeling); we used the same parameters as in Mancin et al. (2021a). The accuracy of the imputations was roughly estimated as a correlation between true and imputed SNPs. To this aim, ten rounds of cross-validation were performed: in each round the overlapping SNPs between the two panels were removed in ten animals and then imputed using the HD panel from young bulls as reference population (Supplementary Material, Table S1). Subsequently, the correlation between the true and the imputed genotypes was calculated on these animals.

After imputation, we performed a second genomic quality control with the preGSf90 program (Aguilar et al., 2018): the SNPs with MAF lower than 0.05 and SNPs that deviated too much for the expected value of heterozygosis (i.e., Hardy-Weinberg Equilibrium) were removed. In accordance

Single step genome-wide association (ssGWAS)

Single step genome-wide association (ssGWAS) models were used to estimate allele substitution effect. In ssGWAS, the estimation of allele substitution effects was obtained from a linear transformation of the BLUP of breeding value under ssGBLUP model (Aguilar et al., 2019). Mancin et al. (2021b) showed the advantages of this method in terms of QTL detection and control of populations structure over two-step methods in which de-regression

of breeding value as pseudo phenotype is required. This issue is particularly evident in the presence of unbalanced data (i.e., sex-limited traits). In fact, the ssGWAS allows the use of both male and female genomes even when analyzing a phenotype collected only in individuals of one sex.

The ssGBLUP model used in this analysis, written in matrix form, is the following:

$$\begin{bmatrix} X'\mathrm{X} & X'\mathrm{Z} \\ \mathrm{Z}'\mathrm{X} & \mathrm{Z}'\mathrm{Z}+\mathrm{H}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ \mathrm{Z}'y \end{bmatrix} \quad [1]$$

Where phenotypes are included in vector y, X is the incidence matrix of fixed effects (group of contemporaries, cow parity class and months of birth), b is the vector of these effects. The contemporary group has 147 levels, with each level consisting of bulls grouped together at the Performance Test because homogeneous by age (i.e., born within 1 month of each other; 82

animals per group on average, minimum 5 and maximum 142). The parity order of cow has four classes (first parity; second parity; third to seventh parity; above the eighth parity), and the classes of months of birth correspond to the single months, as in Guzzo et al. (2019).

Z represents the incident matrix that relates the random genetic additive effects to the phenotype, with effects represented by vector a. The vector of random residual error (e) has a normal distribution $N(0, I\sigma_e^2)$, where σ_e^2 is the residual variance. In the ssGBLUP vector of additive genetic effects is distributed as $N(0, H\sigma_a^2)$, where σ_a^2 is the additive genetic variance and H is the (co)variances structure which combines pedigree and genomic relationships (Aguilar et al., 2010). Its inverse, used in equation [1] is described as:

$$\mathrm{H}^{-1} = \mathrm{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathrm{G}^{-1} - \mathrm{A}_{22}^{-1} \end{bmatrix} \quad [2]$$

where A^{-1} and A_{22}^{-1} are the inverse of the pedigree kindship matrix respectively for all animals and for only genotyped animals. Since the frequencies of current genotyped population are used to center G and pedigree and genomic matrices have different bases, G

was adjusted so the average diagonal and off-diagonal matches the averages of A^{22}. Pedigree kinship (sub) matrix was estimated tracing back the pedigree up to 7 generations, i.e., 6,644 animals. G matrix was built using the methods proposed by VanRaden (2008), as follows:

$$G_0 = \frac{\mathrm{MM}'}{2\sum p_i(1-p_i)} \quad [3]$$

where M is a matrix of SNP content centered by twice the current allele frequencies, and p_i is the allele frequency for the ith SNP (VanRaden, 2008).

Additionally, to avoid singularity problems, the final G was computed as

$$\mathrm{G} = \lambda \mathrm{G}_0 + \beta \mathrm{I} \quad [4]$$

Where G is the matrix present in the equations [2], I is an identity matrix of the same dimensions, λ and β are two weighting coefficients, with λ=0.99 and β=0.01. These values were chosen due to their influence on the power of signal detection of the GWAS, and because they resulted in inflation close to optimum values. In addition, G was adjusted to a better blending with diagonal and off-diagonal of A^{22} as described in Vitezica et al. (2011):

$$\delta = 1 - \frac{0.5}{n^2}\left(\sum_i \sum_j \mathrm{A}_{22(\mathrm{i,j})} - \sum_i \sum_j \mathrm{G}_{\mathrm{i,j}}\right) \quad [5]$$

Then, the vector of estimated breeding values was obtained as:

$$\hat{g} = \lambda\delta \frac{1}{2\sum pq} \mathrm{M}'\mathrm{G}^{-1}\hat{a}_{22} \quad [6].$$

Where $\hat{a}_{22}$is the vector of estimated breeding values of genotyped animals. The prediction error variances $\hat{g}$, necessary to calculate the p-values, were calculated following Gualdrón Duarte et al. (2014) and computed as in Aguilar et al. (2019), where:

$$Var(\hat{g}) = Var\left(\lambda\delta \frac{1}{2\sum pq} \mathrm{M}'\mathrm{G}^{-1}\hat{a}_{22}\right) \quad [7]$$

$$Var(\hat{g}) = \lambda\delta \frac{1}{2\sum pq} \mathrm{M}'\mathrm{G}^{-1} Var(\hat{a}_{22}) \mathrm{G}^{-1}\mathrm{M}\, \lambda\delta \frac{1}{2\sum pq} \quad [8]$$

Since $Var(\hat{a}_{22})$ is equal to PEV ($\hat{a}_{22}$) – Var(a_{22}); thus $Var(\hat{a}_{22}) = \mathrm{G}\hat{\sigma}_a^2 - \mathrm{C}^{22}$. It follows that formula [8] becomes:

$$Var(\hat{g}) = \lambda\delta \frac{1}{2\sum pq} \mathrm{M}'\mathrm{G}^{-1}(\mathrm{G}\hat{\sigma}_a^2 - \mathrm{C}^{22})\mathrm{G}^{-1}\mathrm{M}\, \lambda\delta \frac{1}{2\sum pq} \text{ [9]}$$

C^{22} is a submatrix of C belonging to the genotyped animals and represents the prediction error variances of $\hat{a}_{22}$. The p-values are then calculated as

$$p - value_i = 2\left(1 - \Phi\left(\left|\frac{\hat{g}_i}{sd(\hat{g}_i)}\right|\right)\right) \text{[10]}$$

Where $\hat{g}_i$ is the allele substitution effect of SNP i and $sd(\hat{g}_i)$ represents the square root of [9], Φ (·) is the cumulative density function (CDF) of the normal distribution. Two thresholds were used for the association tests: a genome-wide 5% significant level of -log10(p) = 5.55 (0.05/17,766) and a suggestive association with -log10(p) = 4.29 (0.1/17,766). These are the thresholds corrected for multiple tests i.e., $\frac{p}{n}$where p is the probability level of significance and n is the corresponding number of independent SNPs (n = 17,766) calculated using the 'poolR' R package (https://cran.r-project.org/web/packages/poolr; R Core Team, 2021), according to Li and Ji (2005). The number of independent tests was calculated based on the number of eigenvalues. Instead of the standard approach of Cheverud (2001), we used the approach by Li and Ji (2005), a function that decomposes the eigenvalues in the integral part (Effective Number Independent Test) and the nonintegral part.

The (co)variance components have been estimated with REML using Average-Information algorithm (Gilmour et al., 1995). Approximate standard error of (co)variance components has also been estimated through Monte Carlo sampling as in Houle and Meyer (2015), in which standard deviations were calculated from Monte Carlo chains sampled from multinormal distribution with covariance being the inverse of the Average Information Matrix and the estimated variances as the expectation. Then the heritability for the 3 phenotypes was calculated under single trait models as in equation [1]. Heritability was calculated as:

$h^2 = \frac{\sigma_a^2}{(\sigma_a^2+\sigma_e^2)}$; where σ_a^2 and σ_e^2 are, respectively, the additive genetic and the residual variances.

Genetic and phenotypic correlations were estimated with bi-traits models, which are equivalent to equation [1] except for the animal additive genetic and residual variance, assumed to follow a multivariate normal distribution with mean 0 and variances G ⊗ H, and R ⊗ I, where

$$\mathrm{G} = \begin{vmatrix} \sigma^2_{\mathrm{a1}} & \sigma_{\mathrm{a1a2}} \\ \sigma_{\mathrm{a1a2}} & \sigma^2_{\mathrm{a2}} \end{vmatrix}; \mathrm{R} \begin{vmatrix} \sigma^2_{\mathrm{e1}} & \sigma_{\mathrm{e1e2}} \\ \sigma_{\mathrm{e2e1}} & \sigma^2_{\mathrm{e2}} \end{vmatrix}; \quad [6]$$

where G is the matrix of additive genetic (co)variances σ^2_{a1}, σ^2_{a2}, σ_{a1a2} of traits 1 and 2, R the matrix of residual (co)variances σ^2_{e1}, σ^2_{e2} and σ_{e1e2} of traits 1 and 2. The correlation was estimated as: $cov = \frac{\sigma_{i1i2}}{(\sigma_{i,1}\ \sigma_{i,2})}$ where *i* stands for the genetic and phenotypic correlation; 1 and 2 refer to the different performance test traits, and $\sigma_{i1i2}$is the covariance between traits 1 and traits 2, off diagonal of [6]. For phenotypic (co)variance, we mean the sum of the genetic and the phenotypic (co)variances. Traits that do not include zero in their correlations Higher Posterior Density Interval (HPD) were declared significantly correlated. All the genomic analyses were carried out with BLUPF90 family software (Aguilar et al., 2018) following the procedure described in Lourenco et al. (2020). Manhattan plots were drawn using 'ggplot' R package (Wickham, 2016), as were the LD graphs.

1.6 Pathway analysis

Pathway's enrichment analysis was conducted to identify which biological pathways and functional elements were enriched for the investigated traits. From GWAS results, we selected SNPs with nominal P-values of < 0.01 which were mapped to genes based on a distance of 15 kb from the coding region using the 'biomaRt' R package (Drost and Paszkowski, 2017) and Bos taurus UMD3.1 assembly as in Pegolo et al. (2020). Functional enrichment analysis was carried out on the list of significant genes using the Cytoscape plugin ClueGo (Bindea et al., 2009). As functional categories, we used cellular component, biological process, and molecular functions within the Gene Ontology (GO,

http://geneontology.org) database and the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://genome.jp/kegg/). The Benjamini-Hochberg correction was applied to declare significant pathways: only pathways showing FDR < 0.05 were retained. The minimum number of genes in the pathway was set to 3; the minimum percentage of genes present in the pathway was set to 4%. To simplify the redundance of GO terms we provide figures with similar terms grouped based on their semantic similarity using the R packages 'rrvgo' (Sayols, 2020). In addition, we investigated if the candidate regions declared as significant by our GWAS overlapped with QTL in animal QTLdb, identified with R package 'GALLO' (Fonseca et al., 2021).

RESULTS AND DISCUSSION

Heritability and genetic correlations

Descriptive statistics after data editing of the phenotypes are shown in Table 1. Phenotypic and genetic correlations and the heritability (h^2) for the analyzed traits are reported in Table 2. Body weight traits presented an average value of h^2 lower than other traits: BW_i showed the lower heritability (0.130), while BW_m and BW_f had heritability of 0.220. In fact, as reported in literature, a large discrepancy of values has been observed for heritability of body weights, and generally, traits similar to birth weight or weaning weight have a slightly lower heritability than weight measured in more advanced stages (Yin and König, 2018). Average daily gain (ADG_tot) presented an intermediate heritability of 0.322 partitioned into 0.164 and 0.220 for ADG in the first and last period. As for body weight, ADG presents lower h^2 in first stages of the performance test, and h^2 values agree with what has been found in the literature (Yin and König, 2018). The highest heritabilities were found for the traits related to the carcass conformation, with a value of 0.45 and 0.47 respectively for CF and DP, close to what was observed in other local dual-purpose or beef cattle (Albera et al., 2001; Sbarra et al., 2013; Mancin et al., 2021c). These traits also appeared highly genetic correlated. All ADG traits were moderately genetically correlated with them, with a value of 0.5 on average. On the contrary, body weight measured at the beginning of the performance test was not significantly correlated with CF and DP. Interestingly, the weights measured in more advanced periods showed an increase of genetic correlation with a value close to 0.7. Body weight and ADG also presented a strong genetic correlation with body

weight traits, especially for the traits measured at the final stages of the performance test. In terms of genetic correlations, the results agree with what was found in other local dual-purpose or beef breeds (Veselá et al., 2011; Filipčík et al., 2020). Phenotypic correlation followed the same trends of genetic correlation but with a lower magnitude (Table 2, under diagonal).

Table 1: Summary statistics for phenotypic data of animals with both genotypic and phenotypic information (n = 689).

Traits	Mean	SD	Min.	Max.
BW_i (kg)	65.72	14.64	37	139.0
BW_m (kg)	183.40	30.53	83	317.0
BW_f (kg)	376.20	43.60	203	576.0
ADG_i (g/d)	939.20	167.90	138	1388
ADG_f (g/d)	1082	157.30	365	1756
ADG_tot (g/d)	1024	124.2	474	1562
CF (score)	99.05	3.80	80	111
DP (score)	54.18	0.94	50	57

BW_i, body weight at the entrance at performance test stations; BW_m, body weight at six months; BW_f, at the end of performance test; ADG_i, average daily gains covering the first half of the period (since entering into the PT station until the 6th month); ADG_f, average daily gain covering the daily gain of the second half (since the 6th month to the end of the period), ADG_tot average daily gain covering the entire period; DP, Dressing Percentage; CF, Carcass Fleshiness; [c]SD Standard deviation, [b]Min minimum, [c]Max maximum

Table 2. Mean of genetic (over diagonal) and phenotypic (under diagonal) correlations, and heritability (diagonal) with the respective standard deviations in target traits in Rendena population, estimated under ssGBLUP models. (^{NS}) stands for non-significant correlations.

	BW_i	BW_m	BW_f	ADG_i	ADG_f	ADG_tot	CF	DP
BW_i	0.13 ± 0.08	0.99 ± 0.17	0.80 ± 0.10	0.52 ± 0.96	0.44 ± 0.85	0.50 ± 0.60^{NS}	0.33 ± 0.71	0.53 ± 0.80
BW_m	0.41 ± 0.05	0.22 ± 0.09	0.87 ± 0.11	0.81 ± 0.41	0.68 ± 0.36	0.78 ± 0.59	0.69 ± 0.58	0.73 ± 0.44
BW_f	0.29 ± 0.07	0.79 ± 0.03	0.22 ± 0.09	0.78 ± 0.43	0.97 ± 0.17	0.97 ± 0.28	0.62 ± 0.21	0.63 ± 0.23
ADG_i	0.17 ± 0.07	0.77 ± 0.03	0.86 ± 0.02	0.16 ± 0.10	0.64 ± 0.12	0.81 ± 0.21	0.62 ± 0.43	0.67 ± 0.25
ADG_f	-0.04 ± 0.08	0.09 ± 0.08	0.68 ± 0.04	0.14 ± 0.08	0.23 ± 0.08	0.97 ± 0.1	0.43 ± 0.23	0.47 ± 0.22
ADG_tot	0.11 ± 0.08	0.47 ± 0.06	0.84 ± 0.02	0.68 ± 0.04	0.80 ± 0.03	0.32 ± 0.09	0.55 ± 0.16	0.6 ± 0.15
CF	0.14 ± 0.08	0.4 ± 0.07	0.49 ± 0.09	0.3 ± 0.08	0.37 ± 0.08	0.42 ± 0.08	0.46 ±0.09	0.98 ± 0.02
DP	0.05 ± 0.09	0.35 ± 0.08	0.51 ± 0.07	0.26 ± 0.09	0.38 ± 0.08	0.98 ± 0.02	0.73 ± 0.05	0.46 ±0.09

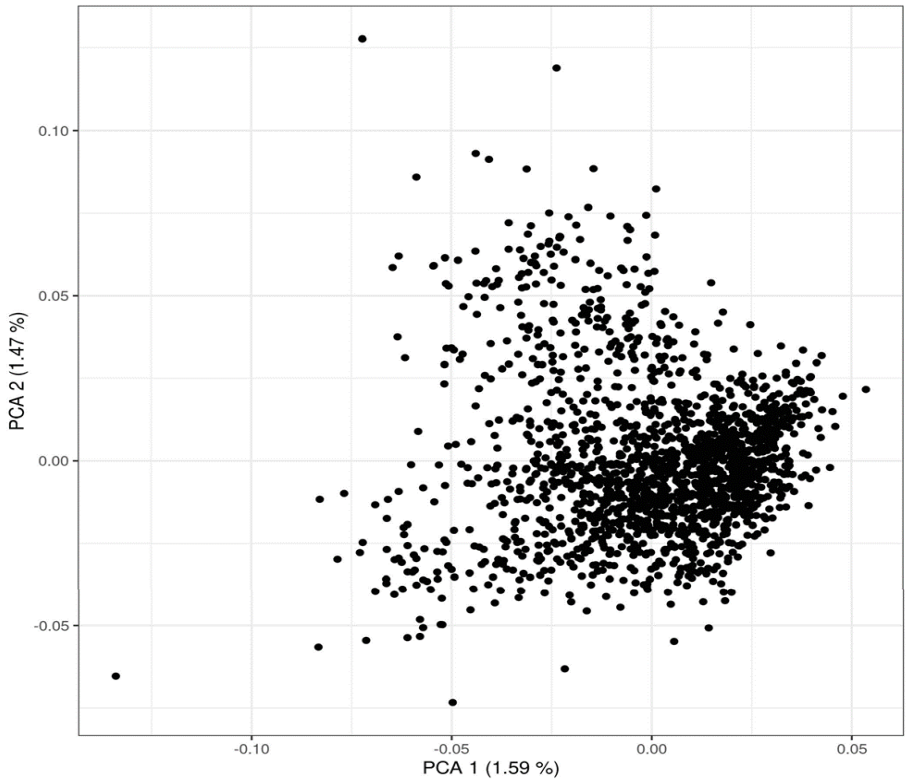
BW_i, body weight at the entrance at performance test stations; BW_m, body weight at six months; BW_f, at the end of performance test; ADG_i, average daily gains covering the first half of the period (since entering into the PT station until the 6th month);

ADG_f, average daily gain covering the daily gain of the second half (since the 6th month to the end of the period), ADG_tot average daily gain covering the entire period; DP, Dressing Percentage; CF, Carcass Fleshiness

Genomic architecture and imputation

A homogeneous density distribution (number of SNPs per Mb) was found throughout the genome, apart from few relatively small blank areas in 12 chromosomes. For further details on SNP density on each chromosome after imputation and quality control, see Supplementary Material, Figure S2. The new imputed panel had a SNPs density close to the one found in the young bulls genotyped with HD platforms. A value of imputation accuracy of 0.95 ± 0.05 was observed via cross-validation in the HD males (Supplementary Material, Figure S2). Combined with the high correlation between the A and G matrix, these results confirm the reliability of the new Alphaimpute2 algorithm for this population.

Figure 1: Scatter plot of first and second principal components of the genomic relationship matrix (the G matrix) used in the ssGBLUP. A total of 113,279 SNPs and 1,690 cattle were used to perform the principal component analysis.



The PCA scatterplots (Figure 1) illustrate a homogenous distribution of allele frequencies in individuals that comprised our study population. No stratification has been observed in the first two components, suggesting that most G matrix variance is explained by many eigenvalues with small effect. Genome-wide linkage disequilibrium and MAF have also been explored since the availability of high-density SNP platforms permits to explore the LD decay at an unprecedented resolution. In addition, MAF and LD are useful for understanding differences in population history and demography and for its impacts for genome-wide mapping studies. LD decay per each chromosome is reported in Supplementary Material, Figure S3. As expected, most tightly linked SNPs presented strong levels of LD while it rapidly declines when the distance increases. A within-chromosome LD average value of 0.19 ± 0.12 has been observed. When the distance between markers is lower than 1 Mb, the LD squared correlation between pairs of loci across autosomes (r^2) (Hill and Robertson, 1968) reached an average value of 0.17 ± 0.27, and when the distance was > 1 Mb LD decreased to 0.04 ± 0.09 (Figure S3). Larger levels of LD have been observed for chromosome 6 (0.20), while lower levels of LD were observed for chromosome 28 (0.18). An average value of 0.29 ± 0.12 was observed for minor allele frequency; no noticeable difference has been observed along the 29 chromosomes, with MAF values ranging from 0.28 ± 0.12 (chromosome 12) to 0.30 ± 0.12 (chromosome 19). With respect to the other local Italian breeds (i.e., Fabbri et al., 2020), Rendena presents a lower level of LD. This issue implicitly underlines the reassuring demographic situation of Rendena compared with other indigenous cattle of Italy, as it demonstrates a lower risk of inbreeding depression.

Heritability and genetic correlations

Descriptive statistics after data editing of the phenotypes are shown in Table 1. Phenotypic and genetic correlations and the heritability (h^2) for the analyzed traits are reported in Table 2. Body weight traits presented an average value of h^2 lower than other traits: BW_i showed the lower heritability (0.130), while BW_m and BW_f had heritability of 0.220. In fact, as reported in literature, a large discrepancy of values has been observed for heritability of body weights, and generally, traits similar to birth weight or weaning weight have a slightly lower heritability than weight measured in more advanced stages (Yin and König, 2018). Average daily gain (ADG_tot) presented an intermediate heritability of 0.322 partitioned into 0.164 and 0.220 for ADG in the first and last period. As for body weight, ADG presents lower h^2 in first stages of

the performance test, and h^2 values agree with what has been found in the literature (Yin and König, 2018). The highest heritabilities were found for the traits related to the carcass conformation, with a value of 0.45 and

0.47 respectively for CF and DP, close to what was observed in other local dual-purpose or beef cattle (Albera et al., 2001; Sbarra et al., 2013; Mancin et al., 2021c). These traits also appeared highly genetic correlated. All ADG traits were moderately genetically correlated with them, with avalue of 0.5 on average. On the contrary, body weight measured at the beginning of the performance test was not significantly correlated with CF and DP. Interestingly, the weights measured in more advanced periods showed an increase of genetic correlation with a value close to

0.7. Body weight and ADG also presented a strong genetic correlation with body weight traits, especially for the traits measured at the final stages of the performance test. In terms of genetic correlations, the results agree with what was found in other local dual-purpose or beef breeds (Veselá et al., 2011; Filipčík et al., 2020). Phenotypic correlation followed the same trends ofgenetic correlation but with a lower magnitude (Table 2, under diagonal).

Genomic architecture and imputation

A homogeneous density distribution (number of SNPs per Mb) was found throughout the genome, apart from few relatively small blank areas in 12 chromosomes. For further details on SNP density on each chromosome after imputation and quality control, see Supplementary Material, Figure S2. The new imputed panel had a SNPs density close to the one found in the young bulls genotyped with HD platforms. A value of imputation accuracy of 0.95 ± 0.05 was observed via cross- validation in the HD males (Supplementary Material, Figure S2). Combined with the high correlation between the A and G matrix, these results confirm the reliability of the new Alphaimpute2 algorithm for this population.

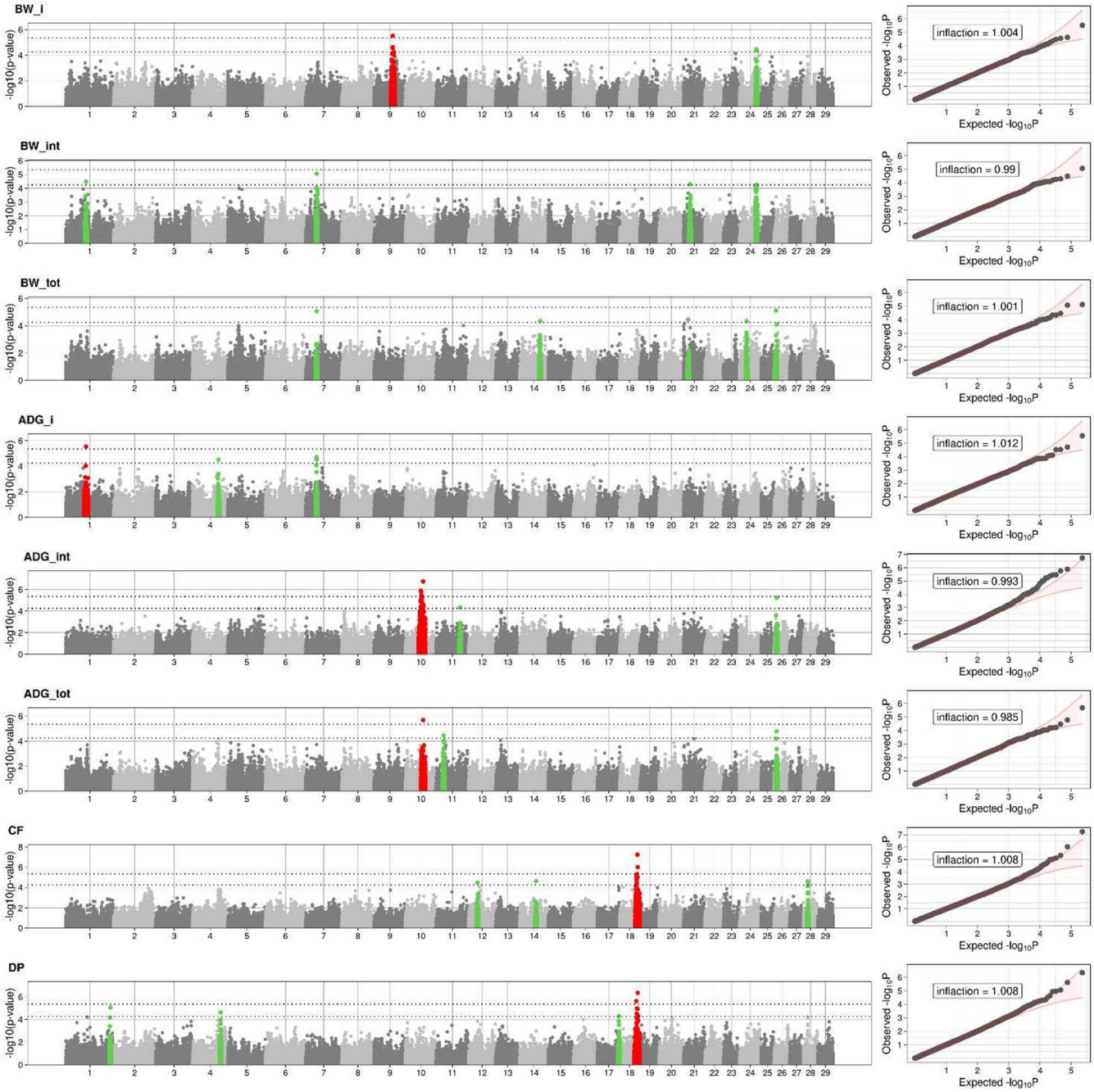
The PCA scatterplots (Figure 1) illustrate a homogenous distribution of allele frequencies in individuals that comprised our study population. No stratification has been observed in the first twocomponents, suggesting that most G matrix variance is explained by many eigenvalues with small effect. Genome-wide linkage disequilibrium and MAF have also been explored since the availability of high-density SNP platforms permits to explore the LD decay at an unprecedented resolution. In addition, MAF and LD are useful for understanding differences in population history

and demography and for its impacts for genome-wide mapping studies. LD decay per each chromosome is reported in Supplementary Material, Figure S3. As expected, most tightly linked SNPs presented strong levels of LD while it rapidly declines when the distance increases. A within- chromosome LD average value of 0.19 ± 0.12 has been observed. When the distance between markers is lower than 1 Mb, the LD squared correlation between pairs of loci across autosomes (r^2)(Hill and Robertson, 1968) reached an average value of 0.17 ± 0.27, and when the distance was > 1Mb LD decreased to 0.04 ± 0.09 (Figure S3). Larger levels of LD have been observed for chromosome 6 (0.20), while lower levels of LD were observed for chromosome 28 (0.18). An average value of 0.29 ± 0.12 was observed for minor allele frequency; no noticeable difference has been observed along the 29 chromosomes, with MAF values ranging from 0.28 ± 0.12 (chromosome 12) to 0.30 ± 0.12 (chromosome 19). With respect to the other local Italian breeds (i.e., Fabbri et al., 2020), Rendena presents a lower level of LD. This issue implicitly underlines the reassuring demographic situation of Rendena compared with other indigenous cattle of Italy, as it demonstrates a lower risk of inbreeding depression.

GWAS and pathway analysis

The full results of GWAS are reported in Table 3. We found a total of 8 SNP significantly associated with 5 of the investigated traits, and 47 SNPs suggestively associated with all 7 investigated traits (Figure 2). Pathway analysis revealed that out of 113,279 SNPs, 77,506 were located within a 15 kb window of annotated genes; in the end, 14,380 annotated genes were used asa background for each trait. On average, 628 genes near significant SNPs (< 0.01) were identified and subsequently used for pathway analysis of each trait. All traits presented an inflation factor close to optimum values of 1 (Figure 2) calculated based on the median chi-squared test. In addition, analysis on localized linkage disequilibrium (0.5 Mb form significant SNP), has been carried out (Figure 3-7), and results indicated that all significant candidate genes are extremely close to the significant SNPs, except for candidate gene *ZNF784*, which is situated between two significant SNP (Figure 6).

Figure 2. Manhattan and Q-Q plots of BW_i: body weight at the entrance at performance test stations; BW_m: body weight at six months; BW_f: body weight at the end of performance test. Average daily gain: ADG_i, covers of the first half of the period (since entering into the PT station until the 6th month); ADG_f, covers the daily gain of the second half (from the 6th month to the end of the period); ADG_tot is the average daily gain throughout the entire period. DP, Dressing Percentage; CF, Carcass Fleshiness. Dotted lines represent the suggestive and the significant threshold. Red dot represented the significant SNPs and neighboring SNPs (± 1 Mb) while green dots are the SNPs and neighboring SNPs (± 1 Mb). Q-Q plots are displayed as scatter plots of observed and expected –log10 (p-values) (right). Values of inflation are reported within the QQplots.



Body weight

Significant SNPs contributing to the genetic effect of body weight are listed in Table 3. Body weight measured at first stage was the only BW trait in which significant SNPs were identified, while body weight measured at the half of the performance test period presented the lowest number of suggestive SNPs and biological pathways enriched. The significant peak for BW_i was located at64 Mb on BTA9, in the vicinity of gene *TBX18* (Table 3; Figure 3). This gene is mainly involved in controlling the first stages of embryonic development and in the morphogeny of the embryonic epithelium (Consortium, 2021). To our knowledge, no previous connection with body weight had ever been found for *TBX18*; however, a study found an association between this gene and development in dual-purpose Simmental breed but not in other specialized breeds (Doyle et al., 2020a). We hypothesize that a possible mechanism for the connection between *TBX18* and body weight could lie in the fact that it is a strict paralogue of *TBX15*, a gene linked to obesity-related traits in humans and mice (Ejarque et al., 2019; Sun et al., 2019); it is demonstrated that *TBX15* regulates processes related to the skeletal muscles metabolism, which is in turn linked to animals' body size (Lee et al., 2015). However, studies on the relationship between *TBX15* and *TBX18* in cattle and the impact of *TBX15/18* on the regulation of muscle metabolism are needed to validate this hypothesis. We identified several known cattle QTLs in QTLdb overlapping with our candidate region (Supplementary Material Table S2a): the majority of these QTLs were linked to morphology (47.5%), followed by beef production (22.5%).

MYO5B is a candidate gene for both BW_m and BW_i (Table 3), identified by the presence of twosuggestively associated SNPs located on chromosome 24. *MYO5B* is related to the development ofskeletal muscle for what concerns actin and myosin organization and with the binding of ATP (Consortium, 2021). Interestingly, this gene was also identified in GWAS conducted on dual-purpose Simmental breeds (Doyle et al., 2020b).

The analysis of the enriched pathways, represented in Figure 8, reinforced what has been mentionedfor the single genes, namely that in our study the mechanisms regulating body weight were mainly those linked to the development of muscle masses. Among the GO terms enriched (Figure 8 and S4a-b), there were: organization of cytoskeleton (GO:0007010), actomyosin structure (GO:0031032), actin filament bundle (GO:0061572), and contractile actin filament bundle assembly (GO:0051017). The pathways analysis revealed a further biological process

related to the metabolism of lipids on skeletal muscles (GO:0055088, GO:0055092, GO:0042632). Regulation of the selection of appropriate nutrients by the skeletal muscle is essential both in terms of muscle energy metabolism and in terms of general regulation of whole-body supply and use of fuel (Hocquette et al., 1998): again, this enriched pathway was also found in Srivastava et al. (2020).

Aside from the already mentioned *MYOB5,* two candidate genes within suggestively associated SNPs were identified for BW_m: *CPEB1* and *DIRC2*, found on BTA1 and 21, respectively (Table 3). While these genes are not directly involved with body weight, we found them related to factors with a potential secondary impact on growth. For example, the *CPEB1* gene is involved in the regulation of mRNA translation and cell proliferation, with an influence on the molecular mechanisms associated with superior resilience to heat stress in cattle (Livernois et al., 2021). Moreover, *CPEB1* was also detected by other GWAS studies in cattle in which the target phenotype was residual feed intake (Lapierre et al., 1995). *DIRC2* has been associated with lipid storage in geese's (*Anser anser domesticus*) liver (Yang et al., 2020), given its role as a substrate carrier.

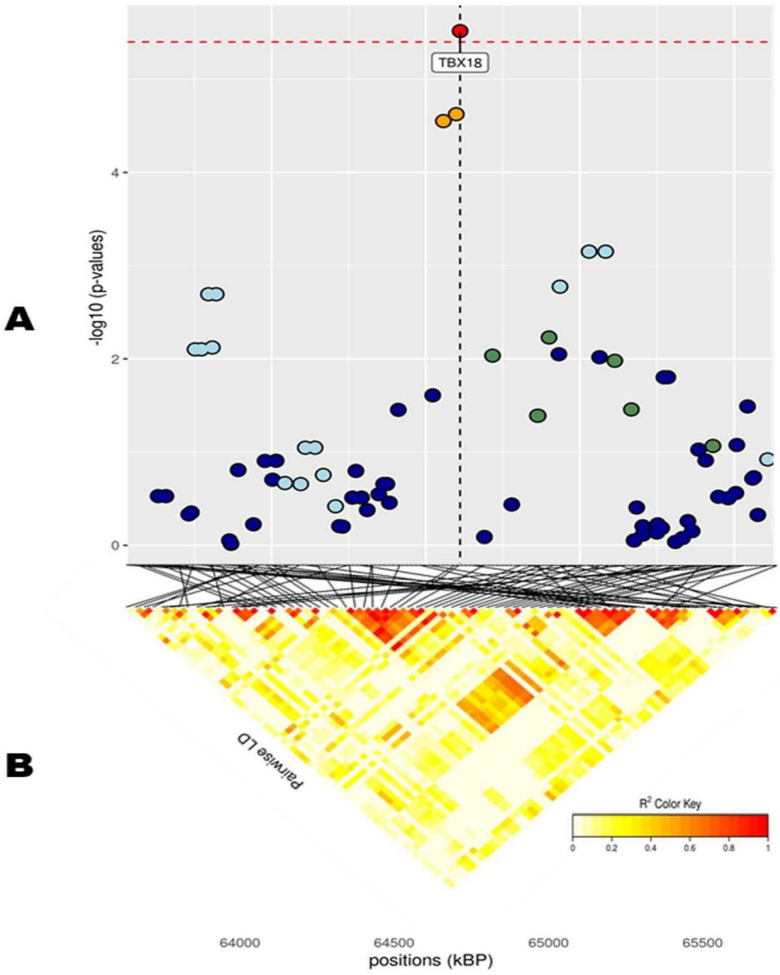
In BW_f, as in the other phenotypes, several genes identified by suggestively associated SNPs (Table 3) had never been associated before with body size traits. Moreover, connections between such candidate genes and body weight were not straightforward. One suggestively associated gene for BW_f, *CCDC178*, was identified in some GWA studies on disease resistance in local cattle (Kosińska-Selbi et al. 2020). The *MBL2* gene, a candidate gene suggestively associated to BW_f (and almost suggestive for ADG_tot), also seems to have an indirect connection with body weight: *MBL2* plays a central role in the activation of the mannose-binding lectin or mannose-binding protein; this protein is involved in processes that regulate the immune system, preventing infection from bacteria, virus, and yeast (Consortium, 2021).

No biological process strictly related to muscle mass development was identified (Figure 8 and S4c), but many processes related to other aspects of growth and body weight have been found. Several pathways were involved in GABA processes (Figure 8 and S4a-c): GABA is actively involved in regulating leptin, the satiety hormone, which has an essential role in nutrient intake and feeding motivation (Miller 2017). Some pathways also appear to be associated with processes such as morphogenesis of the epithelium (GO:0048791, GO:0007492, GO:0048332, GO:0001707 GO:0035987; mesoderm morphogenesis in Figure 8), which has a connection with

body weight (increased paracellular permeability for the absorption of nutrients leads to augmented energy intake (Vanvanhossou et al., 2020).

Finally, many enriched terms were related to neuronal aspects (i.e., GO:0043005 GO:0097060, GO:0099537; Figure 8 and S4a-c): this may find justification in the many studies underlining how these pathways are linked to the complex interaction between physio- and behavioral componentsthat control the intake of food and energy expenditure (Martinez, 2000).

Figure 3 (A) Localized linkage disequilibrium analysis of BW_i. Manhattan plots displaying the level of significance (y-axis) over genomic positions (x-axis) in a window of 0.5 Mb upstream and downstream of the most significantly SNP. Vertical line represents the position of candidate gene *TBX18*. Different colors are used to represent the pairwise LD with the closest significant SNPs: blue < 0.2; light blue < 0.4; green < 0.6; yellow < 0.8 and red > 0.8. **(B)** Represents linkage disequilibrium of that area.



Average Daily Gain

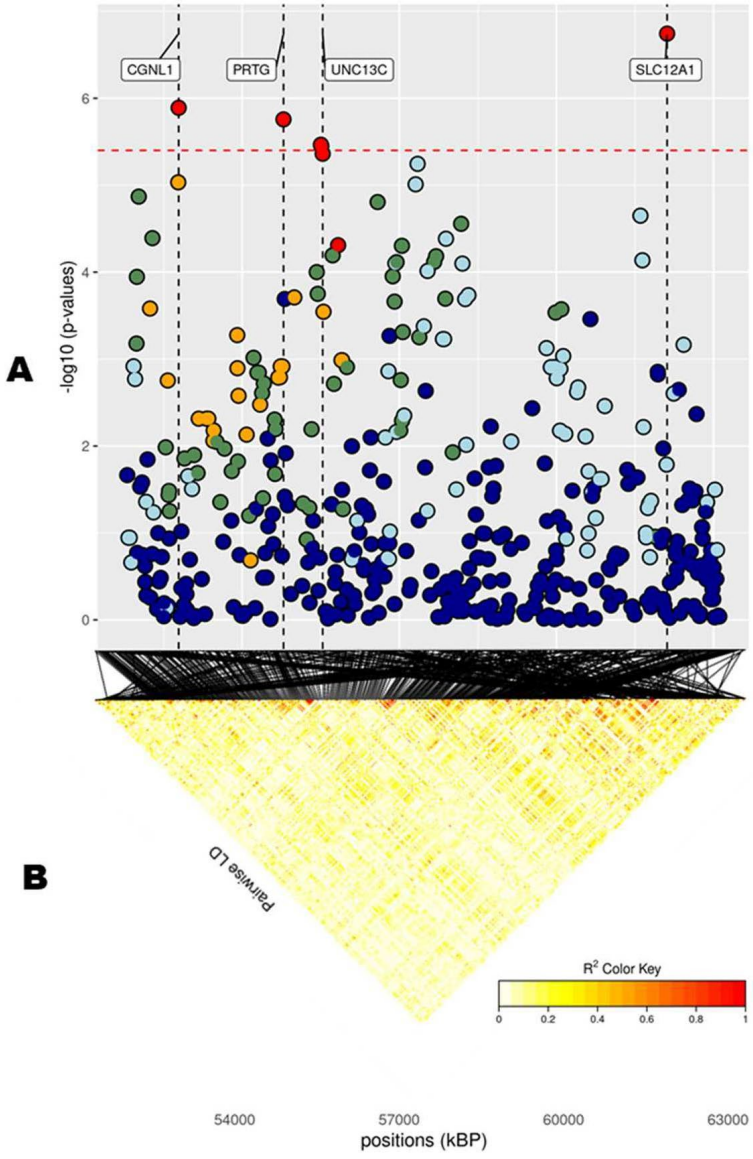
Both GWAS and pathway analyses of Average Daily Gain showed different results depending on the age at which the trait was recorded, similarly to what resulted from our analysis of BW. In particular, the only GO terms in common between ADG_i and ADG_f were GO:0031175 (neuron projection development) and its associated terms; all the other 105 GO, and KEGG terms were not (Figure S4d). The result of the GWAS also highlighted SNPs present in wholly different BTAs (Table 3). ADG_i had only one significant SNP (also suggestively associated with BW_m) situated on BTA1 (Figure 4), 0.2 Mb away from gene *DIRC2* (also associated with BW_m) and 1.1 Mb away from gene *HSPBAP*. Both loci can be in some ways considered candidate genes for growth, as also *HSPBAP* has already been associated with residual feed intake from birth to 12 months(Cohen-Zinder et al., 2016). One suggestively associated SNP for ADG_i on BTA4 (Table 3) waswithin candidate gene *GRM8*, associated with body size in cattle (Chen et al., 2020) and eating behavior in other mammals (Gast et al., 2013). Again, in agreement with what was found for BW_m (the measure of ADG_i is based on the difference between BW_m and BW_i measurement), the results of the pathway analysis for ADG_i were less extensive than for other ADG traits (Figure 9 and S4d-f); moreover, out of 20 pathways (Figure S4d), those readily associable with ADG were GO:0004629 phospholipase activity (crucial for lipid metabolism) and GO:0043124, responsible for negative regulation of I-kB kinase/NF-κB signaling (involved with metabolic regulation, especially in cases of overnutrition; Kracht et al., 2020).

The same trait recorded at a later age, ADG_f, showed a much greater number of results, similarly to what transpired with BW_f (Table 3; Figure 9 and S4e). For trait ADG_f the region with the greatest number of signals was on BTA10, roughly between 50 and 60 Mb (Table 3; Figure 5). This region contains a QTL that has already been associated to growth in cattle (Mao et al., 2016), although not in the present study. The three significant SNPs and 14 out of 16 suggestively associated SNPs were found in this region. Significant SNPs were situated within *SLC12A1*, *CGNL1* and *PRTG* genes (Figure 5). While the latter two have already been associated respectively with growth (Londoño-Gil et al., 2021) and backfat thickness in cattle (Júnior et al., 2016), *SLC12A1,* to our knowledge, has never been associated with growth or weight traits in cattle (but see Kemter et al., 2014, for evidence in mice). However, among the suggestively associated SNPs on BTA10 (Table 3), several were within or close genes highly important for ADG, such as *ALDH1A2*, *FBN1,* and *AQP9* (Hirano et al., 2012; Liu et al., 2019; Londoño-Gil et al., 2021; Zhang et al., 2021). Figure 9 shows that enriched pathways spanned

several macro-categories (Figure 9 and S4e): these results suggest that, as for BW, during the late months of the first year, a complex interplay of different biological processes takes place in growing bulls. For what concerned the overlapping of our QTLs associated with ADG_f with the animal QTLdb, we identified QTLs from several studies: 28.77% associated with morphology, 21.92% associated with beef production, 19.18% associated with milk, and 8.22% associated with meat and carcass (Supplementary Material, Table S2b).

Finally, for the total ADG, ADG_tot, the results obtained mirrored those obtained with final ADG, both in terms of significant and suggestive SNPs (on BTA10 and BTA26; Table 3) and in terms of GO terms (Figure 9 and S4f) and candidate genes, such as *SLC12A1*. Interestingly, one signal reported in ADG_tot was not present in ADG_f: on BTA11, one single suggestively associated SNP was located close to two genes well known for their effect on feed intake and weight (*CDKL4* and *MAP4K3*; Edea et al., 2020). Apart from this exception, our results show conclusively that total average daily gain mirrored the final part of the daily gain, i.e., that the last months were decisive inshaping the total weight gain trajectory of the bulls.

Figure 5 (A) Localized linkage disequilibrium analysis of ADG_f. Manhattan plots displaying the level of significance (y-axis) over genomic positions (x-axis) in a window of 0.5 Mb upstream and downstream of the most significantly SNP. Vertical line represents the position of candidate genes *CGNL1*, *PRTG*, *UNC13C* and *SLC12A1*. Different colors are used to represent the pairwise LD with the closest significant SNPs: blue< 0.2; light blue < 0.4; green < 0.6; yellow < 0.8 and red > 0.8. **(B)** the represents Linkage disequilibriumpresent of that area.



Carcass Traits

The main region of interest for both CF and DP traits was situated on a gene-rich region of BTA18, between 55 Mb and 62 Mb, where 3 significant and 9 suggestively associated SNPs allowed to locate several candidate genes (Table 3; Figure 6). The QTL with the highest significance for CF (suggestively associated for DP) was located within candidate gene *LOC513941* (Figure 7), translating into a cationic amino acid transporter 3-like. This type of transporters regulates the metabolism of cationic amino acids, a key factor for growth and beef characteristics in cattle (Liao et al., 2009). Further corroboration of the importance of this metabolic pathway for CF was the enrichment of 10 GO terms (Figure 10 and S4h), within the group of 'amino acid transport', such as ammino acid transmembrane transporter activity (GO:0015171), and amino acid transmembranetransport (GO:0003333).

A second SNP in the same region (significant for DP and suggestively associated for CF; Table 3) was located within gene *CCDC155* (Coiled-coil domain containing 155). This gene encodes for a protein involved in dynein complex binding and actin filament organization and it has been associated with beef conformation (Lemos et al., 2016; Hardie et al., 2017). Apart from being the main component of the cytoskeleton, actin constitutes together with myosin the myofilaments, which grant muscle cells their mobility and thus ultimately their organization and dynamics. The association of actin filaments and carcass traits was again made apparent also by the number (morethan 30) and diversity of enriched GO terms related to actin (Figure 10 and S4g-h): for example, those related to GO:0098858 (CF), actin-based cell projection; GO:0030048 (CF and DP), actin filament-based movement; GO:0070161 (CF and DP), anchoring junction; GO:0030833 regulation of actin filament polymerization; GO:0005912 (CF and DP), adherens junction (Londoño-Gil et al., 2021). Similarly, for DP 20 terms were enriched for pathways associated with actin filament-based GO terms (Figure S4g).

In the same region of BTA18, our analysis found two more candidate genes with a known association with size and growth traits, all with one or more suggestively associated SNPs for CF. *Siglec-5* is a gene commonly found in GWAS concerning cattle size and growth traits; its over- expression indicates a deficiency of leptin, and thus longer gestation time and bigger fetuses (Hardie et al., 2017). *KLK12* is a kallikrein gene, a serin protease associated with food intake and feed efficiency at the transcript level in backfat and rumen (Kern et al., 2016).

LOC101904435 and *ZNF784* are zinc-finger proteins: the former is suggestively associated with both CF and DP; the latter only with CF but is linked to food intake in cattle (Oliviera et al., 2016).

Finally, three more SNPs (one significant both for CF and DP and two SNPs suggestively associated for DP) were situated within *NLRP2* gene (NACHT, LRR and PYD domains-containing protein 2), a key player in early embryogenesis, maternal effects, immune response, and inflammasome (Peng et al., 2012). Taken together, these results about carcass traits have numerous substantial implications. Firstly, we highlight how the 57-62 Mb region on BTA18 can truly be considered a hotspot of genetic diversity in this breed (as it is for several others; Grigoletto et al., 2020; Purfield et al., 2020). Secondly, as expected with strongly correlated traits, CF and DP shared part of their genetic architecture, as significant SNPs for the two traits are mostly in the same region. Only another region was shared, as both traits reported two suggestively associated SNPs close to each other on BTA28 (Table 3)

The region encompasses the *PHYHIPL* gene, which influences feed efficiency (Abo-Ismail et al., 2018), whose link with carcass traits has recently been established (Seabury et al., 2017). CF was associated only with two more SNPs, one on BTA12 and the other on BTA14 (Table 3). While the former was more than 1 Mb far away from any annotated functional element, the latter fell within *SAMD12*, a gene already found to have a significant dominance signal to chuck roll and be associated with 18-months weight in Simmental (Zhuang et al., 2020). On the other hand, DP had an almost significant signal on BTA1: the gene closest to the SNP was *SIM2*, already known to be associated with carcass quality, differentiation of *longissimus,* and *semimembranosus* muscle (De Las Heras-Saldana et al., 2019; Edea et al., 2020). To conclude, the strongest of the remaining suggestively associated signals for DP came from BTA4, within *LOC112446424*, a non-coding RNA close to candidate gene *SLC13A4*, a cationic canal important both for muscle traits in sheep and growth and development in cattle (Carvalho et al., 2020; Kaur et al., 2020).

While, as we mentioned, results from pathway analyses (represented in Figure 10 and figure S4g-h), and GWAS were often complementary, pathway analyses for both CF and DP resulted in the enrichment of a robust number of pathways related to neuron activity, not really pointed out by GWAS results. Such pathways referred to the regulation of neuroblast proliferation (GO:1902692 for CF), chemical synaptic transmission (GO:0007268 for CF), neurogenesis (GO:0022008 for CF and DP), neuron projection (GO:0043005 for DP), synapse

(GO:0045202 for DP) and especially synaptic transmission, glutamatergic (GO:0035249 for DP and, to a lesser extent, CF). Glutamatergic synapses guide the development of growth neurons and regulate feeding motivation in the hippocampus (Huang et al., 2017). The relation between feeding motivation and nutrient intake is crucial to maintaining energy intake and storage (Illius et al., 2002). Such relationship is complex, involving leptin (see above-mentioned gene *Siglec-5*), and the NPY/AgRP system, which makes food intake-stimulating peptides, which can dramatically influence metabolism and consequently carcass traits (Seabury et al., 2017; Ruud et al., 2020). Among the genes more often represented in the glutamatergic synapse network enriched in our analysis, several were linked with food intake and metabolism (for example, *GRM8*), eating behavior (*GRIK3*), insulin secretion, and lipolysis (*ADCY1*, Olivieri et al. 2016). In support of this hypothesis, we also found out that the enriched KEGG term for DP Glutamatergic synapse (KEGG:04724) belonged to the same group of Circadian entrainments (KEGG:04713) and Apelin signaling pathway (KEGG:04371), both also enriched. Circadian rhythm has a strong connection with feeding behavior (Mrode et al., 2019), and apelin is a peptide connected with food intake and lipid metabolism (Bertrand et al., 2015). The same was true also for CF, with KEGG term Hippo signaling pathway (KEGG:04390) appearing multiple times (Figure S4h). This might reflect a greater role of regulatory systems of feeding motivation, nutrient intake, and storage in shaping the variability of these traits. On the other hand, glutamatergic synapses are also involved in physiological responses to stressors and environmental changes. QTLs from the QTLdb associated to our candidate regions for these two traits are reported in Supplementary Material, Table S2c-d.

Figure 4 (A) Localized linkage disequilibrium analysis of ADG_i. Manhattan plots displaying the level of significance (y-axis) over genomic positions (x-axis) in a window of 0.5 Mb upstream and downstream of the most significantly SNP. Vertical line represents the position of candidate gene *DIRC2*. Different colors are used to represent the pairwise LD with the closest significant SNPs: blue < 0.2; light blue < 0.4; green < 0.6; yellow < 0.8 and red > 0.8. **(B)** Represents linkage disequilibrium of that area.

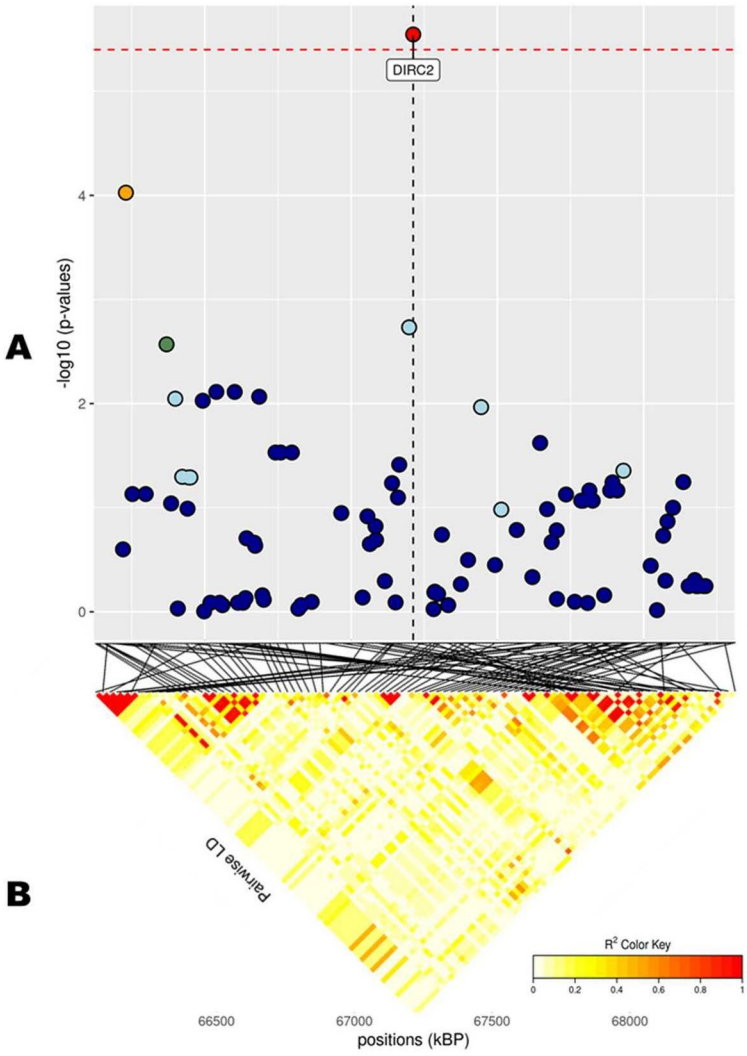


Table 3 Significant and suggestively SNPs found on the GWAS study. Significant SNPs are reported in **bold**. Gene with * were just outside suggestive association range for one trait; it was retained in the table because significant for another trait.

Trait	BTA	Position of the SNP (bp)	Significance of the SNP (-log(p-value))	Nearest gene(s)	Distance to nearest gene (kb)	Other traits associated	Variance explained (%)
Body Weight							
BW_i	**9**	**64611352**	**3.04E-06**	***TBX18***	**0.589**		0.22 %
BW_i	9	64599056	2.37E-05	*TBX18*	12.885		
BW_i	9	64557321	2.81E-05	*TBX18*	54.620		
BW_i	24	49394386	3.43E-05	*ACAA2*	48.389		
BW_i	24	49493559	4.43E-05	*MYO5B*	within		
BW_m	7	32306269	8.65E-06	*FTMT*	321.80	BW_f; ADG_i	
BW_m	1	67212088	3.27E-05	*DIRC2*	2.783	ADG_i	
BW_m	21	22956171	5.11E-05	*CPEB1*	within		
BW_m	24	49735783	5.55E-05	*MYO5B*	within		
BW_f	26	6437290	7.50E-06	*MBL2*	3.483	ADG_tot	
BW_f	7	32306269	8.39E-06	*FTMT*	321.80	BW_m, ADG_i	
BW_f	21	17568377	3.44E-05	*AGBL1*	within		
BW_f	24	24130452	4.56E-05	*CCDC178*	within		
BW_f	14	60644816	4.62E-05	*RIMS2*	within		
Average Daily Gain							
ADG_i	**1**	**67212088**	**2.84E-06**	***DIRC2***	**2.783**	**BW_m**	**0.441 %**
ADG_i	7	32306269	1.99E-05	*FTMT*	321.80	BW_m; BW_f	
ADG_i	7	32009625	3.03E-05	*FTMT*	25.152		
ADG_i	4	91417417	3.11E-05	*GRM8*	within		
ADG_f	**10**	**62113751**	**1.81E-07**	***SLC12A1***	within	**ADG_tot**	**0.073 %**
ADG_f	**10**	**52785760**	**1.29E-06**	***CGNL1***	within		**0.203 %**
ADG_f	**10**	**54787499**	**1.75E-06**	***PRTG***	within		**0.435 %**
ADG_f	10	55502036	3.42E-06	*UNC13C*	135.046		
ADG_f	10	55510249	3.56E-06	*UNC13C*	126.833		
ADG_f	10	55535781	4.35E-06	*UNC13C*	101.301		
ADG_f	10	57348706	6.68E-06	*LOC101904374*	248.031		
ADG_f	26	8564813	5.92E-06	*A1CF; ASAH2*	17.739; 32.479	ADG_tot	
ADG_f	10	52777666	9.27E-06	*CGNL1*	within		
ADG_f	10	57311183	9.77E-06	*LOC101904374*	285.554		
ADG_f	10	52023061	1.35E-05	*AQP9*	65.881		
ADG_f	10	56585283	1.56E-05	*WDR72*	within		
ADG_f	10	61604387	2.24E-05	*LOC104973175; FBN1*	20.944; 51.118		

ADG_f	10	58180258		*MYO5C; GNB5*	1.494; 11.943		
ADG_f	10	63669471	3.56E-05	-			
ADG_f	10	52284899	4.06E-05	*ALDH1A2*	within		
ADG_f	10	57890651	4.13E-05	*MYO5A*	within		
ADG_f	11	78877665	4.48E-05	*WDR35*	within		
ADG_f	10	55830543	4.90E-05	*UNC13C*	within		
ADG_f	10	57048787	4.98E-05	*LOC101904374*	547.950		
ADG_tot	**10**	**62113571**	**2.07E-06**	***SLC12A1***	within	**ADG_f**	**0.501 %**
ADG_tot	26	8564813	1.66E-05	*A1CF; ASAH2*	17.739; 32.479	ADG_f	
ADG_tot	11	21542682	3.41E-05	*CDKL4; MAP4K3*	7.971; 11.618		
ADG_tot	26	6437290	6.03E-05*	*MBL2*	3.483	BW_f	
Dressing Percentage							
DP	**18**	**62412976**	**4.51E-07**	***NLRP2***	within	**CF**	**0.640 %**
DP	**18**	**55878286**	**2.40E-06**	***CDC155***	within	**CF**	**0.731 %**
DP	1	148893434	8.77E-06	*SIM2*	80.004		
DP	18	58645859	1.06E-05	*LOC101904435*	within	CF	
DP	18	61137684	1.15E-05	*LOC513941*	within	CF	
DP	4	99574406	2.34E-05	*LOC112446424*	within		
DP	18	57735853	3.03E-05	*LOC787554*	within	CF	
DP	18	62427814	4.49E-05	*NLRP2*	within		
DP	18	63362491	4.97E-05	*LOC107131476*	560		
DP	17	72055006	5.07E-05	*YPEL1*	23.650		
DP	18	62428754	5.25E-05	*NLRP2*	within		
Carcass Fleshiness							
CF	**18**	**61137684**	**5.62E-08**	***LOC513941***	within	**DP**	**0.450%**
CF	**18**	**62412976**	**9.40E-07**	***NLRP2***	within	**DP**	**0.670 %**
CF	18	58645859	4.71E-06	*LOC101904435*	within	DP	
CF	18	55878286	7.67E-06	*CCDC155*	within	DP	
CF	18	61920892	9.57E-06	*ZNF784*	895		
CF	18	57735853	1.05E-05	*LOC787554*	within	DP	
CF	18	57516245	1.66E-05	*LOC618268*	within		
CF	14	45804718	2.30E-05	*SAMD12*	within		
CF	28	14722675	2.48E-05	*LOC101906006*	within		
CF	18	57565406	3.23E-05	*SIGLEC5*	within		
CF	12	27043078	3.38E-05	-			
CF	18	57008781	4.83E-05	*KLK12*	within		
CF	28	14788560	5.31E-05	*PHYHIPL*	within		

The threshold of significance chosen for our analysis was p = 3.162 * 10-6, obtained through Bonferroni correction, while threshold for Bonferroni suggestive p-values was p = 5.629

Traits and Time Stratification

The results of our study can help frame the genetic architecture of our between-traits correlation, including such traits that are measures of the same trait in different time points or intervals (the three BW and the three ADG). Within BW, we demonstrated how also from the genomic point of view the weight at the half of the PT was underlined by a mixture of QTLs that were also found either at the start or at the end of the PT. On the other hand, no common SNPs resulted significant both for BW_i and BW_f, and the number of enriched pathways in common was very low (Table 3; Figure S4a-c). For what concerns ADG, there was also a deep difference between the signals found for ADG_i and ADG_f, with the latter reflecting much more closely the total ADG, and again no SNPs were shared by ADG_i and ADG_f (Table 3). Moreover, the lowest number of significant SNPs and pathways for BW was at BW_m, and for ADG was ADG_i, with these two traits sharing a temporal correspondence.

Interestingly, we found many genes in common between measures of different traits taken at the same time. For example, both SNPs on BTA7 and BTA1 were significant both for BW_m and ADG_i. Also, one SNP on BTA26 was suggestively associated both for BW_f and ADG_f (Table 3). These results have several implications: firstly, from an economic point of view, they show that the timing of the trait measurement is crucial. Different life stages can result in different genetic signals; if used for a selection program, this can have an economic and conservation impact. While this is of course expected, given the succession of different biological processes during development, very few studies include such a time stratification in their analysis of productive traits. Even if such a process is difficult to infer, our results show that complexity – intended as the number of functional elements, their diversity, and pathways involved – might increase with age.

Figure 8. Scatter plot representing the main groups of biological pathways enriched for Body Weight traits measured at first, half and final period of performance test (BW_i, BW_m, BW_f); the area represents the number of pathways in that group, among the total. For a detailed list of the pathways enriched by these traitssee Supplementary Material, figure S4a-c

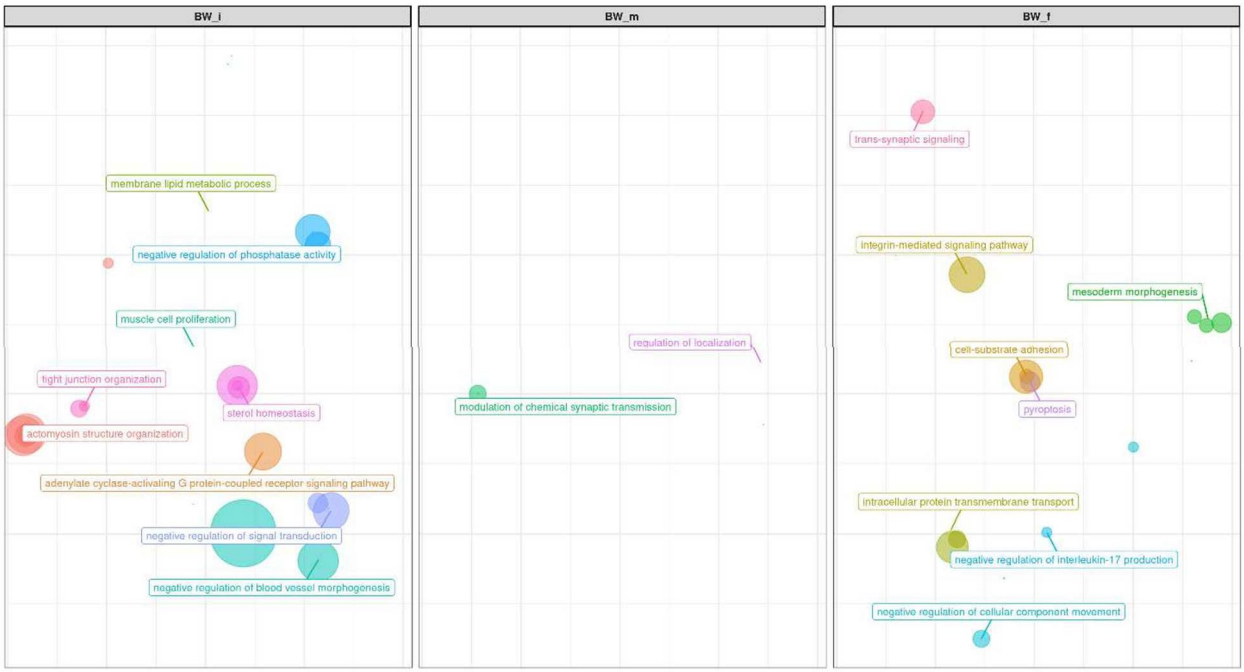


Figure 9. Scatter plot representing the main groups of biological pathways enriched for average daily gain traits measured at first, half and total period of performance test (ADG_i, ADG_f, ADG_tot); the area represents the number of pathways in that group, among the total. For a detailed list of the pathways enrichedby these traits see Supplementary Material, figure S4d-f.

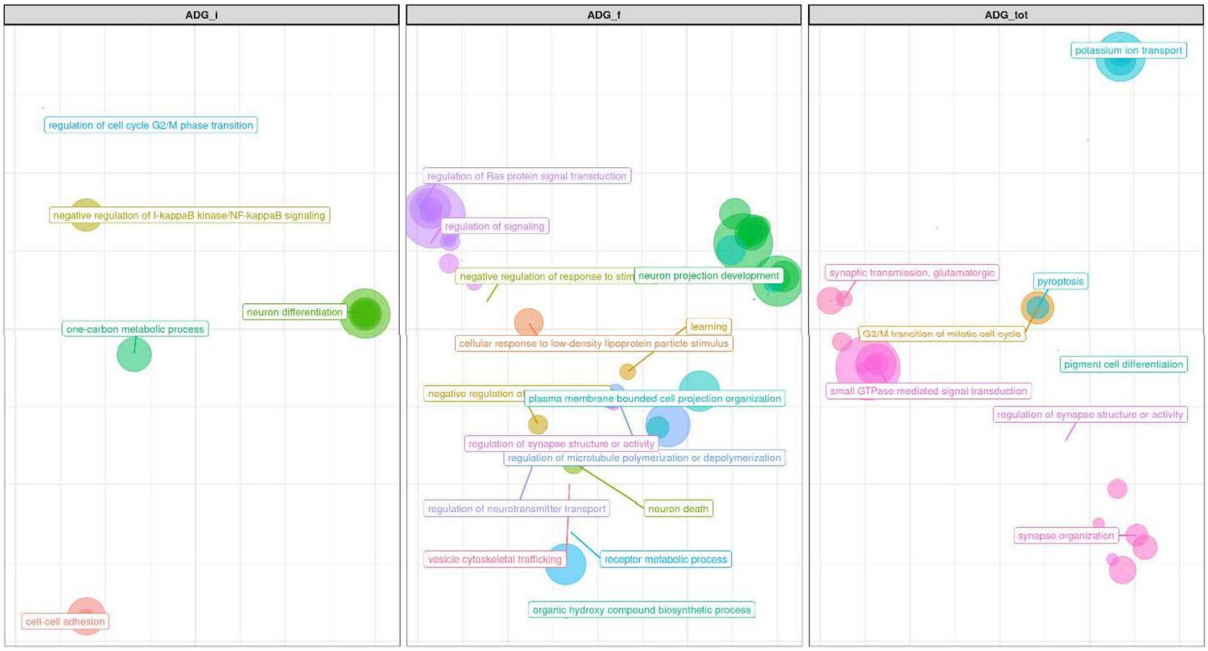


Figure 10. Scatter plot representing the main groups of biological pathways enriched for carcass traits (carcass fleshiness and dressing percentage). For a detailed list of the pathways enriched by these traits see Supplementary Material, figure S4g-h.

CONCLUSIONS AND IMPLICATIONS FOR LOCAL BREEDS

There are four main takeaways that could be extracted from our study. Firstly, our analysis detected a significant signal for body weight (recorded when bulls were one month old) on BTA9; a significant signal of average daily gain (recorded at seven months of age) on BTA1 and three significant signals of average daily gain (recorded at one year of age) on BTA10. Three significant signals for carcass traits (one signal each for dressing percentage and carcass fleshiness, plus one incommon between the two) were all situated on BTA18.

Secondly, the variety of GO terms and functional elements involved in the beef-related traits under study was staggering. We could detect in multiple traits key roles of pathways related to actin, lipid transport, and several types of channels. Moreover, our analysis detected – alongside many genes often found in relation to the investigated traits – multiple pathways, genes, and functional elements of unclear attribution, for example with links to early development and maternal effect (such as *TBX18, NLRP2, SLCA12*), or to pathogen resistance (*MBL2*). This issue underlines how even research of well-studied traits can turn out unexpected results, especially if performed in rarely investigated breeds. In additions, the fact that Rendena has been bred not only for the considered traits, but also for antagonistic could have added a layer of complexity to our results.

Thirdly, we detected for almost all traits several pathways and genes linked with neuroblast development and synaptic transmission, especially (but not exclusively) glutamatergic, which added to the the intricacy of the gene networks involved in these traits. Pathways linked to both neuroblast proliferation and synaptic communication have been tied in recent years to selection for environmental condition (Rowan et al., 2020) differences in behavioral temperament (Gutiérrez-Gilet al., 2008) and adaptability (Taye et al., 2017).

Finally, as discussed above, we found that even when focusing on widely investigated traits the influence of time stratification was fundamental. We argue that future studies on this issue should include an analysis of time stratification of their trait to fully report their complexity during development.

A greater diffusion of adaptable and diversified local breeds, with characteristics allowing for lower environmental impact, better survival and greater production in challenging environments might be crucial in staving off the negative effects of intensive beef farming. To

achieve this, however, there is urgent need for further studies of the genetic basis of productive and life-history trait, which are still lacking. Moreover, these studies could help uncovering several novel gene networks associations and pathways, thanks to the less intensive selection for production occurring in local breed. Finally, they would help to map the diversity of such breeds, in an unvaluable help for their conservation.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financialrelationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, R.M. B.T. E.M.; methodology, B.T. and E.M.; formal analysis, E.M. and B.T.; support to analysis S.P., investigation, S.P., B.T., C.S., R.M., and E.M.; resources, R.M.; data curation, E.M., and R.M.; writing original draft preparation, E.M. and B.T. writing—review and editing S.P., C.S., and R.M. All authors have read and agreed to the published version of the manuscript.

FUNDING

The study was funded by the DUALBREEDING project (CUP J61J18000030005) and byBIRD183281.

ACKNOWLEDGMENTS

Authors are grateful to National Breeders Association of Rendena cattle breed (ANARE) for datasupport.

SUPPLEMENTARY MATERIALS

TABLE S1. Accuracy of imputation over the 5 iterations of cross validation.

TABLE S2. Table representing the overlapping of our candidate region with QTL in animal QTLdb (animal QTL database), i.e., QTLs discovered in the other studies and summarized within the QTL database. (a) BW_i; body weight at first stage of performance test; (b) ADG_f; average daily gain from intermediate to final weighing (c) CF; *In vivo* Carcass Fleshiness; (f) DP; *in vivo* Dressing Percentage.

FIGURE S1. Diagram representing the Rendena selection scheme; young bulls are constantly used (80%) as sires of bulls, while in other breeds only proven bulls are used to father bulls.

FIGURE S2. Bar plot representing the density of genomic data after quality control and imputation for the 1,690 animals, divided in the 29 autosomes. Density is represented as number of SNPs within 1Mb, representing indirectly the performance of imputations.

FIGURE S3. Linkage disequilibrium decay for the genomic dataset for each of the 29 chromosomes. Red lines represent the regression of LD and distance. Differences in LD can be dueto various factors, among them chromosome length.

FIGURE S4: Bar plot representing the significantly enriched GO terms and KEGG pathways for the investigated traits. (a) Body weight at first stages of performance test; (b) body weight at intermediate period of performance test; (c) body weight at final period of performance test; (d) average daily gain to entrance to intermediate period of performance test; (e) average daily gain from intermediate period to final; (f) average daily gain in the whole performance test; (g) in vivoCarcass Fleshiness; (h) in vivo Dressing Percentage.

REFERENCES

Abo-Ismail, M. K., Lansink, N., Akanno, E., Karisa, B. K., Crowley, J. J., Moore, S. S., et al. (2018). Development and validation of a small SNP panel for feed efficiency in beef cattle. J. Anim. Sci. 96, 375–397. doi:10.1093/jas/sky020.

Aguilar, I., Tsuruta, S., Masuda, Y., Lourenco, D. A. L., Legarra, A., and Misztal, I. (2018). BLUPF90 suite of programs for animal breeding. 11th World Congr. Genet. Appl. to Livest. Prod., 11.751.

Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., and Lourenco, D. A. L. (2019). Frequentist p - values for large - scale - single step genome - wide association, with an application to birth weight in American Angus cattle. Genet. Sel. Evol., 1–8. doi:10.1186/s12711-019-0469-3.

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93, 743–752. doi:10.3168/jds.2009-2730.

Albera, A., Mantovani, R., Bittante, G., Groen, A. F., and Carnier, P. (2001). Genetic parameters for daily live-weight gain, live fleshiness and bone thinness in station-tested Piemontese young bulls. Anim. Sci. 72,449–456. doi:10.1017/S1357729800051961.

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465, 627—631. doi:10.1038/nature08800.

Begum, F., Ghosh, D., Tseng, G. C., and Feingold, E. (2012). Comprehensive literature review and statistical considerations for GWAS meta-analysis. Nucleic Acids Res. 40, 3777–3784. doi:10.1093/nar/gkr1255.

Bertrand, C., Valet, P., and Castan-Laurell, I. (2015). Apelin and energy metabolism. Front. Physiol. 6, 1–5.doi:10.3389/fphys.2015.00115.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.

Bioinformatics 25, 1091–1093. doi:10.1093/bioinformatics/btp101.

Biscarini, F., Nicolazzi, E., Alessandra, S., Boettcher, P., and Gandini, G. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. Front. Genet. 5, 1–16. doi:10.3389/fgene.2015.00033.

Buitenhuis, B., Janss, L. L. G., Poulsen, N. A., Larsen, L. B., Larsen, M. K., and Sørensen, P. (2014). Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics 15, 1–11. doi:10.1186/1471-2164-15-1112.

Carvalho, F. E., Espigolan, R., Berton, M. P., Neto, J. B. S., Silva, R. P., Grigoletto, L., et al. (2020). Genome-wide association study and predictive ability for growth traits in Nellore cattle. Livest. Sci. 231,103861. doi:10.1016/j.livsci.2019.103861.

Chen, Q., Huang, B., Zhan, J., Wang, J., Qu, K., Zhang, F., Shen, J., Jia, P., Ning, Q., Zhang, J. and Chen, N., (2020). Whole-genome analyses identify loci and selective signals associated with body size in cattle. J.Anim. Sci. 98(3), p.skaa068. doi:10.1093/jas/skaa068.

Cheverud, J. M. (2001). A simple correction for multiple comparisons in interval mapping genome scans.Hered. (Edinb) 87, 52–58. doi:10.1046/j.1365-2540.2001.00901.x.

Cohen-Zinder, M., Asher, A., Lipkin, E., Feingersch, R., Agmon, R., Karasik, D., et al. (2016). FABP4 is a leading candidate gene associated with residual feed intake in growing holstein calves. Physiol. Genomics 48, 367–376. doi:10.1152/physiolgenomics.00121.2015.

Consortium, T. U. (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49,D480–D489. doi:10.1093/nar/gkaa1100.

De Las Heras-Saldana, S., Chung, K. Y., Lee, S. H., and Gondro, C. (2019). Gene expression of Hanwoo satellite cell differentiation in longissimus dorsi and semimembranosus. BMC Genomics 20, 1–15. doi:10.1186/s12864-019-5530-7.

de Oliveira Silva, R., Bonvino Stafuzza, N., de Oliveira Fragomeni, B., de Camargo, G., Matos Ceacero, T., dos Santos Gonçalves Cyrillo, J., et al. (2017). Genome-Wide Association Study for Carcass Traits in an Experimental Nelore Cattle Population. PLoS One 12, 1–14. doi:10.1371/journal.pone.0169860.

Doyle, J. L., Berry, D. P., Veerkamp, R. F., Carthy, T. R., Walsh, S. W., Evans, R. D., et al. (2020a). Genomic Regions Associated With Skeletal Type Traits in Beef and Dairy Cattle Are Common to Regions Associated With Carcass Traits, Feed Intake and Calving Difficulty. Front. Genet. 11. doi:10.3389/fgene.2020.00020.

Doyle, J. L., Berry, D. P., Veerkamp, R. F., Carthy, T. R., Evans, R. D., Walsh, S. W., & Purfield, D. C. (2020b). Genomic regions associated with muscularity in beef cattle differ in five contrasting cattle breeds. Genetics, selection, evolution: GSE, 52(1), 2. https://doi.org/10.1186/s12711-020-0523-1

Drost, H.-G., and Paszkowski, J. (2017). Biomartr: genomic data retrieval with R. Bioinformatics 33, 1216–1217. doi:10.1093/bioinformatics/btw821.

Edea, Z., Jung, K. S., Shin, S. S., Yoo, S. W., Choi, J. W., and Kim, K. S. (2020). Signatures of positive selection underlying beef production traits in Korean cattle breeds. J. Anim. Sci. Technol. 62, 293–305. doi:10.5187/JAST.2020.62.3.293.

Ejarque, M., Ceperuelo-Mallafré, V., Serena, C., Maymo-Masip, E., Duran, X., Díaz-Ramos, A., et al. (2019). Adipose tissue mitochondrial dysfunction in human obesity is linked to a specific DNA methylation signature in adipose-derived stem cells. Int. J. Obes. 43, 1256–1268. doi:10.1038/s41366-018-0219-6.

Fabbri, M. C., Dadousis, C., and Bozzi, R. (2020). Estimation of linkage disequilibrium and effective population size in three Italian autochthonous beef breeds. Animals 10, 1–14. doi:10.3390/ani10061034.

Falker-Gieske, C., Blaj, I., Preuß, S., Bennewitz, J., Thaller, G., and Tetens, J. (2019). GWAS for meat and carcass traits using imputed sequence level genotypes in pooled F2-designs in pigs. G3 Genes, Genomes, Genet. 9, 2823–2834. doi:10.1534/g3.119.400452.

Fonseca P, Suarez-Vega A, Marras G, Cánovas A (2020). “GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci.” *GigaScience*, 9(12). doi:10.1093/gigascience/giaa149.

Filipčík, R., Falta, D., Kopec, T., Chládek, G., Večeřa, M., and Rečková, Z. (2020). Environmental factors and genetic parameters of beef traits in fleckvieh cattle using field and station testing. Animals 10, 1–19. doi:10.3390/ani10112159.

Frigo, E., Samorè, A. B., Vicario, D., Bagnato, A., and Pedron, O. (2013). Heritabilities and genetic correlations of body condition score and muscularity with productive traits and their trend functions in Italian Simmental cattle. Ital. J. Anim. Sci. 12, 240–246. doi:10.4081/ijas.2013.e40.

Gast, M.T., Tönjes, A., Keller, M., Horstmann, A., Steinle, N., Scholz, M., Müller, I., Villringer, A., Stumvoll, M., Kovacs, P. and Böttcher, Y., 2013. The role of rs2237781 within GRM8 in eating behavior. Brain Behav., 3(5), pp.495-502. doi:10.1002/brb3.151.

Gershoni, M., Weller, J. I., and Ezra, E. (2021). Genetic and Genome-Wide Association Analysis of YearlingWeight Gain in Israel Holstein Dairy Calves

Grigoletto, L., Ferraz, J. B. S., Oliveira, H. R., Eler, J. P., Bussiman, F. O., Abreu Silva, B. C., et al. (2020). Genetic Architecture of Carcass and Meat Quality Traits in Montana Tropical® Composite Beef Cattle.Front. Genet. 11, 1–13. doi:10.3389/fgene.2020.00123.

Gualdrón Duarte, J. L., Cantet, R. J. C., Bates, R. O., Ernst, C. W., Raney, N. E., and Steibel, J. P. (2014). Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. BMC Bioinformatics 15, 1–11. doi:10.1186/1471-2105-15-246.

Gutiérrez-Gil, B., Ball, N., Burton, D., Haskell, M., Williams, J. L., and Wiener, P. (2008). Identification of quantitative trait loci affecting cattle temperament. J. Hered. 99, 629–638. doi:10.1093/jhered/esn060.

Guzzo, N., Sartori, C., and Mantovani, R. (2018). Heterogeneity of variance for milk, fat and protein yield in small cattle populations: The Rendena breed as a case study. Livest. Sci. 213, 54–60. doi.org/10.1016/j.livsci.2018.05.002.

Guzzo, N., Sartori, C., and Mantovani, R. (2019). Analysis of genetic correlations between beef traits in young bulls and primiparous cows belonging to the dual-purpose Rendena breed. animal 13, 694–701. doi:10.1017/S1751731118001969.

Hardie, L. C., VandeHaar, M. J., Tempelman, R. J., Weigel, K. A., Armentano, L. E., Wiggans, G. R., et al. (2017). The genetic and biological basis of feed efficiency in mid-lactation Holstein dairy cows. J. Dairy Sci. 100, 9061–9075. doi:10.3168/jds.2017-12604.

Helgeland, Ø., Vaudel, M., Juliusson, P. B., Holmen, O. L., Juodakis, J., Bacelis, J., et al. (2019). Genome-wide association study reveals dynamic role of genetic variation in infant anHelgeland, Øyvind, Marc Vaudel, Petur B Juliusson, Oddgeir Lingaas Holmen, Julius Juodakis, Jonas Bacelis, Bo Jacobsson, Haakon Lindekleiv, and Kristian Hveem. “Genome-Wid. Nat. Commun. doi:10.1038/s41467-019-12308-0.

Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38,226–231. doi:10.1007/BF01245622.

Hirano, T., Matsuhashi, T., Kobayashi, N., Watanabe, T., and Sugimoto, Y. (2012). Identification of an FBN1 mutation in bovine Marfan syndrome-like disease. Anim. Genet. 43, 11–17. doi:10.1111/j.1365-2052.2011.02209.x.

Hocquette, J. F., Ortigues-Marty, I., Pethick, D., Herpin, P., and Fernandez, X. (1998). Nutritional and hormonal regulation of energy metabolism in skeletal muscles of meat-producing animals. Livest. Prod. Sci. 56, 115–143. doi:10.1016/S0301-6226(98)00187-0.

Huang, S., He, Y., Ye, S., Wang, J., Yuan, X., Zhang, H., et al. (2018). Genome-wide association study on chicken carcass traits using sequence data imputed from SNP array. J. Appl. Genet. 59, 335–344. doi:10.1007/s13353-018-0448-3.

Huang, W., Guo, Y., Du, W., Zhang, X., Li, A., and Miao, X. (2017). Global transcriptome analysis identifies differentially expressed genes related to lipid metabolism in Wagyu and Holstein cattle. Sci. Rep. 7, 1–11. doi:10.1038/s41598-017-05702-5.

Illius, A. W., Tolkamp, B. J., and Yearsley, J. (2002). The evolution of the control of food intake. Proc. Nutr. Soc. 61, 465–472. doi:10.1079/pns2002179.

Júnior, G. A. F., Costa, R. B., De Camargo, G. M. F., Carvalheiro, R., Rosa, G. J. M., Baldi, F., et al. (2016). Genome scan for postmortem carcass traits in nellore cattle. J. Anim. Sci. 94, 4087–4095. doi:10.2527/jas.2016-0632.

Kaur, M., Kumar, A., Siddaraju, N. K., Fairoze, M. N., Chhabra, P., Ahlawat, S., et al. (2020). Differential expression of miRNAs in skeletal muscles of Indian sheep with diverse carcass and muscle traits. Sci. Rep. 10, 1–11. doi:10.1038/s41598-020-73071-7

Kemter, E., Rathkolb, B., Becker, L., Bolle, I., Busch, D.H., Dalke, C., Elvert, R., Favor, J., Graw, J., Hans,

W. and Ivandic, B., (2014). Standardized, systemic phenotypic analysis of Slc12a1 I299F mutant mice. J. Biom. Sci., 21(1), 1-10. doi:10.1186/s12929-014-0068-0.

Kern, R. J., Lindholm-Perry, A. K., Freetly, H. C., Kuehn, L. A., Rule, D. C., and Ludden, P. A. (2016). Rumen papillae morphology of beef steers relative to gain and feed intake and the association of volatile fatty acids with kallikrein gene expression. Livest. Sci. 187, 24–30. doi:10.1016/j.livsci.2016.02.007.

Kosińska-Selbi, B., Suchocki, T., Egger-Danner, C., Schwarzenbacher, H., Frąszczak, M., and Szyda, J. (2020). Exploring the Potential Genetic Heterogeneity in the Incidence of Hoof Disorders in Austrian Fleckvieh and Braunvieh Cattle. Front. Genet. 11, 1423. doi:10.3389/fgene.2020.577116

Kracht, M., Müller-Ladner, U., & Schmitz, M. L. (2020). Mutual regulation of metabolic processes and proinflammatory NF-κB signaling. J All. Cl. Imm, 146(4), 694-705. doi:10.1016/j.jaci.2020.07.027

Krupová, Z., Krupa, E., Michaličková, M., Wolfová, M., and Kasarda, R. (2016). Economic values for health and feed efficiency traits of dual-purpose cattle in marginal areas. J. Dairy Sci. 99, 644–656. doi:10.3168/jds.2015-9951.

Lapierre, H., Pelletier, G., Abribat, T., Fournier, K., Gaudreau, P., Brazeau, P., et al. (1995). The effect of feed intake and growth hormone-releasing factor on lactating dairy cows. J. Dairy Sci. 78, 804–815. doi:10.3168/jds.S0022-0302(95)76692-9.

Lee, K. Y., Singh, M. K., Ussar, S., Wetzel, P., Hirshman, M. F., Goodyear, L. J., et al. (2015). Tbx15 controls skeletal muscle fibre-type determination and muscle metabolism. Nat. Commun. 6. doi:10.1038/ncomms9054.

Lemos, M. V. A., Chiaia, H. L. J., Berton, M. P., Feitosa, F. L. B., Aboujaoud, C., Camargo, G. M. F., et al. (2016). Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. BMC Genomics 17, 1–16. doi:10.1186/s12864-016-2511-y.

Li, J., and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb). 95, 221–227. doi:10.1038/sj.hdy.6800717.

Liao, S. F., Vanzant, E. S., Harmon, D. L., McLeod, K. R., Boling, J. A., and Matthews, J. C. (2009). Ruminal and abomasal starch hydrolysate infusions selectively decrease the expression of cationic amino acid transporter mRNA by small intestinal epithelia of forage-fed beef steers. J. Dairy Sci. 92, 1124–1135. doi:10.3168/jds.2008-1521.

Liu, Y., Xu, L., Wang, Z., Xu, L., Chen, Y., Zhang, L., et al. (2019). Genomic prediction and association analysis with models including dominance effects for important traits in Chinese simmental beef cattle. Animals 9. doi:10.3390/ani9121055.

Livernois, A. M., Mallard, B. A., Cartwright, S. L., and Cánovas, A. (2021). Heat stress and immune response phenotype affect DNA methylation in blood mononuclear cells from Holstein dairy cows. Sci. Rep. 11, 11371. doi:10.1038/s41598-021-89951-5.

Londoño-Gil, M., Rincón Flórez, J. C., Lopez-Herrera, A., and Gonzalez-Herrera, L. G. (2021). Genome- Wide Association Study for Growth Traits in Blanco Orejinero (Bon) Cattle From Colombia. Livest. Sci. 243, 1–9. doi:10.1016/j.livsci.2020.104366.

Lourenco, D. A. L., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., Misztal, I. (2020). Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in BLUPF90. Genes, 11, 790. doi:10.3390/genes11070790

Mancin E, Tuliozi B, Sartori C, Guzzo N, Mantovani R. Genomic Prediction in Local Breeds: The Rendena Cattle as a Case Study. (2021a) Animals doi: 10.3390/ani11061815

Mancin, E., Lourenco, D. A. L., Bermann, M., and Mantovani, R., Misztal I. (2021b). Accounting for Population Structure and Phenotypes From Relatives in Association Mapping for Farm Animals: A Simulation Study. doi:10.3389/fgene.2021.642065.

Mancin, E., Sartori, C., Guzzo, N., Tuliozi, B., and Mantovani, R. (2021c). Selection Response Due to Different Combination of Antagonistic Milk, Beef, and Morphological Traits in the Alpine Grey Cattle Breed. Animals 11. doi:10.3390/ani11051340.

Mantovani R, Gallo L, Carnier P, Cassandro M and Bittante G 1997. The use of a juvenile selection scheme for genetic improvement of small populations: the example of Rendena breed. Proceedings of the 48th EAAP Annual Meeting,25–28 August 1997, Vienna, Austria.

Marshall, K., Gibson, J. P., Mwai, O., Mwacharo, J. M., Haile, A., Getachew, T., et al. (2019). Livestock Genomics for Developing Countries – African Examples in Practice. Front. Gen. 10, 297. doi:10.3389/fgene.2019.00297

Mao, X., Sahana, G., De Koning, D.-J., and Guldbrandtsen, B. (2016). Genome-wide association studies of growth traits in three dairy cattle breeds using whole-genome sequence data. J. Anim. Sci. 94, 1426. doi:10.2527/jas.2015-9838.

Mateescu, R. G., Garrick, D. J., and Reecy, J. M. (2017). Network Analysis Reveals Putative Genes Affecting Meat Quality in Angus Cattle. 8. doi:10.3389/fgene.2017.00171.

Mazza, S., Guzzo, N., Sartori, C., and Mantovani, R. (2016). Genetic correlations between type and test-day milk yield in small dual-purpose cattle populations: The Aosta Red Pied breed as a case study. J. Dairy Sci. 99, 8127–8136. doi:10.3168/jds.2016-11116.

Miller G. D. (2017). Appetite Regulation: Hormones, Peptides, and Neurotransmitters and Their Role in Obesity. American journal of lifestyle medicine, 13(6), 586–601. https://doi.org/10.1177/1559827617716376

Mrode, R., Ojango, J. M. K., Okeyo, A. M., and Mwacharo, J. M. (2019). Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects. Front. Genet. 10. doi:10.3389/fgene.2018.00694.

Mudadu, M. A., Porto-Neto, L. R., Mokry, F. B., Tizioto, P. C., Oliveira, P. S. N., Tullio, R. R., et al. (2016). Genomic structure and marker-derived gene networks for growth and meat quality traits of Brazilian Nelore beef cattle. BMC Genomics 17, 1–16. doi:10.1186/s12864-016-2535-3.

Olivieri, B. F., Mercadante, M. E. Z., Cyrillo, J. N. D. S. G., Branco, R. H., Bonilha, S. F. M., De Albuquerque, L. G., et al. (2016). Genomic regions associated with feed efficiency indicator traits in an experimental nellore cattle population. PLoS One 11. doi:10.1371/journal.pone.0164390.

Ovaska, U., and Soini, K. (2017). Local Breeds – Rural Heritage or New Market Opportunities? Colliding Views on the Conservation and Sustainable Use of Landraces. Sociol. Ruralis 57, 709–729. doi:10.1111/soru.12140.

Pegolo, S., Momen, M., Morota, G., Rosa, G. J. M., Gianola, D., Bittante, G., et al. (2020). Structural equation modeling for investigating multi-trait genetic architecture of udder health in dairy cattle. Sci. Rep. 10, 7751. doi:10.1038/s41598-020-64575-3.

Peng, H., Chang, B., Lu, C., Su, J., Wu, Y., Lv, P., et al. (2012). Nlrp2, a maternal effect gene required for early embryonic development in the mouse. PLoS One 7. doi:10.1371/journal.pone.0030344.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. doi:10.1086/519795.

Purfield, D. C., Evans, R. D., and Berry, D. P. (2020). Breed- And trait-specific associations define the genetic architecture of calving performance traits in cattle. J. Anim. Sci. 98, 1–18. doi:10.1093/JAS/SKAA151.

Rowan, T., Durbin, H., Seabury, C., Schnabel, R., and Decker, J. (2020). Powerful detection of polygenic selection and evidence of environmental adaptation in US beef cattle. doi:10.1101/2020.03.11.988121.

Ruud, L.E., Pereira, M.M., de Solis, A.J., Fenselau, H. and Brüning, J.C., 2020. NPY mediates the rapid feeding and glucose metabolism regulatory functions of AgRP neurons. Nat. Comm., 11(1), 1-12. doi:10.1038/s41467-020-14291-3.

Samorè, A. B., Fontanesi, L., Fontanesi, L., and Samore, A. B. (2016). Genomic selection in pigs: state of the art and perspectives. doi:10.1080/1828051X.2016.1172034.

Sartori, C., Guzzo, N., Mazza, S., and Mantovani, R. (2018). Genetic correlations among milk yield, morphology, performance test traits and somatic cells in dual-purpose Rendena breed. Animal 12, 906–914. doi:10.1017/S1751731117002543.

Sayols S 2020. rrvgo: a Bioconductor package to reduce and visualize Gene Ontology terms. https://ssayols.github.io/rrvgo

Sbarra, F., Mantovani, R., Quaglia, A., and Bittante, G. (2013). Genetics of slaughter precocity, carcass weight, and carcass weight gain in Chianina, Marchigiana, and Romagnola young bulls under protected geographical indication. J. Anim. Sci. 91, 2596–2604. doi:10.2527/jas.2013-6235.

Schmid, M., and Bennewitz, J. (2017). Invited review: Genome-wide association analysis for quantitative traits in livestock – a selective review of statistical models and experimental designs. Arch. Anim. Breed. 60,335–346. doi:10.5194/aab-60-335-2017.

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., et al. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. BMC Genomics 18,1–25. doi:10.1186/s12864-017-3754-y.

Senczuk, G., Mastrangelo, S., Ciani, E., Battaglini, L., Cendron, F., Ciampolini, R., et al. (2020). The genetic heritage of Alpine local cattle breeds using genomic SNP data. Genet. Sel. Evol. 52, 1–12. doi:10.1186/s12711-020-00559-1.

Srivastava, S., Srikanth, K., Won, S., Son, J.-H., Park, J.-E., Park, W., et al. (2020). Haplotype-Based Genome-Wide Association Study and Identification of Candidate Genes Associated with Carcass Traits in Hanwoo Cattle. Genes (Basel). 11. doi:10.3390/genes11050551.

Sun, W., Zhao, X., Wang, Z., Chu, Y., Mao, L., Lin, S., et al. (2019). Tbx15 is required for adipocyte browning induced by adrenergic signaling pathway. Mol. Metab. 28, 48–57. doi:10.1016/j.molmet.2019.07.004.

Sutera, A. M., Moscarelli, A., Mastrangelo, S., Sardina, M. T., Di Gerlando, R., Portolano, B., et al. (2021). Genome-Wide Association Study Identifies New Candidate Markers for Somatic Cells Score in a Local Dairy Sheep. Front. Genet. 12, 409. doi:10.3389/fgene.2021.643531.

Taye, M., Lee, W., Jeon, S., Yoon, J., Dessie, T., Hanotte, O., et al. (2017). Exploring evidence of positive selection signatures in cattle breeds selected for different traits. Mamm. Genome 28, 528–541. doi:10.1007/s00335-017-9715-6.

Tiezzi, F., and Maltecca, C. (2015). Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. Genet. Sel. Evol. 47, 24. doi:10.1186/s12711-015-0100-1.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. J. Dairy Sci. 91, 4414–4423. doi:10.3168/jds.2007-0980.

Vanvanhossou, S. F. U., Scheper, C., Dossa, L. H., Yin, T., Brügemann, K., and König, S. (2020). A multi- breed GWAS for morphometric traits in four Beninese indigenous cattle breeds reveals loci associated with conformation, carcass and adaptive traits. BMC Genomics 21, 783. doi:10.1186/s12864-020-07170-0.

Veselá, Z., Vostrý, L., and Šafus, P. (2011). Linear and linear-threshold model for genetic parameters for SEUROP carcass traits in Czech beef cattle. Czech J. Anim. Sci. 56, 414–425. doi:10.17221/1292-cjas.

Vitezica, Z. G., Aguilar, I., Misztal, I., and Legarra, A. (2011). Bias in genomic predictions for populations under selection. Genet. Res. 93, 357–366. doi:10.1017/S001667231100022X..

Whalen, A., and Hickey, J. M. (2020). AlphaImpute2: Fast and accurate pedigree and population based imputation for hundreds of thousands of individuals in livestock populations. bioRxiv. doi:10.1101/2020.09.16.299677.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New Tork. Retrieved from https://ggplot2.tidyverse.org.

Yang, Y., Wang, H., Li, G., Liu, Y., Wang, C., and He, D. (2020). Exploring the genetic basis of fatty liver development in geese. Sci. Rep. 10, 1–12. doi:10.1038/s41598-020-71210-8.

Yin, T., and König, S. (2018). Genetic parameters for body weight from birth to calving and associations between weights with test-day, health, and female fertility traits. J. Dairy Sci. 101, 2158–2170. doi:10.3168/jds.2017-13835.

Yin, T., and König, S. (2019). Genome-wide associations and detection of potential candidate genes for direct genetic and maternal genetic effects influencing dairy cattle body weight at different ages. Genet. Sel. Evol. 51, 1–14. doi:10.1186/s12711-018-0444-4.

Zhang, J., Tan, J., Zhang, C., Wang, Y., Chen, X., Lei, C., et al. (2021). Research on associations between variants and haplotypes of Aquaporin 9 (AQP9) gene with growth traits in three cattle breeds. Anim. Biotechnol. 32, 185–193. doi:10.1080/10495398.2019.1675681.

Zhuang, Z., Xu, L., Yang, J., Gao, H., Zhang, L., Gao, X., et al. (2020). Weighted Single-Step Genome-Wide Association Study for Growth Traits in Chinese Simmental Beef Cattle. Genes (Basel). 11, 189. doi:10.3390/genes11020189

G.R. Wiggans, T.A. Cooper, P.M. VanRaden, K.M. Olson, M.E. Tooker. (2012). Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation1, J. of Dairy Sci. 95(3), 1552-1558. doi:10.3168/jds.2011-4985.

12. GENERAL CONCLUSION

The current studies covered some possible enhancement of selection plans in local breeds through genetic and genomic selection and characterization. Following this general objective, a first suitable selection approach has been conducted on Reggiana and Alpine Grey cattle breeds by the implementation of classical selection indexed accounting for several antagonistic traits. The specific chapter of the thesis on this topic demonstrated that a genetic selection that considered functional, morphological, beef traits besides milk production is feasible and allows the preservation of the breed identity. Indeed, using the restriction selection index is a good trade-off that guarantees the selection of economically important traits, the preservation of “peculiar” traits belonging to the local breeds appreciated by breeders. In Reggiana, we also demonstrated the feasibility to account for GxE in routinary selection plans. Further studies included in the present thesis demonstrated the possibility to improve EBVs’ accuracy through the uses of genomic data in Rendena breeds, despite the reduced size of genotyped animals. We demonstrated that the integrations of different sources of genomic/phenotype information (animals with performance test data and their relatives) can considerably increase the accuracy of EBVs for young proven bulls. However, it must be pointed out that to preserve this accuracy gap it is necessary to continuously genotype animals over years.

Additionally, we provided an ad-hoc/naïve approach that constructed a G matrix using only the most informative SNPs. This has ensured a further increase of EBVs’ accuracy, especially when Elastic Net algorithms were used. Identifying new strategies that increase EBVs’accuracy in local breeds is a necessity as these breeds are small populations with potentially large numbers of traits to include in the breeding index.

In the last chapter, we demonstrated that single-step GWAS can be a suitable strategy that can account for population structure and could be considered a straightforward method for an association analysis when only a fraction of the population is genotyped and/or when phenotypes are available on non-genotyped relatives. The results of this ssGWAS led to the detection of a variety of both known and new genes. These can broaden characteristics allowing for lower environmental impact, better survival, and higher production in harsh environments. Such aspects are crucial to avoid the negative effects of intensive beef farming on a world scale. Furthermore, theese knowledges can greatly aid efforts to map the genomic complexity of traits

of interest and to make appropriate breeding decisions. Additionally, such studies could help uncover several novel genes and pathways associated with beef traits measured at different times that are due to the less intensive selection, as occurred for other breeds. In general, both classical genetic approach and genomic information have been proven to be useful tools in understanding the genetic architecture of many traits belonging to dual-purpose cattle breeds and all these strategies could become interesting approaches for a continuous selection process of these small cattle populations, allowing their maintenance and the maintenance of all cultural traditions and history linked to them.