Assessment of the accuracies of estimated breeding values and genomic enhanced breeding values in selection of Bonsmara cattle

by

Antoinette van Wyk

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae Agriculturae
MSc(Agric) Animal Science: Animal Breeding and Genetics

in the Faculty of Natural and Agricultural Sciences
Department of Animal Sciences
University of Pretoria
Pretoria

November 2020

Assessment of the accuracies of estimated breeding values and genomic enhanced breeding values in selection of Bonsmara cattle

by

Antoinette van Wyk

Supervisor: Prof E. van Marle-Köster

Co-supervisors: Dr R. R. van der Westhuizen and Prof D. P. Berry

Department: Animal Science

Degree: MSc(Agric) Animal Breeding and Genetics

Acknowledgements

The completion of this dissertation would not have been possible without God and the strength He gave me to push through all obstacles along the way.

This dissertation is dedicated to my dad who always believed in me and played an integral role in my success.

I would like to thank the following people and organisations for support during this study:

Thank you, Prof Esté van Marle-Köster for believing in me, always being available to assist me, always making time for a conversation and truly investing in me as a person.

Thank you, Dr Bobbie van der Westhuizen and Prof Donagh Berry, for always assisting me and guiding me in the right direction.

Thank you, mom for always supporting me, believing in me and encouraging me when I needed it most.

Thank you, Jacques for your endless support, love and encouragement. I could not have done this without you.

Thank you to SA Stud Book for allowing me to use your computer facilities and the South African Bonsmara Breeders' Society for allowing me to use your data.

"Be still and know that I am God" – Psalm 46:10

Declaration

I, the undersigned, hereby declare that this thesis, submitted for the MSc(Agric) Animal Science: Animal Breeding and Genetics degree at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at any other University.

Signature: 
Name: Antoinette van Wyk
Date: November 2020

Abstract

Traditionally, the selection of beef cattle was based on the quantitative animal breeding theory and principles. The development of estimated breeding values (EBVs) resulted in accelerated genetic progress in most traits of economic importance. The advent of molecular technology, completion of the bovine genome sequence and single nucleotide polymorphism (SNP) marker discoveries facilitated the use of genomic selection as a selection tool which increased breeding value accuracies. In South Africa, the Beef Genomic Programme enabled the establishment of a reference population for the Bonsmara breed facilitated by large datasets containing performance data due to mandatory performance recording in the breed and availability of biological samples for genotyping. In 2017, the first genomic enhanced breeding values (GEBVs) were made available to breeders. This study aimed to assess the accuracies of EBVs and GEBVs in the selection of Bonsmara cattle for growth traits. The study was conducted in two parts; the dataset for analysis I and II consisted of 4128 and 4189 genotypes, respectively and 2 018 052 phenotypic records. In analysis I, EBVs and GEBVs were estimated for 4128 animals and were correlated to determine if including genomic information in the breeding value estimates influenced the ranking of the animals and breeding value accuracies. In analysis II, a forward validation scheme was applied using validation populations, which consisted of the youngest 500 animals with phenotypic and genotypic information for each trait. Traditional parental averages (TPAs) without SNP information of the parents, genomic-based parental averages (GPAs) with SNP information of the parents, parental averages with genomic information (PAGs) which include SNP information of the parents and the animals themselves, EBVs and GEBVs were estimated. These breeding values were correlated to determine the predictive ability of the breeding value models. In analysis I, the ranking of the animals based on the GEBVs differed from the EBV rankings. The increase in the average GEBV accuracy was between 2.7% to 5.3% compared to the EBVs. In analysis II, the predictive ability of the GEBV models were 8.4% to 78.1% higher compared to the TPA models for all the traits. Additionally, the predictive ability of the PAG models were 5.3% to 11.0% higher compared to the TPA and GPA models for height, direct and maternal weaning weight. The results indicated that genomic information plays an important role in the breeding value estimation and should be included in routine genomic evaluations for growth traits in the Bonsmara breed. This study confirmed the value of genomic information in the breeding value estimation for Bonsmara cattle.

Table of Contents

List of figures

List of tables

Chapter 2: Literature review

Chapter 3: Materials and methods

Chapter 4: Results

List of abbreviations

ADG	Average Daily Gain
AI	Artificial Insemination
A-matrix	Additive relationship matrix
BGP	Beef Genomic Programme
BLUP	Best Linear Unbiased Prediction
DGP	Dairy Genomic Programme
DNA	Deoxyribonucleic acid
EBV/s	Estimated Breeding Value/s
GBLUP	Genomic Best Linear Unbiased Prediction
GEBV/s	Genomic Enhanced Breeding Value/s
GGP-HD	GeneSeek® Genomic Profiler Bovine HD™
G-matrix	Genomic relationship matrix
GPA	Genomic-based parental average
GS	Genomic Selection
h^2	Narrow sense heritability
PAG	Parental Average with Genomic information
PCR	Polymerase Chain Reaction
PEV	Prediction Error Variance
LD	Linkage Disequilibrium
MACE	Multiple Across Country Evaluation
MWW	Maternal Weaning Weight
QTL	Quantitative Trait Loci
r	Pearson correlation coefficient
RAPD	Random Amplification of Polymorphic DNA
REML	Restricted Maximum Likelihood
RFLP	Restriction Fragment Length Polymorphism
RMRDT	Red Meat Research and Development Trust
S	Selection differential
SA	South African
SNP/s	Single Nucleotide Polymorphism/s
ssGBLUP	single-step Genomic Best Linear Unbiased Prediction Method
TIA	Technology Innovation Agency
TPA	Traditional Parental Average
WW	Direct Weaning Weight
YD/s	Yield Deviation/s

Chapter 1: Introduction

1.1 Introduction

The South African population is expected to increase by approximately 16 million people, reaching a population size of 72 million by 2050 (U. N., 2019). It has been noted in a recent study by Queenan *et al.* (2020) that the expenditure on livestock-derived foods has increased by 39% over the past two decades, emphasising the growth in the demand for animal protein and production. In South Africa, the beef industry is one of the principal industries in the agricultural sector due to its important role in food production and support of livelihoods (Nevondo *et al.*, 2019). Only 20% of the agricultural land in South Africa is suitable for crop production, with the majority of land (approximately 80%) only suited for grazing by livestock (Department of Agriculture Forestry and Fisheries, 2017).

The South African beef industry consists of two sectors, the developed (i.e. commercial) sector and the developing sector, which include market-oriented farmers and subsistence farming (Department of Agriculture Forestry and Fisheries, 2017). The commercial beef production sector is divided into stud farmers (i.e. seed stock), commercial farmers, feedlots and abattoirs, similar to the beef industry in the United States (Garrick & Golden, 2009; Department of Agriculture Forestry and Fisheries, 2017). The stud and commercial farmers represent the cow-calf sector that consists of the stud farmers responsible for producing sires for breeding and the commercial farmers that focus on calf production using sires that are generated by the stud farmers (Garrick & Golden, 2009). The feedlot and abattoir industry also play an important role in the beef value chain as the majority of South African beef is produced by feeding and finishing of weaner calves in feedlots (Vermeulen *et al.*, 2008; Department of Agriculture Forestry and Fisheries, 2017). Feedlots form part of intensive production systems which are widespread and usually found near metropolitan markets and companies that supply the required animal feed (Meissner *et al.*, 2013).

Several beef cattle breeds are used to produce high quality meat suitable for finishing in feedlots. In South Africa, there are approximately 30 beef breeds which include British, European, composite and indigenous breeds (Van Marle-Köster *et al.*, 2013). The seven most important beef cattle breeds in the feedlot industry include Bonsmara, Hereford, Simmentaler, Limousin, SA Angus, Beefmaster and Drakensberger (Scholtz *et al.*, 2008). Bonsmara cattle have been dominating the feedlot industry with up to 15.9% of weaners and both Hereford and Simmentaler are prevalent in crossbred calves (Scholtz *et al.*, 2008). South Africa has several abattoirs forming part of the beef value chain for humane

slaughtering and delivering safe beef to the market (Red Meat Producers' Organisation, 2017).

The overall breeding objectives of the beef sector, including seed stock and commercial farmers, are usually directed at the improvement of fertility and growth efficiency (Garrick, 2011). Breeding objectives need to be clearly defined with specific and measurable selection criteria in order to ensure genetic improvement in the beef sector (Garrick & Golden, 2009). Breeding objectives may include a range of aspects and traits and differ between production systems (Amer *et al.*, 2001). South African beef farmers had access to national animal recording schemes since the early 1960's, followed by the development of estimated breeding values (EBVs) in the 1980's which resulted in genetic improvement in most of the production traits with high heritability (Bergh, 1999; Mokoena *et al.*, 1999).

Since the completion of the bovine genome in 2009 (Matukumalli *et al.*, 2009), followed by single nucleotide polymorphism (SNP) marker discovery, high density commercial arrays have been developed (Van Tassell *et al.*, 2008; Eck *et al.*, 2009; Seidel, 2010). New possibilities arose for using deoxyribonucleic acid (DNA) information in genetic evaluations for cattle (Meuwissen *et al.*, 2001; Schaeffer, 2006; Harris *et al.*, 2008; Garrick, 2009; VanRaden *et al.*, 2009; Calus, 2010; Koopaee & Koshkoiyeh, 2014). Genomic selection (GS) entails using SNP genotyping information together with phenotypic and pedigree information to estimate genomic enhanced breeding values (GEBVs) which provide an additional selection tool (Meuwissen *et al.*, 2001; Calus, 2010; Goddard, 2012). Since the first application of GS in a simulation study in dairy cattle in 2006 (Schaeffer, 2006), GS has been widely implemented in both dairy and beef cattle populations in most of the developed countries in the world (Koivula *et al.*, 2012). A major advantage of GS is the increase in the accuracy of breeding values when genomic information is included in genetic evaluations (Berry *et al.*, 2016).

The success of GS in developed countries, led to the establishment of a state-funded project in South Africa in 2015, namely the Beef Genomic Programme (BGP), in order to establish training/reference populations for beef breeds (Van Marle-Köster & Visser, 2018a). More than 7000 genotypes were generated over three years and 16 breed societies participated in the BGP funded by the Technology Innovation Agency (TIA) (Van Marle-Köster & Visser, 2018a). The Bonsmara breed is a composite beef breed that was developed with the intention to improve growth traits and to enhance adaptation to South African climatic conditions (Bonsma, 1980). Registering and recording the performance of Bonsmara cattle are mandatory in South Africa, which has led to a large population of approximately 108 198

registered Bonsmara animals with available phenotypic and pedigree information (Bosman *et al.*, 2017; SA Stud Book, 2019).

Animals of both sexes were genotyped in the BGP, which included high impact sires with EBVs for maternal weaning weight (MWW) with ≥ 65% accuracy (Bosman *et al.*, 2017). Additionally, the cows were chosen based on their age (older than 6 years) and the number of calves (≥ 3) that they have weaned with recorded weights (Bosman *et al.*, 2017). The large phenotypic and pedigree datasets that were available, for several production traits such as growth and fertility, together with the genotypic information from the BGP allowed the establishment of a reference population for the Bonsmara breed (Bosman *et al.*, 2017). The reference population made it possible to include genomic information in the genetic evaluations to estimate GEBVs (SA Stud Book, 2017).

1.2 Aim

Initially, only EBVs were estimated for the Bonsmara breed and were useful in selecting high quality individuals for production traits (SA Stud Book, 2017). However, when research indicated the value of including genomic information in the breeding value estimation, the BGP was established to genotype beef breeds in South Africa (Van Marle-Köster & Visser, 2018a). The BGP aimed to gather enough genotypes to establish reference populations for the beef breeds (Van Marle-Köster & Visser, 2018a). The Bonsmara Breed Society further invested additional funds to genotype as many animals as possible in order to establish a large enough reference population to estimate GEBVs for the breed (Bosman *et al.*, 2017). The Bonsmara breed had large datasets with phenotypic and pedigree information available due to the mandatory performance recording of the animals (Bosman *et al.*, 2017). Therefore, once the genotypes were obtained, a reference population for the Bonsmara breed was established and resulted in the estimation of GEBVs for the Bonsmara breed (SA Stud Book, 2017). Since 2017, GEBVs have been provided to Bonsmara breeders for application in their herds (SA Stud Book, 2017). This study is the first attempt to evaluate the effect of genomic information on the breeding value estimation in Bonsmara cattle.

The aim of this study was to assess the accuracies of estimated breeding values and genomic enhanced breeding values in the selection of Bonsmara cattle.

In order to achieve the aim, the following objectives were set:

Analysis I:

1. Determine whether including genomic information in the breeding value estimation influences the ranking of Bonsmara animals with phenotypic and genotypic information for growth traits (direct weaning weight, maternal weaning weight, average daily gain and height).

2. Determine whether including genomic information in the breeding value estimation influences the accuracy of the breeding values of Bonsmara animals for growth traits.

Analysis II:

1. Determine whether including genomic information in the breeding value estimation influences the predictive ability of the breeding values of Bonsmara animals for growth traits.

Chapter 2: Literature review

2.1 Introduction

Selective breeding in livestock plays an important role in improving economically important traits (Meuwissen *et al.*, 2001; Goddard & Hayes, 2007). In the past, selection for economically important quantitative traits was performed using phenotypic information from the individuals and their relatives and making use of pedigree information (Meuwissen *et al.*, 2001; Goddard & Hayes, 2007). The development of DNA technology lead to the discovery of DNA markers that have been used to identify multiple genes which affect traits of interest in livestock (Koopaee & Koshkoiyeh, 2014). Mapping of the bovine genome in 2009 (Matukumalli *et al.*, 2009) resulted in the discovery of single-gene traits, quantitative trait loci (QTL) and genomic regions that affect quantitative traits (Van Marle-Köster *et al.*, 2013; Koopaee & Koshkoiyeh, 2014; Rothschild & Plastow, 2014).

The discovery of SNPs and the development of commercial SNP arrays for cattle (Van Tassell *et al.*, 2008; Eck *et al.*, 2009; Seidel, 2010), resulted in genotyping of thousands of cattle. The development of appropriate statistical methods facilitated the application of genomic selection (GS) in dairy and beef cattle (Meuwissen *et al.*, 2001; Schaeffer, 2006; Harris *et al.*, 2008; Garrick, 2009; VanRaden *et al.*, 2009; Calus, 2010; Koopaee & Koshkoiyeh, 2014). This literature review provides a brief overview on the development of quantitative animal breeding and discusses the role genomics has played in the advancement of GS in cattle.

2.2 Brief overview of the history and development of quantitative and molecular genetics focusing on animal breeding

Quantitative genetic principles laid the foundation for the development of statistical models which enabled the estimation of breeding values for livestock (Patterson & Thompson, 1971; Henderson, 1984; Meuwissen *et al.*, 2001). However, studies suggested that the breeding value accuracy could be increased by including SNP genotyping information in the breeding value estimation (Meuwissen *et al.*, 2001; Schaeffer, 2006). Therefore, the developments in molecular genetics that made SNP genotyping possible were essential for the estimation of

GEBVs for livestock and the implementation of GS in cattle (Gray *et al.*, 2000; Harris *et al.*, 2008; Garrick, 2009, 2011; Matukumalli *et al.*, 2009; VanRaden *et al.*, 2009; Calus, 2010; Goddard, 2012; Koopaee & Koshkoiyeh, 2014).

2.2.1 Quantitative principles

Dr Jay Lush established the field of animal breeding (Lush, 1935, 1936; Hill, 2014). He is known for explaining the concept of additive genetic variance and the role of the environment in the observed differences among traits (Lush, 1935, 1936; Hill, 2014). The term 'narrow sense heritability', which indicated the ratio of additive genetic variance to the total phenotypic variance within a population, was introduced in 1936 (Lush, 1936; Bell, 1977), followed by the "breeder's equation", consisting of narrow sense heritability (h^2) and the selection differential (S) which could be applied to predict the response to selection (Lush, 1937).

Hazel and Lush, as early as 1942, compared three selection methods to determine the best method for efficient selection of individual traits (Hazel & Lush, 1942). These methods included selection for one trait at a time, selection for a total score and selection using independent culling levels (Hazel & Lush, 1942). The selection method that assigned an appropriate weight to each trait i.e. the selection for total score was the most efficient (Hazel & Lush, 1942). However, the standard index theory used for selection could only be used in the case where all the animals were reared under the same environmental conditions, due to the assumption that the selection method already corrected for the environmental effects (Hazel & Lush, 1942). Therefore, the selection method could not be applied to bulls that had daughters distributed among many different herds and environments as a result of artificial insemination (AI) (Hazel & Lush, 1942). Additionally, Hazel (1943) explained the role of genetic correlations on selection and application thereof in the calculation of multi-trait selection indices.

Henderson postulated the inclusion of fixed effects such as herd or season and random effects in the breeding value models, which lead to the initial development of the best linear unbiased prediction (BLUP) method (Henderson, 1948; Hill, 2014). Estimating variance and covariance components of the random effects in the models were necessary to estimate breeding values for selection candidates (Hofer, 1998). However, the variance component analyses at the time mainly focussed on biological data with a balanced structure (Crump, 1946, 1951; Eisenhart, 1947). Henderson aimed to design efficient testing and selection programs for the New York State dairy bulls where the data had an unbalanced structure

and required high quality variance estimates of the production records of the daughters (Henderson, 1953). Therefore, Henderson continued to develop mixed model methods to estimate variance and covariance components for data with an unbalanced structure, as is commonly the case with livestock data (Henderson, 1953). The development of computer software programs made the use of these methods possible, due to the ability of the program to handle large datasets for the estimation of genetic parameters (Harvey, 1960, 1977). Robertson & Rendel (1954) introduced the concept of contemporary comparisons that grouped animals according to similar environmental and management conditions which influenced the phenotypes of the animals. Contemporary comparisons resulted in the determination of breeding values with a higher accuracy (Robertson *et al.*, 1956).

Patterson and Thompson (1971) developed the restricted maximum likelihood (REML) method, which became the preferred theoretical method to estimate variance and covariance components in animal breeding (Meyer, 1989; Hofer, 1998). However, REML was not frequently implemented in practice due to it having extensive computational requirements and complex algorithms (Meyer, 1986, 1989). Therefore, Meyer (1986) developed a simple model that could facilitate the use of large amounts of data to estimate variance and covariance components between and within genetic groups. The development of computational resources and specialized algorithms at the time greatly facilitated the use of REML (Meyer, 1989).

Initially, REML was used most often in sire models, where dairy cattle data was analysed using information of the progeny to gain information on half of the progenies' sire breeding values (Meyer, 1989). However, the animal model was theoretically superior in estimating variance and genetic parameters compared to the sire model (Everett *et al.*, 1979; Schaeffer, 1983; Hudson & Schaeffer, 1984; Sun *et al.*, 2009) and therefore animal models started to gain more interest (Meyer, 1989). These models could account for all records and was therefore able to provide insight on the measured animal's additive genetic merit (Meyer, 1989). Parents without phenotypic information could be included in the animal models, which ensured that all relationship information could be taken into account (Meyer, 1989). This led to Graser *et al.* (1987) suggesting a derivative-free REML algorithm using an animal model for univariate analyses that could use data from thousands of animals from large selection experiments to estimate the additive genetic and error variance. Thereafter, a derivative-free REML approach was developed for the estimation of variance components including the animals' additive genetic merit and random effects for univariate and extended multi-variate analyses (Meyer, 1989, 1991).

In 1984, Henderson presented the BLUP method, based on a linear model that uses phenotypic and pedigree information to calculate the inverse of the additive relationship matrix (A-matrix) to estimate breeding values for animals (Calus, 2010; Forni *et al.*, 2011). The A-matrix uses pedigree data to calculate the probability that gene pairs are identical by descent (Wright, 1922; Forni *et al.*, 2011). In traditional animal breeding models the base population is assumed to be unrelated to each other and share no common alleles or genes (VanRaden, 2007). The A-matrix therefore assigns a 0 to off-diagonal elements indicating no relationship between individuals and a 1 for diagonal elements where the individual is compared to itself (VanRaden, 2007). Once the A-matrix is calculated, it is used to estimate EBVs using the traditional BLUP method (Meuwissen *et al.*, 2016).

The basis of the BLUP linear model is expressed as follows,

$$\mathbf{y = X\beta + Zu + e},$$

with **X** representing an incident matrix (*n x p*) that is known and fixed, **β** representing a fixed vector (*p x 1*), **Z** representing an incidence matrix (*n x q*) and **u** representing a random vector (*q x 1*) with null means (Henderson, 1984). **Xβ** defines the fixed effects and **Zu** defines the random effects (VanRaden, 2008).

2.2.2 Molecular developments

Since 1980 several technological developments made it possible to study the genome. The development of the polymerase chain reaction (PCR) made a major contribution to the accelerated amplification of DNA (Mullis *et al.*, 1986) and use of several DNA markers which included restriction fragment length polymorphism (RFLP) markers, random amplification of polymorphic DNA (RAPD) markers and microsatellite markers (Botstein *et al.*, 1980; Weller *et al.*, 1984; Williams *et al.*, 1990). The sequencing of the entire human genome by the human genome sequencing consortium led the way for other genomes to be sequenced such as the mouse genome in 2002 (Mouse Genome Sequencing Consortium, 2002) and the chicken genome in 2004 (Hillier *et al.*, 2004).

The bovine genome assembly was completed in 2009 and the sequence was based on a Hereford cow, which was partially inbred, as well as her sire (Matukumalli *et al.*, 2009). The sequence provided insight on the cattle genome and was useful for comparison with other species and studying the evolution of cattle (Matukumalli *et al.*, 2009). A study indicated that the cattle genome consists of a minimum of 22 000 genes and approximately 14 000

orthologs are shared between cattle and seven mammalian species (Elsik *et al.*, 2009). Once the bovine genome sequence became available several studies were performed to identify SNPs in the genome (Van Tassell *et al.*, 2008; Eck *et al.*, 2009; Seidel, 2010).

SNPs had several advantages as opposed to using microsatellite markers which lead to SNPs becoming the preferred genotyping method (Gray *et al.*, 2000; Garrick, 2011). SNPs are usually bi-allelic due to mutation bias and single nucleotide substitutions having a low frequency at the origin of SNPs (Vignal *et al.*, 2002). In many genomes these nucleotide changes are found every 300 to 1000 base pairs (Aitken *et al.*, 2004). A base pair can be classified as a SNP when the allele with the lowest frequency has a frequency greater than or equal to 1% (Vignal *et al.*, 2002).

The bi-allelic nature of SNPs lead to low individual locus information and therefore many more SNPs compared to microsatellites need to be used in population analyses (Aitken *et al.*, 2004). Depending on the frequency of the alleles, when SNPs are used for parentage analyses and individual identification, it is estimated that 30-50 SNPs are needed compared to only 10-15 microsatellites to allow for equal statistical power (Aitken *et al.*, 2004). SNPs are genotyped using high-throughput methods such as allele-specific hybridization, DNA arrays and pyrosequencing (Koopaee & Koshkoiyeh, 2014). SNPs are better suited for high-throughput genetic analysis compared to microsatellites and the inheritance of SNPs is more stable compared to other DNA markers. Therefore, it is more appropriate to be used as long time selection markers (Koopaee & Koshkoiyeh, 2014). SNPs are highly abundant and can therefore be found closer to the loci of interest compared to other DNA markers (Koopaee & Koshkoiyeh, 2014).

The Bovine Hapmap consortium discovered that SNPs occur approximately every 300 base pairs in *Bos indicus* and every 700 base pairs in *B. taurus* cattle breeds (The Bovine HapMap Consortium, 2009; Seidel, 2010). It is estimated that the *B. taurus* genome has approximately 4 milllion SNPs (The Bovine HapMap Consortium, 2009). Due to SNPs being discovered in the bovine genome, Affymetrix and Illumina were able to develop several SNP arrays for *B. indicus* and *B. taurus* cattle breeds (Table 2.1).

Table 2.1 Available bovine SNP arrays with various numbers of SNPs from Illumina and Affymetrix.

Platform	SNP chip name	Number of SNPs
Affymetrix ®	Axiom ® Genome-Wide BOS 1	>640 000
Illumina ®	BovineSNP50	53 714
	BovineHD BeadChip	> 777 000
	BovineLD Genotyping BeadChip	200
	GGP Bovine 150K Array	> 134 000
	GGP Bos Indicus HD Array	74 000
	GGP Bovine LD Array	26 000

(https://www.affymetrix.com/products_services/arrays/specific/axiom_gwas_bovine.affx; https://www.illumina.com/products/all-products.html) – Accessed October 2019

High-density SNP markers are used in genomic prediction in livestock, due to higher linkage disequilibrium (LD) being observed in physically close loci compared to distant loci and therefore family structures are not required (Garrick, 2011). Parentage validation, breed assignment and screening for lethal mutations or congenital defects in beef breeding programs were facilitated by the use of genomic information (Berry *et al.*, 2016). Multiplex SNP genotyping technology can be used to improve animal production, health and selection accuracy within genetic improvement programs that use a modified genome-wide approach known as genome-wide selection (Meuwissen *et al.*, 2001). SNP markers were therefore the most appropriate for application in genomic selection.

2.2.3 Genomic selection

Genomic selection (GS) can be defined as the prediction of breeding values of animals using dense marker maps across the genome (Meuwissen *et al.*, 2001; Calus, 2010; Goddard, 2012). Meuwissen & Goddard (1996) predicted that including marker information in the breeding value estimation as opposed to only using phenotypic information could increase genetic gain by 8% to 38%, which led to Meuwissen *et al*. (2001) introducing the concept of GS. However, GS could not be implemented at the time, due to it requiring a large number of SNPs and an affordable manner to genotype animals (Meuwissen *et al.*, 2001; Goddard & Hayes, 2007; Calus, 2010; Goddard, 2012; Boichard *et al.*, 2016). The sequencing of livestock genomes led to thousands of SNPs being available for livestock species (Hillier *et al.*, 2004; Matukumalli *et al.*, 2009). The development of commercial SNP arrays made it affordable to genotype animals, resulting in the inclusion of SNPs in breeding

value predictions and GS (Meuwissen *et al.*, 2001; Goddard & Hayes, 2007; Calus, 2010; Goddard, 2012; Boichard *et al.*, 2016).

GS has several advantages such as increasing the rate of genetic improvement and enabling breeders to preselect animals with high genetic potential based on genomic information which, in turn, reduces the costs related to progeny testing (Meuwissen *et al.*, 2001; Schaeffer, 2006). It is a valuable method used in animal breeding that provides a higher accuracy (~0.31) compared to traditional pedigree indices (Meuwissen *et al.*, 2001; Calus, 2010; Goddard, 2012). GS is the most advantageous for the selection of young animals with no phenotypic information for the trait at a young age or when the phenotype is difficult and costly to measure (Goddard, 2012). For example, sex-limited traits such as milk production can only be measured in females and meat quality traits such as meat tenderness can only be observed once the animal is slaughtered (Goddard, 2012).

GS is based on a statistical method that requires thousands of SNP markers across the genome without identifying the genes or regions that are responsible for the variation in the trait (Goddard & Hayes, 2007; Goddard, 2012). The statistical method used for GS entails estimating the genetic effects of each marker and using this information together with phenotypic and genotypic information to predict the overall breeding value of any animal (Boichard *et al.*, 2016). A reference population is required to estimate the marker effects and this information is applied to selection candidates with genotypic marker information with no phenotypic information (Figure 2.1) (Calus, 2010; Boichard *et al.*, 2016).

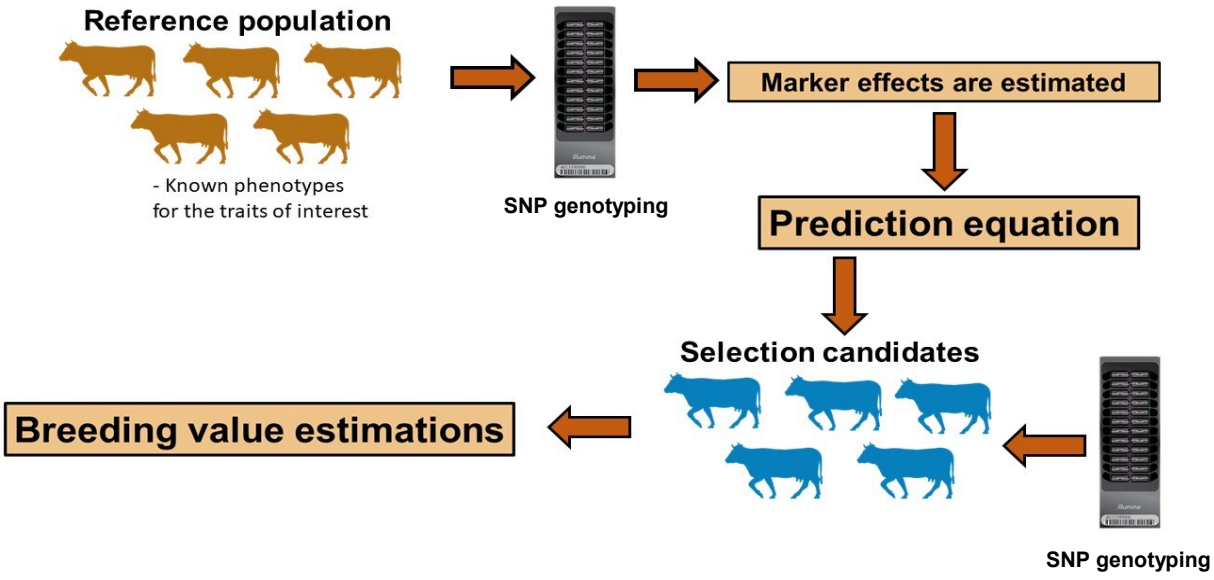

Figure 2.1. A brief overview of the genomic selection process (Adapted from Boichard *et al.*, 2016).

The reference population consists of animals with known genotypes and reliable phenotypic information which can be obtained from several sources such as phenotypic recording, deregressed proofs, daughter yield deviations or the average performance of the offspring (González-Recio *et al.*, 2008; Berry *et al.*, 2009; VanRaden *et al.*, 2009; Calus, 2010; Boichard *et al.*, 2016). The reference population should ideally be comprised of animals that represent all the phenotypes and genotypes that are present in the breed for the traits of interest (Calus, 2010). The characteristics of the reference population, which include the number of animals and markers as well as the heritability of the phenotype of interest, influence the accuracy of the genomic predictions (Calus, 2010). The accuracy of GS increases as the size of the reference population increases (Goddard, 2012). The reference population should consist of approximately 1000 animals or more, depending on the heritability of the trait of interest (Blasco & Toro, 2014). In order to obtain the same GS accuracies for lowly and highly heritable traits, a larger reference population is required for low heritability traits compared to highly heritable traits when the GEBV is estimated using only genomic information (Oldenbroek & Van der Waaij, 2015) (Figure 2.2).

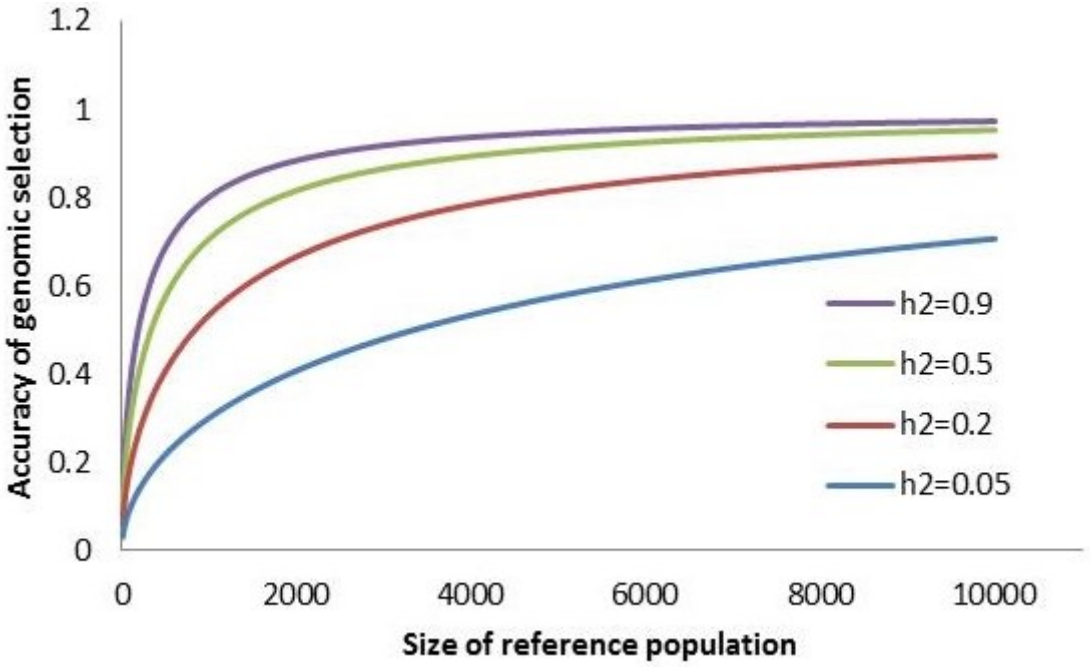

Figure 2.2. The relationship between the size of the reference population and the accuracy of genomic selection for lowly, moderately and highly heritable traits (Oldenbroek & Van der Waaij, 2015).

The predictive ability of the genomic predictions are influenced by the number of SNPs that is used to estimate the marker effects in the reference population (Blasco & Toro, 2014). An increase in the number of SNPs that are used to estimate marker effects leads to an

increase in the predictive ability of genomic predictions (Blasco & Toro, 2014). The standard amount of SNPs that are currently used are 50 000 SNPs, due to a insignificant increase in the predictive ability of genomic predictions when more than 50 000 SNPs are used (VanRaden *et al.,* 2011; Blasco & Toro, 2014).

There are a variety of statistical methods that can be used for GS such as the non-linear Bayesian approach and an extension of the BLUP method known as the Genomic Best Linear Unbiased Prediction (GBLUP) method (Table 2.2) (Goddard, 2012; Meuwissen *et al.*, 2016).

Table 2.2. A comparison between the non-linear Bayesian method and linear Genomic Best Linear Unbiased Prediction (GBLUP) method (Goddard, 2012; Meuwissen *et al.*, 2016).

Methods	Appropriate use	Assumption	Breeds	Examples
Bayesian	Small number of QTL with large effects	Only a fraction of the SNPs have an effect	Across breeds	BayesR, BayesC, BayesB
GBLUP	Large number of QTL with small effects	All SNPs have small effects	Within breeds	G-BLUP, ssGBLUP

GBLUP – Genomic Best Linear Unbiased Prediction, ssGBLUP – single-step Genomic Best Linear Unbiased Prediction, QTL – Quantitative Trait Loci

The genetic architecture of the traits determines which method is the best to use (Hayes *et al.*, 2010). Non-linear Bayesian methods have a higher accuracy than the GBLUP method when computer simulations are used instead of real data (Meuwissen & Goddard, 2010). In some cases when real data is used to compare the accuracy of the methods, non-linear methods do not always have the highest accuracy compared to the GBLUP method (Erbe *et al.*, 2012). This is due to economically important traits being influenced by several genes and the GBLUP method assuming that all SNPs have an effect on the trait (Meuwissen *et al.*, 2016). Additionally, many SNPs are associated with a gene due to LD being present across large regions in livestock genomes and the effect of a QTL can only be explained by using several SNPs (Meuwissen *et al.*, 2016).

The GBLUP method uses the genomic relationship matrix (G-matrix) to estimate GEBVs (Meuwissen *et al.*, 2016). The G-matrix utilises genotypic data obtained from SNPs across the genome to determine genomic relationships between animals by estimating the proportion of DNA that is shared between two individuals (VanRaden, 2007; Meuwissen *et al.*, 2016). Several different methods that rely on the M-matrix, P-matrix and Z-matrix can be used to obtain the G-matrix.

The M-matrix and the P-matrix is needed to obtain the Z-matrix. The M-matrix specifies the marker alleles that were inherited by each individual (VanRaden, 2008). The rows of the M-matrix represent the number of individuals (n) and the columns represent the number of loci (m) (VanRaden, 2008). Marker information can be included in the equations to obtain matrix **MM'** (n x n) and matrix **M'M** (m x m) (Legarra & Misztal, 2008). If the homozygote, heterozygote and other homozygote is defined as -1, 0, and 1, respectively in the M-matrix, the diagonals of **MM'** count the number of homozygous loci for each individual and the off-diagonals measure the number of alleles shared by relatives (VanRaden, 2007; Legarra & Misztal, 2008). Whereas the diagonals of **M'M** count the number of homozygous individuals for each locus, and the off-diagonals measure the number of times alleles at different loci were inherited by the same individual (VanRaden, 2007; Legarra & Misztal, 2008).

For the P-matrix, column **i** of **P** is **2(pi – 0.5)** with **pi** representing the frequency of the second allele at locus **i** and the P-matrix contains the allele frequencies expressed as a difference from 0.5 and multiplied by 2 (VanRaden, 2008). In order to obtain the Z-matrix, the P-matrix is subtracted from the M-matrix which sets the mean values of the allele effects to 0 (VanRaden, 2008). The allele frequencies that are used in the P-matrix should be from the unselected base population (VanRaden, 2008).

The base population that is used influences the number of relationships between the individuals and the level of inbreeding (VanRaden, 2008). When the P-matrix is subtracted from the M-matrix there is a bias towards rare alleles compared to common alleles when genomic relationships are determined (VanRaden, 2008). The genomic inbreeding coefficient differs between individuals that are homozygous for rare alleles and those that are homozygous for common alleles (VanRaden, 2008). If an individual is homozygous for rare alleles then the genomic inbreeding coefficient is greater than for an individual that is homozygous for common alleles (VanRaden, 2008).

The formula of the first G-matrix method is as follows:

$$\mathbf{G} = \frac{\boldsymbol{ZZ'}}{\mathbf{2}\Sigma \boldsymbol{pi}(\mathbf{1} - \boldsymbol{pi})}$$

with **Z** representing the Z-matrix and **pi** representing the frequency of the second allele at locus **i** (VanRaden, 2008). The division by the elements, $\mathbf{2}\Sigma \boldsymbol{pi}(\mathbf{1} - \boldsymbol{pi})$**,** scales **G** to be analogous to **A** (VanRaden, 2008). The G-matrix is semi-definite but can be singular (VanRaden, 2007). The G-matrix is singular when the total number of alleles is less than the number of genotyped individuals (VanRaden, 2007).

The second method is mainly used in human genetics studies (Leutenegger *et al.*, 2003; Amin *et al.*, 2007). The formula is **G = ZDZ'** and **D** is the Diagonal with $\boldsymbol{Dii} = \frac{\mathbf{1}}{\boldsymbol{m}[\mathbf{2}\boldsymbol{pi}(\mathbf{1}-\boldsymbol{pi})]}$ (Leutenegger *et al.*, 2003; Amin *et al.*, 2007). The formula for the third method is,

$$\boldsymbol{G} = \frac{\boldsymbol{MM'} - \boldsymbol{g}\mathbf{0}\,(\mathbf{11'})}{\boldsymbol{g}\mathbf{1}}$$

and uses the model **MM' = $g_0$11' + g_1A + E,** which adjusts the mean homozygosity by regressing **MM'** on the A-matrix to obtain the G-matrix. The elements **g_0** represents the intercept and **g_1** represents the slope. The E-matrix includes the differences between true and expected fractions of DNA in common as well as measurements error (VanRaden, 2008). **MM'** is the dependent variable and the A-matrix is the independent variable in the regression (VanRaden, 2008).

Although traditional animal breeding models assume that the individuals in a base population are not related, genomic analyses have determined that genes are always shared between the base population individuals due to common ancestors (VanRaden, 2007). Additionally, pedigree relationships assume that siblings with the same parents share 50% of their genomic information, however siblings can share more or less than 50% of their genetic material (Meuwissen *et al.*, 2016). Therefore, GEBVs have a higher accuracy compared to EBVs due to the GBLUP method using genomic relationships to build the relationship matrix as opposed to the traditional BLUP method relying on pedigree relationships (Meuwissen *et al.*, 2016). The estimation of the extent of heterozygosity and the number of alleles shared between individuals is not always accurately calculated in the BLUP method and therefore it is important to include genomic information in the breeding value estimation (VanRaden, 2007; Meuwissen *et al.*, 2016).

GS predictions using the GBLUP method can be performed using a multiple-step or a single-step approach (Meuwissen *et al.*, 2016). The multiple-step approach firstly calculates pseudo-phenotypes for the genotyped animals, which includes information on their relatives that do not have genotypic information, thereafter the pseudo information together with the animals genotypes are used for the genomic prediction and finally breeding values are estimated by combining the traditional EBV and GEBV (VanRaden, 2008). This multiple-step method leads to loss of information, inaccuracy and bias (Legarra *et al.*, 2014).

The single-step genomic best linear unbiased prediction (ssGBLUP), includes all the data in one single estimation step (Meuwissen *et al.*, 2016). The relationship matrix that is used in the calculation is most correct when the genotyped animals' relatedness is taken into account (Meuwissen *et al.*, 2016). The genomic relationships are taken into account first and

then pedigree information is used to determine the remaining relationships of non-genotyped offspring of the genotyped animals (Meuwissen *et al.*, 2016).

The matrix that is used in this method is the inverse of the H-matrix which includes the difference between the inverse of the A-matrix and the inverse of the G-matrix, due to constraints that prevent the genotyping of all animals in a population (Forni *et al.*, 2011). The inverse of the H-matrix is expressed as follows,

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

with $\mathbf{G}^{-1}$ representing the inverse of the G-matrix and $\mathbf{A}_{22}^{-1}$ representing the inverse of the A-matrix for the genotyped animals (Forni *et al.*, 2011). The development of GBLUP and software packages that focus on mixed model equations such as MiX99 and BLUPF90 suite facilitated the estimation of GEBVs (Table 2.3) (Garrick *et al.*, 2018).

Table 2.3. A summary of available software packages for genetic and genomic evaluations.

Software	Website	Main use	Free resource	Reference
DMU	https://omictools.com/dmu-tool	Analysis of multivariate mixed models using Restricted Maximum Likelihood	No	Madsen *et al.*, 2006
BLUPF90 suite	http://nce.ads.uga.edu/software/	Collection of software for mixed model analyses and can be used for ssGBLUP	Yes	Aguilar *et al.*, 2018
TASSEL	https://www.maizegenetics.net/tassel	Implements general linear model and mixed model approaches for controlling population and family structure	Yes	Bradbury *et al.*, 2007
ASREML	http://www.vsn-intl.com/asreml	General mixed models analysis program focusing on the estimation of variance components	No	Gilmour *et al.*, 2002
MiX99	https://www.luke.fi/en/business-solutions/expertise-areas/livestock-and-feed/mix99-solving-large-mixed-model-equations/	Solves large mixed model equations	No	Lidauer *et al.*, 2015

In order to measure the precision of the breeding values, the reliability of the breeding values is estimated using the prediction error variance (PEV) (Henderson, 1975, 1984). The PEV is calculated as part of the genetic evaluation and the reliability of the estimations are derived

from the PEV by using the calculation 1 minus the ratio between PEV and the additive genetic variance (Gorjanc *et al.*, 2015). The reliability of the breeding values also give an indication of the potential response to selection (Gorjanc *et al.*, 2015) and can be expressed as the square of the accuracy (Mrode, 1996; Meuwissen *et al.*, 2001).

The predictive ability of the models are validated by the correlation between the EBV or GEBV and the adjusted phenotypes, deregressed EBVs or yield deviations (YDs) using validation populations (Legarra *et al.*, 2008; Legarra & Reverter, 2018). The YD of an animal is the phenotype of the animal adjusted for all the model effects excluding the additive genetic effects and the errors (Lourenco *et al.*, 2015). The YD of an animal is therefore based on its own phenotypic performance (Lourenco *et al.*, 2015). The equation used to calculate the YD is expressed as follows,

$$\mathbf{YD = y - X\hat{b} - Z\hat{p}},$$

where **y** represents the phenotypic measurements, $\hat{\mathbf{b}}$ represents all the fixed effect factors and $\hat{\mathbf{p}}$ represents the non-genetic animal effect factors (Lidauer *et al.*, 2017).

Validation studies have been performed to determine the influence that genotypes have on the breeding value estimation (Bolormaa *et al.*, 2013; Ni *et al.*, 2017). There are two types of studies that can be performed, a random cross-validation and a forward prediction. The forward prediction validation method divides the population into two groups, a reference or training population and a validation population (Ni *et al.*, 2017). In this method, the youngest individuals are used as the validation set (Ni *et al.*, 2017). In contrast, the random cross-validation method divides the population into smaller groups (Bolormaa *et al.*, 2013; Ni *et al.*, 2017). A few of the groups are used as the training population and one group is used as the validation population (Bolormaa *et al.*, 2013; Ni *et al.*, 2017). This process is repeated until all the groups have been used as a validation population and a training population, respectively (Bolormaa *et al.*, 2013).

2.3. The application of genomic selection methods in dairy and beef cattle

Performance and pedigree data of dairy cattle obtained before the development of genotyping technologies together with the adoption of AI in the 1940's have been essential for the implementation of GS in dairy cattle (Weigel *et al.*, 2017). Selection in dairy cattle mainly focussed on sex-limited traits which can't be measured on a bull, therefore progeny testing, which aimed to estimate a bull's genetic merit based on its offspring, was

implemented (Weigel *et al.*, 2017). However, progeny testing was time consuming, since the genetic merit of the bull was based on the performance of his daughters, which took several years to obtain and AI was costly to implement (Schefers & Weigel, 2012).

The study by Schaeffer (2006) demonstrated that selecting bull calves at 2 years of age using GS could potentially double the rate of genetic gain compared to selecting bulls at 5 years of age or older using progeny testing. If progeny testing was not necessary anymore it would save the bull breeding companies 92% of their expenses and some of the saved costs could be spent on genotyping the bulls (Schaeffer, 2006). A simulation study on dairy cattle also indicated that the GEBV accuracy for a bull calf could be as high as the EBV accuracy after a progeny test (Meuwissen *et al.*, 2001).

Additionally, GS in dairy cattle would allow the breeders to identify genetically superior animals for sex-limited traits at a young age before sexual maturity, decrease the generation interval needed to obtain performance information and increase selection intensity (Schefers & Weigel, 2012). Therefore, when the bovine SNP array became available, thousands of progeny tested Holstein bulls were genotyped to allow the establishment of a reference population for the breed which facilitated the application of GS in dairy cattle (Taylor *et al.*, 2016). Since GS has been implemented in dairy cattle, several studies have demonstrated that GEBVs have a higher reliability compared to parent averages and EBVs obtained from progeny tests (Table 2.4) (Harris *et al.*, 2008; Hayes *et al.*, 2009; VanRaden *et al.*, 2009).

Table 2.4. A summary of studies that focused on the reliabilities of GEBVs compared to EBVs and/or parent averages for genomic selection in dairy cattle.

Country	Aim	Number of animals genotyped	SNP array	Methods	Summary of traits	Main outcome	Reference
Australia	Determining the reliability of the GEBVs compared to the EBVs from progeny tests	798	SNP50TM array	BLUP, BayesA	Fertility, milk protein yield	Fertility had the lowest GEBV reliability (0.18) compared to the other traits such as milk protein yield that had a GEBV reliability of 0.45.	Hayes *et al.*, 2009
New Zealand	Determining the effect of genomic information on the accuracy of breeding values in dairy cattle by comparing the reliabilities of GEBVs and parent averages for unproven sires	4500	Bovine SNP50 BeadChip	BLUP, BayesA, BayesB	Milk production traits, fertility, somatic cell count, longevity, live weight	Reliabilities of the GEBVs varied between 50% to 67% for production traits compared to an average of 34% for parent averages. GEBV reliabilities' for the linear type traits varied between 40% to 50% compared to an average of 31% for parent averages.	Harris *et al.*, 2008
USA	Determining the gain in the reliability of genomic evaluations compared to traditional parent average evaluations	3576 (Training population), 1759 (Test population)	BovineSNP50 BeadChip	BLUP, Bayesian	Milk production traits and linear type traits	Average reliability of the GEBVs averaged across all the traits was 50%, compared to an average reliability of the parent averages of 27%.	VanRaden *et al.*, 2009

BLUP – Best Linear Unbiased Prediction, EBV - Estimated Breeding Value, GEBV – Genomic Enhanced Breeding Value

A subsequent study optimised the method used to estimate GEBVs for the application of genomic evaluations in dairy cattle in New Zealand, which could lead to further success in the application of GS in dairy cattle (Winkelman *et al.,* 2015). The success of GS in dairy cattle over time was confirmed when García-Ruiz *et al.,* (2016) reviewed two decades of genomic selection data in Holstein cattle and determined that annual rates of genetic gain increased from approximately 50% to 100% for highly heritable traits and increased 3 to 4 times for traits with low heritability such as fertility traits (Figure 2.3).

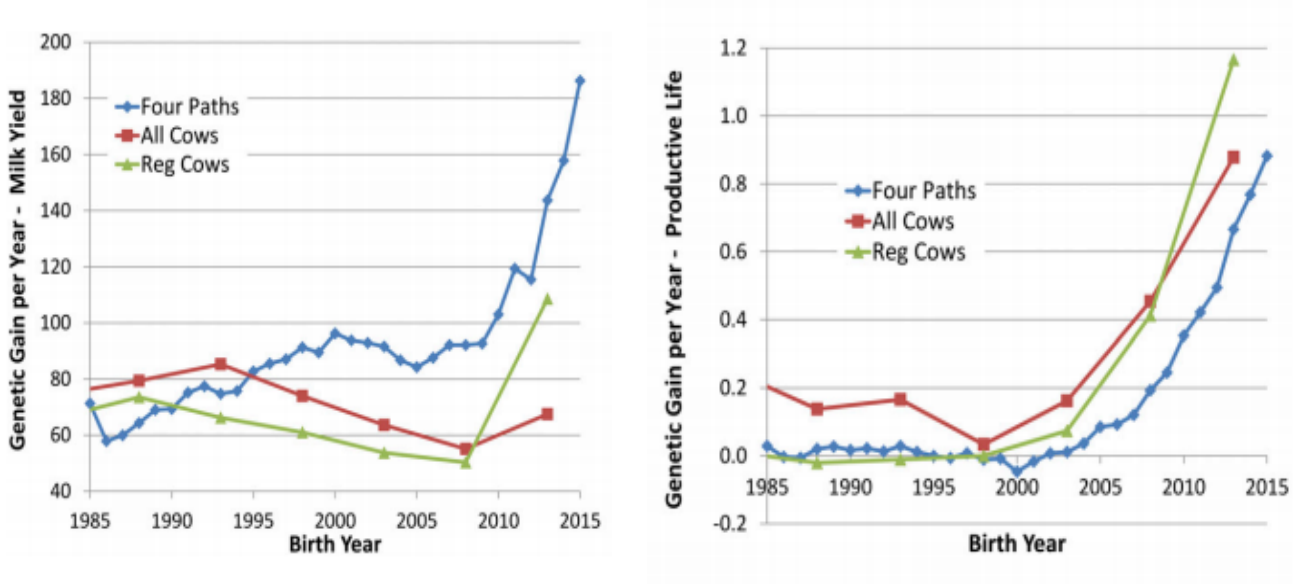

Figure 2.3. The genetic gain per year as a result of genomic selection for milk yield and productive life in Holstein cattle. (Adapted from García-Ruiz *et al.,* 2016).
Four Paths = selection based on the four paths which include sires of bulls, sires of cows, dams of bulls and dams of cows. All cows and Reg cows = segmented regressions of predicted breeding values on birth year for all cows and cows that are registered in the national herd book, respectively.

In South Africa, dairy cattle receive genetic breeding values as part of the Multiple Across Country Evaluation (MACE) for several traits such as fertility and longevity (Van der Westhuizen & Mostert, 2020). A South African state-funded Dairy Genomic Programme (DGP) was established in 2016 with the aim to genotype as many animals as possible to ensure that South African reference populations could be established (Van Marle-Köster & Visser, 2018b). The DGP made it possible to start establishing reference populations for the Jersey and Holstein breeds and these reference populations will aid the implementation of GS using South African reference populations (Van der Westhuizen & Mostert, 2020).

The success of GS in dairy cattle populations across the world led to the development and implementation of GS in beef cattle (Garrick, 2011; Berry *et al.*, 2016). However, the development and implementation of genomic selection in beef cattle was more challenging compared to dairy cattle as a result of several differences between dairy and beef cattle populations (Garrick, 2011; Berry *et al.*, 2016). Beef cattle consist of several pure breeds as well as crossbreeds that are bred to withstand several different environments and

temperatures across the world as opposed to the dairy cattle industry that is mainly dominated by one breed, the Holstein-Friesan (Berry *et al.*, 2016).

The small effective population size of close to 100 of the Holstein breed easily facilitated the establishment of a reference population (McParland *et al.*, 2007; De Roos *et al.*, 2008). In contrast, larger reference populations were required for beef breeds due to their large effective population sizes (Berry *et al.*, 2016). Obtaining larger reference populations for beef breeds were challenging and more expensive due to the number of animals that had to be genotyped (Goddard, 2012; Berry *et al.*, 2016). Additionally, the lack of AI in beef cattle proved challenging due to poor genetic relation between similar breeds in different countries (Berry *et al.*, 2016). Therefore, the accuracy of GS was higher in dairy cattle compared to beef cattle (Goddard, 2012).

Another aspect that played a role was that daughter averages are used as phenotypes in dairy cattle whereas in beef cattle phenotypes of both sexes are used for many traits of importance (Goddard, 2012). The different ways of determining the phenotypes lead to dairy cattle having a higher heritability value compared to beef cattle (Goddard, 2012). In beef cattle it is also a common occurrence to use a mixture of breeds in genomic selection calculations, which decreases the LD (Goddard, 2012; Berry *et al.*, 2016). However, in dairy breeds with a small effective population size only one breed is used in genomic selection calculations and therefore the LD is higher in dairy cattle compared to beef cattle (Goddard, 2012).

The American Angus Association was the first beef breed society to include genomic information in their genetic evaluations in North America, because the Angus breed had contributed significantly to the North American performance recordings and investments in genotyping were made to establish a reference population for the breed (Garrick, 2009; Berry *et al.* 2016). Initially, due to the Hereford breed not having enough genotypes to establish a reference population of its own, the prediction equations used for the Angus breed was tested on the Hereford breed, but the prediction equations had no predictive power in the Hereford breed (Garrick, 2009; Berry *et al.* 2016). Across-breed predictions are not ideal to implement, but can be used to either increase the reference population or account for composite breeds (Kizilkaya *et al.*, 2010; Bolormaa *et al.*, 2013).

Thereafter, the Hereford breed association invested in genotyping animals to allow the establishment of a reference population for the breed (Garrick, 2009; Berry *et al.* 2016). The initial development of reference populations for the Angus and Hereford breeds and the implementation of GS in these breeds led to other breed societies also investing in the

establishment of reference populations (Garrick, 2009; Berry *et al.* 2016). Thereafter, studies analysed the influence of genomic information on the breeding value estimation in beef breeds to determine if the results obtained in dairy cattle could also be obtained in beef cattle (Table 2.5) (Garrick, 2009; Onogi et *al.,* 2014; Neves et *al.,* 2014; Chen *et al.,* 2015; Júnior *et al.*, 2016).

Table 2.5. A summary of several studies performed to determine the accuracy of genomic predictions in beef breeds.

Breed	Aim	Number of animals genotyped	SNP array	Genomic prediction method	Validation method	Traits	Outcome	Reference
Angus	Determining the accuracy of genomic predictions in Angus cattle	2100	50k SNP panel	BayesC	Cross-validation	Scrotal circumference, Weaning weight direct, Yearling weight	Breeding values were correlated with deregressed expected progeny differences of the validation population; 0.5 to 0.7, depending on the trait. Correlations were a guide to accuracy, not applied to breeding values of animals without progeny tests.	Garrick, 2009
Japanese Black	Determining the differences in the accuracies between EBVs and GEBVs	1,576	Illumina BovineSNP50 beadchip	BLUP, ssGBLUP	Cross-validation	Beef marbling score, carcass weight, ribeye area	GEBV accuracies were 5% to 8% higher compared to EBV accuracies for all the traits.	Onogi *et al.*, 2014
Nellore	Implementing genomic predictions in Nellore cattle determining the empirical accuracies of the predictions	685	Illumina Bovine HD Chip	GBLUP, Bayesian	Forward prediction	Traits such as scrotal circumference, weight and carcass traits	The empirical accuracies of the genomic predictions ranged from 0.17 to 0.74.	Neves *et al.*, 2014
Nellore	Evaluate the accuracy of genomic predictions in Nellore cattle	1756	Illumina Bovine HD chip	Bayesian	Cross-validation	Carcass traits such as hot carcass weight and rib eye area	Genomic predictions had a moderate to high accuracy and could be applied to improve carcass traits in Nellore cattle.	Júnior *et al.*, 2016

Breed	Aim	Number of animals genotyped	SNP array	Genomic prediction method	Validation method	Traits	Outcome	Reference
Angus and Charolais	To evaluate the breeding value accuracies in Angus and Charolais cattle using within breed, across breed and combined analyses	543 (Angus steers), 400 (Charolais steers)	Illumina Bovine SNP50 beadchip	BLUP, GBLUP, Bayesian	Cross-validation	Carcass traits	Average accuracy of the GEBVs using within-breed training populations were 0.35 (Angus) and 0.36 (Charolais) whereas using across-breed training populations lead to accuracies of close to 0. The results emphasised the importance of genetic relation between selection candidates and the training population.	Chen *et al.*, 2015
Angus and multi-breed	To determine whether a purebreed beef cattle training population can be used for multi-breed genomic selection performance and vice versa	1086 (Purebred Angus), 924 (Multi-breed, representing breeds such as Angus, Brahman, Hereford and Shorthorn)	50k SNP panel	Bayesian	Cross-validation	Simulated traits	Results indicated that using the purebred Angus cattle as the training population to predict breeding values in the multi-breed population were more successful compared to using the multi-breed training population to predict breeding values for the Angus population.	Kizilkaya *et al.*, 2010
Bos taurus, *B. indicus* and composite beef cattle	Determining the accuracy of GEBV predictions for several traits such as meat quality and carcass traits	10,181	Five different panels with varying number of SNPs from 7K to 800K	GBLUP, BayesR	Cross-validation	Feed efficiency, growth, carcass and meat quality traits	Accuracy of the GEBVs varied between traits. Higher number of animals with genotypic and phenotypic information and traits with high heritability had the highest GEBV accuracies.	Bolormaa *et al.*, 2013

BLUP – Best Linear Unbiased Prediction, EBV – Estimated Breeding Value, GBLUP – Genomic Best Linear Unbiased Prediction, GEBV – Genomic Enhanced Breeding Value

The successful implementation of GS in dairy and beef cattle populations across the world initiated the establishment of the BGP in South Africa which aimed to genotype several beef breeds in order to establish reference populations (SA Stud Book, 2017). The BGP facilitated the establishment of reference populations for the Bonsmara and Beefmaster breeds in South Africa, which allowed the estimation of GEBVs for these breeds (Table 2.6) (SA Stud Book, 2018).

Table 2.6. A summary of the major beef cattle breeds in South Africa participating in the Beef Genomic Programmme (BGP) (SA Stud Book, 2018).

Breed	**Classification**	**GEBVs estimated**
Bonsmara	Composite	✓
Beefmaster	Composite	✓
Drakensberger	Indigenous	x
Hereford	British	x
Charolais	Continental European	x
Nguni	Indigenous	x
Tuli	Indigenous	x

GEBVs – Genomic Enhanced Breeding Values

For a number of beef breeds that participated in the BGP, relatively high costs of genotyping and incomplete recording have been limiting the estimation of GEBVs (Van Marle-Köster *et al.*, 2013; Van Marle-Köster *et al.,* 2018a; Van Marle-Köster *et al.,* 2018b).

2.4 Conclusion

The development of genotyping technologies such as SNP arrays had a significant impact on selection of cattle. The development of SNP technology made it possible to include genomic information in genetic evaluations, which, in turn, increased the accuracy of breeding values and facilitated the implementation of GS in dairy and beef cattle. In order to implement GS in cattle breeds, reference populations needed to be established. Reference populations were first successfully established in dairy cattle breeds and led the way for the implementation of GS in beef cattle breeds across the world. Although South Africa has limited resources they have managed to establish a reference population for the Bonsmara breed, which made it possible to estimate GEBVs for production and fertility traits in the breed.

Chapter 3: Materials and methods

3.1 Introduction

EBVs and GEBVs and the associated accuracy metrics for direct weaning weight, maternal weaning weight, average daily gain and height were estimated. Analysis I focussed on correlating EBVs and GEBVs to determine if including genomic information in the traditional breeding value estimation would influence the ranking of the animals. Additionally, the mean EBV and GEBV accuracies were compared to determine if genomic information increased the accuracy of the estimated breeding values. Analysis II used a forward prediction scheme to determine if genomic information increased the predictive abilities of the breeding value models. Phenotypic and genotypic data were obtained with consent from the South African Bonsmara Breed Society and ethical approval was granted by the Ethical Committee of the University of Pretoria for use of external datasets (NAS156/2019).

3.2 Materials

Genotypic and phenotypic information

The genotypic data for these analyses were generated using the GeneSeek® Genomic Profiler Bovine HD™ (GGP-HD) Chip 150K (Neogen,Lincoln, NE, USA) as part of the BGP over a period of three years. Additional genotypes were also provided by a research project (Bosman *et al.*, 2017) as well through private investment of the South African Bonsmara Breed Society. The additional genotypes were generated using either the 80K GGP-HD Chip (Neogen,Lincoln, NE, USA) or the IDB 53K SNP array (Table 3.1).

Table 3.1. A summary of the available genotypes for the study.

SNP array	Number of genotypes	Funding
GGP-HD 80K	597	RMRDT (*Bosman et al.* 2017) & SA Bonsmara Breed Society
Bovine 150K	1949	BGP
IDB 53K	1643	SA Bonsmara Breed Society

RMRDT – Red Meat Research and Development Trust, BGP – Beef Genomic Programme, SA – South African

The Bonsmara phenotypic datasets contained 2 018 052 phenotypic records as well as pedigree information. Separate datasets were used for analysis I and II. The total number of available phenotypes and genotypes for the traits investigated in analysis I and II are summarised in Table 3.2 and Table 3.3, respectively. The shoulder height of the animals are referred to as height in the study. Height and average daily gain were measured post-weaning when the animals were approximately 12 months old.

Table 3.2. Number of phenotypes and genotypes used in analysis I.

Trait	Genotypes(N)	Phenotypes(N)	*Heritability
Direct weaning weight	4128	1 242 158	0.24
Maternal weaning weight	4128	1 242 158	0.15
Height	4128	221 042	0.33
Average daily gain	4128	225 779	0.26

*Heritability calculated by the South African Stud Book and Animal Improvement Association for the Bonsmara breed

Table 3.3. Number of genotypes and the number of animals with phenotypic and genotypic information for the traits in analysis II.

Trait	Genotypes(N)	Animals with phenotypes and genotypes(N)	*Heritability
Direct weaning weight	4189	3955	0.24
Maternal weaning weight	4189	3955	0.15
Height	4189	1747	0.33
Average daily gain	4189	1748	0.26

*Heritability calculated by the South African Stud Book and Animal Improvement Association for the Bonsmara breed

Models

The single- and multi-trait models used in the present study are part of routine models used by the South African Stud Book and Animal Improvement Association. A summary of the fixed effects, random effects and covariates in the models are shown in Table 3.4.

Table 3.4. Fixed effects, random effects and covariates included in the models.

	Models		
	WW and MWW	**ADG**	**Height**
Fixed effects			
Sex of animal	✓	x	x
Contemporary group of animal	✓	✓	✓
Random effects			
sire of animal	✓	x	x
dam of animal	✓	x	x
genotype of animal	✓	✓	✓
herd of the animal	x	✓	✓
Covariates			
Age of animal when measured	✓	✓	✓
Age of cow	✓	✓	✓
Animal born to heifer or cow	✓	x	x

WW – Direct weaning weight, MWW – Maternal Weaning Weight, ADG – Average Daily Gain

BLUP and ssGBLUP models were used to estimate breeding values and reliabilities. BLUP was based on the linear model,

$$\mathbf{y = X\beta + Zu + e},$$

where **Xβ** defined the fixed effects, **Zu** defined the random effects and **e** represented the random error vector (*n x 1*) with null means (Henderson, 1984; VanRaden, 2008). **X** was an incidence matrix (*n x p*) that was known and fixed and **β** was a fixed vector (*p x 1*). **Z** was a known incidence matrix (*n x q*) and **u** was a random vector (*q x 1*) with null means (Henderson, 1984).

The relationship matrix used in the BLUP model was the inverse of the A–matrix whereas in ssGBLUP the inverse of the H–matrix was used. The inverse of the H-matrix can be written as follows,

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

with $\mathbf{G^{-1}}$ representing the inverse of the G-matrix among genotyped animals and $\mathbf{A_{22}^{-1}}$ representing the inverse of the A-matrix for the genotyped animals (Forni *et al.*, 2011).

3.3 Methodology

Breeding values and reliabilities were estimated using MiX99 software (Lidauer *et al.*, 2015) for analysis I and II. For analysis I, multi-trait models were used for the breeding value estimation while single-trait models were used to calculate the associated reliabilities. For analysis II, single-trait models were used for the breeding value estimation as well as the reliabilities. The MiX99 package Apaxx was used to calculate all the reliabilities of both analyses using Misztal and Wiggans type accuracies (Lidauer *et al.*, 2015). The reliabilities were converted to accuracies using the formula; accuracy = $\sqrt{\text{reliability}}$ (Mrode, 1996).

3.3.1 Comparing breeding value accuracies and animal rankings based on EBVs and GEBVs

BLUP was performed for each trait to estimate the breeding values of the animals as well as the reliabilities. In order to estimate the GEBVs of the animals and the associated reliabilities, ssGBLUP was performed for each trait. Pearson correlations were performed between the EBVs and GEBVs of the genotyped animals. The top twenty animals with the largest GEBVs for each trait were selected and ranked accordingly, thereafter, each animal's GEBV ranking was compared to their EBV ranking to determine if the ranking of the animals changed when genomic information was included in the breeding value estimation. Additionally, the mean EBV and GEBV accuracies were calculated for each trait using the individual accuracies of each breeding value. The EBV and GEBV accuracies were also plotted against each other to compare the EBV and GEBV accuracies for each trait. RStudio (v1.1.456) was used for all correlation steps.

3.3.2 Comparing the predictive abilities of the breeding value models

Selecting the validation populations

In order to select the validation populations, the following steps were followed: A forward prediction scheme was used where the youngest animals with phenotypic and genotypic information were identified as the validation population (Ni *et al.*, 2017). The 500 youngest animals with phenotypic and genotypic information for each trait was identified and extracted from the file containing all the genotyped individuals using Perl (v5.31003).

Breeding value estimation

For each trait, a series of breeding values were estimated using different sources of information in order to determine the impact of genomic information on the breeding values. The Traditional Parental Average (TPA) of the validation population was estimated using only the phenotypic data and pedigree information of the parents. The Genomic-based Parental Average (GPA) was estimated using only the phenotypic data, pedigree and SNP information of the parents. The Parental Average with Genomic information (PAG) was estimated using the phenotypic data, pedigree and SNP information of the parents as well as the SNP information of the validation animals themselves. The EBV was estimated using the pedigree and measurements of the parents, the validation animal and their offspring. The GEBV was estimated using the measurements, pedigree and SNP information of the parents, the validation animals and their offspring. A summary of the breeding values and different sources of information are shown in Table 3.5.

Table 3.5. Summary of the sources of information included in the breeding value estimation using validation populations.

Breeding value	Validation animal		Offspring of validation animal		Parents of validation animal	
	Measurements	SNP information	Measurements	SNP information	Measurements	SNP information
TPA	x	x	x	x	✓	x
GPA	x	x	x	x	✓	✓
PAG	x	✓	x	x	✓	✓
EBV	✓	x	✓	x	✓	x
GEBV	✓	✓	✓	✓	✓	✓

TPA – Traditional Parental Average, PAG - Parental average with Genomics, GPA – Genomic-based Parental Average, EBV – Estimated Breeding Value, GEBV – Genomic Enhanced Breeding Value, SNP – Single Nucleotide Polymorphisms

*Pedigree information included in each breeding value estimation

In order to estimate TPA and GPA all the measurements for the trait of interest of the validation population were removed from the data file and replaced with missing values using Perl (v5.31003). The validation populations' relationships with their offspring were also removed using Perl (v5.31003). BLUP was performed to determine the TPA of the validation population. Additionally, to determine the GPA using ssGBLUP, the SNP information of the validation population was removed. The reliabilities were also performed for TPA and GPA using the edited files.

ssGBLUP was performed to determine the PAG. The same procedure used to determine the GPA was followed, but the validation populations' SNP information was not removed during the estimation procedure. Lastly, in order to estimate EBVs and GEBVs, all the measurements and the relationships with the parents and offspring were restored and BLUP and ssGBLUP was performed. Thereafter, Perl (v5.31003) was used to extract the breeding values of the validation population from the files containing the breeding values of all the animals. Pearson correlations were performed between all of the breeding values of the validation populations for each trait. The correlations were performed between the TPA, GPA, PAG, EBVs and GEBVs.

Yield deviations (YDs) were also calculated for direct weaning weight, maternal weaning weight, average daily gain and height using the formula,

$$\mathbf{YD = y - X\hat{b} - Z\hat{p}},$$

where $\mathbf{y}$ represented the phenotypic measurements, $\hat{\mathbf{b}}$ represented all the fixed effect solutions and $\hat{\mathbf{p}}$ represented the non-genetic animal effect solutions (Lidauer *et al.*, 2017). The yield deviations were calculated by adjusting the phenotypes of the animals for non-genetic random effects and fixed effects (Lidauer *et al.*, 2017) as shown in Table 3.4. The yield deviations were correlated with each breeding value (TPA, GPA, PAG, EBVs and GEBVs) in order to determine the predictive ability of the models using Pearson correlations (Legarra *et al.*, 2008; Legarra & Reverter, 2018).

Chapter 4: Results

4.1 The rankings of the animals according to EBVs and GEBVs for four traits

The ranking of the animals were determined by estimating the EBVs and GEBVs for height, average daily gain, direct and maternal weaning weight and correlating the estimated breeding values using Pearson correlations. Additionally, the mean EBV and GEBV accuracies were also estimated and compared.

The Pearson correlations of the EBVs and the GEBVs for height and average daily gain, respectively, was significantly different from 0 ($r = 0.96 \pm 0.002$, $p<0.001$) as well as for direct weaning weight and maternal weaning weight ($r = 0.95 \pm 0.003$, $p<0.001$). The R^2 for the regression of GEBVs onto EBVs for all the traits were ≥ 0.8977 (Figure 4.1).

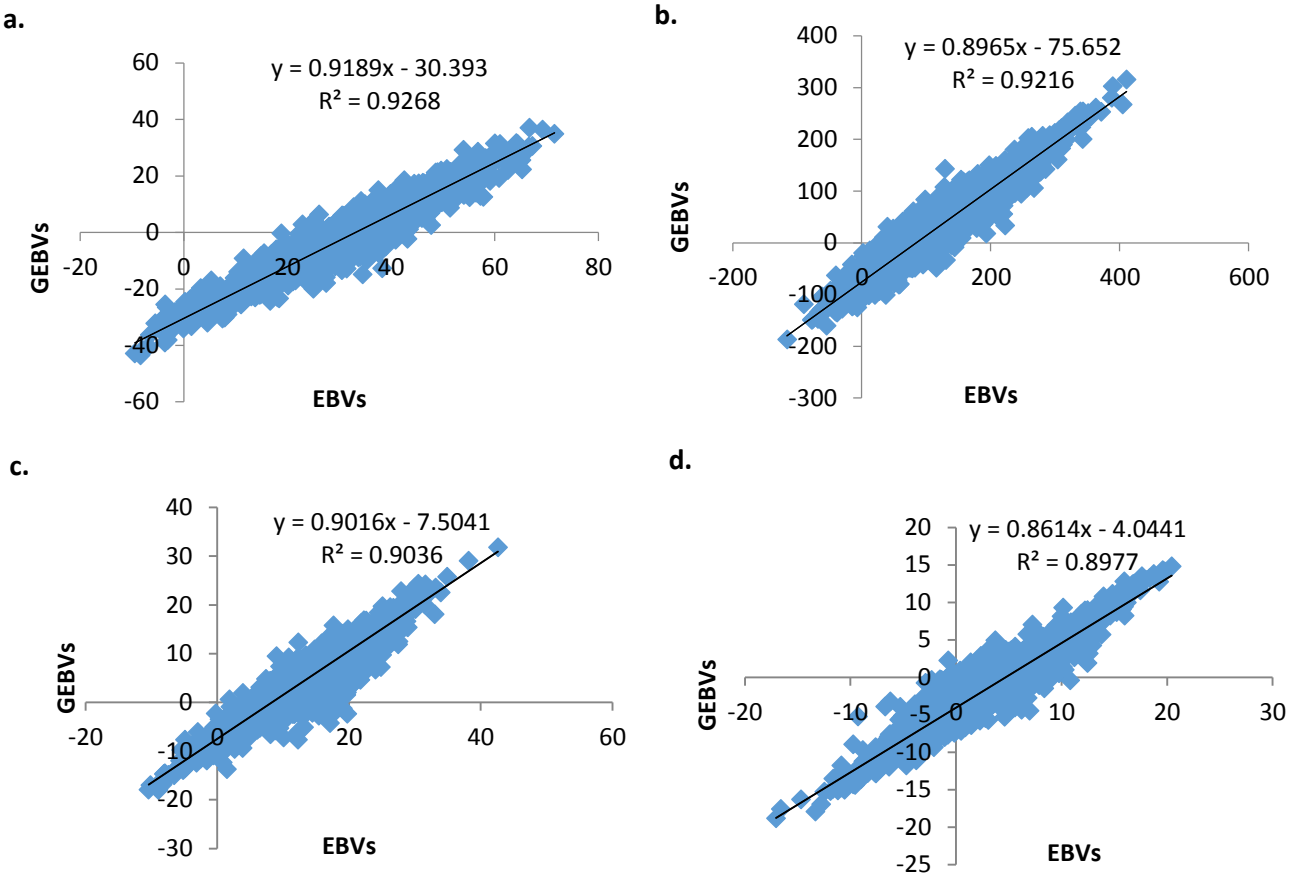

Figure 4.1. Pearson correlations between EBVs and GEBVs for (a) height, (b) average daily gain, (c) direct weaning weight and (d) maternal weaning weight in South African Bonsmara cattle with phenotypic and genotypic information.

Although the Pearson correlations were close to 1, the ranking of the animals did not remain consistent between the GEBVs and the EBVs for the traits (Table 4.1). The top animals tended to remain in the first place for all traits except for height.

Table 4.1. The top ranking animals based on their GEBVs compared to the ranking of the same animals according to their EBVs for the four traits.

WW	Rank	GEBV	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
		EBV	1	2	3	12	9	18	5	7	54	4	8	19	15	21	13	34	22	17	45	28
MWW	Rank	GEBV	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
		EBV	1	3	2	8	11	5	7	6	9	10	20	4	13	15	14	23	12	29	38	26
Height	Rank	GEBV	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
		EBV	4	2	1	12	33	27	11	14	13	15	3	5	18	102	16	6	60	22	21	19
ADG	Rank	GEBV	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
		EBV	1	3	4	2	6	16	12	5	9	8	19	7	11	10	23	14	17	24	20	18

ADG - Average daily gain, EBV – Estimated breeding value, GEBV – Genomic enhanced breeding value, MWW - Maternal weaning weight, WW - Direct weaning weight

For all traits there was an improvement in the accuracies of the GEBVs compared to the EBVs (Table 4.2).

Table 4.2. The mean EBV and GEBV accuracies for height, average daily gain, direct weaning weight and maternal weaning weight for all the animals.

Trait	**Mean accuracy**	
	EBV	**GEBV**
Direct weaning weight	0.629 (0.156)	0.656 (0.144)
Maternal weaning weight	0.519 (0.207)	0.564 (0.185)
Height	0.678 (0.140)	0.721 (0.112)
Average daily gain	0.633 (0.137)	0.686 (0.110)

EBV – Estimated breeding value, GEBV – Genomic enhanced breeding values
*Standard deviation is represented in the brackets

The EBV accuracies plotted against the GEBV accuracies, followed the same trend as seen in Table 2.4 (Figure 4.2).

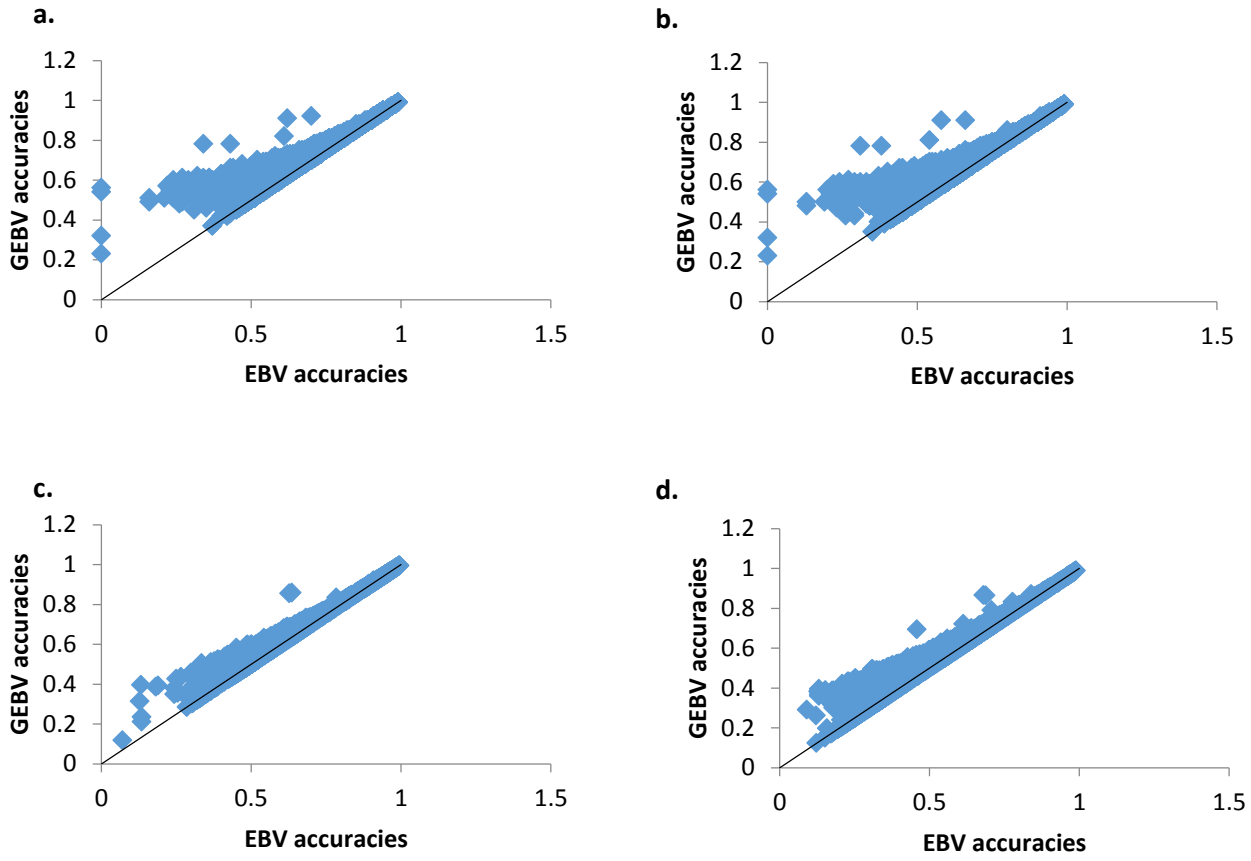

Figure 4.2. EBV accuracies plotted against GEBV accuracies for (a) height, (b) average daily gain, (c) direct weaning weight and (d) maternal weaning weight in South African Bonsmara cattle with phenotypic and genotypic information.

4.2 The predictive abilities of the models for four traits

The predictive abilities of the breeding value models were determined by estimating TPA, GPA, PAG, EBVs, GEBVs and YDs for height, average daily gain, direct and maternal weaning weight and correlating these breeding values using Pearson correlations.

The Pearson correlations between all the breeding values for average daily gain for the validation population was significantly different from 0 ($p<0.001$) (Table 4.3). The relationship between the GEBVs and the PAG was stronger compared to the relationship between the GEBVs and the TPA or GPA (Table 4.3). In contrast, the relationship between the YDs and TPA was stronger than the relationship between the YDs and the GPA or the PAG (Table 4.3). The YDs had the strongest relationship with the GEBVs compared to all the breeding values (Table 4.3). The results are summarised in table 4.3.

Table 4.3. Pearson correlations (r) between breeding values and yield deviations (YDs) for average daily gain.

	r*		
	EBV	**GEBV**	**YD**
TPA	0.713	0.646	**0.196**
GPA	0.687	0.644	0.189
PAG	0.637	**0.710**	**0.188**
EBV		0.953	0.799
GEBV	0.953		**0.807**

TPA – Traditional Parental Average, GPA – Genomic - based Parental Average, PAG – Parental average with Genomics, EBV – Estimated Breeding Value, GEBV – Genomic Enhanced Breeding Value, YD – Yield Deviation

* r = Pearson correlation coefficient

Similar to the results for average daily gain, the Pearson correlations between all the breeding values for height for the validation population was significantly different from 0 ($p<0.001$) (Table 4.4). The same trend was observed in height where the relationship between the GEBVs and PAG was stronger compared to the relationship between the GEBVs and the TPA or GPA (Table 4.4). However, in contrast to the results observed in average daily gain, the relationship between the YDs and the PAG was stronger compared to the relationship between the YDs and the TPA or GPA for height (Table 4.4).

Table 4.4. Pearson correlations (r) between breeding values and yield deviations (YDs) for height.

	r*		
	EBV	**GEBV**	**YD**
TPA	0.710	0.630	**0.305**
GPA	0.702	0.641	0.307
PAG	0.697	**0.734**	**0.368**
EBV		0.966	0.880
GEBV	0.966		**0.893**

TPA – Traditional Parental Average, GPA – Genomic - based Parental Average, PAG – Parental average with Genomics, EBV – Estimated Breeding Value, GEBV – Genomic Enhanced Breeding Value, YD – Yield Deviation

* r = Pearson correlation coefficient

The results for direct weaning weight is summarised in table 4.5. The Pearson correlations between all the breeding values for direct weaning weight for the validation population was significantly different from 0 ($p<0.05$) (Table 4.5). For direct weaning weight there was no significant relationship between the YDs and TPA or GPA ($p>0.05$) (Table 4.5).

Table 4.5. Pearson correlations (r) between breeding values and yield deviations (YDs) for direct weaning weight.

	r*		
	EBV	**GEBV**	**YD**
TPA	0.599	0.471	**-0.004**
GPA	0.532	0.489	0.011
PAG	0.556	**0.660**	**0.106**
EBV		0.924	0.756
GEBV	0.924		**0.777**

TPA – Traditional Parental Average, GPA – Genomic - based Parental Average, PAG – Parental average with Genomics, EBV – Estimated Breeding Value, GEBV – Genomic Enhanced Breeding Value, YD – Yield Deviation

* r = Pearson correlation coefficient

The maternal weaning weight correlations followed the same trend as direct weaning weight and height. However, the relationship between the YDs and the EBVs for maternal weaning weight was stronger compared to the relationship between the YDs and the GEBVs. Furthermore, there was a significant relationship between the YDs and the EBVs ($p<0.05$) and no significant relationship between the GEBVs, PAG, GPA, TPA and YDs of the validation population ($p>0.05$) for maternal weaning weight (Table 4.6). Additionally, the

Pearson correlations between the breeding values for maternal weaning weight for the validation population was significantly different from 0 ($p<0.05$) (Table 4.6). The results are summarised in table 4.6.

Table 4.6. Pearson correlations (r) between breeding values and yield deviations (YDs) for maternal weaning weight.

	r*		
	EBV	**GEBV**	**YD**
TPA	0.965	0.789	**-0.0343**
GPA	0.917	0.829	0.00001
PAG	0.816	**0.979**	**0.019**
EBV		0.825	0.097
GEBV	0.825		**0.05**

TPA – Traditional Parental Average, GPA – Genomic - based Parental Average, PAG – Parental average with Genomics, EBV – Estimated Breeding Value, GEBV – Genomic Enhanced Breeding Value, YD – Yield Deviation

* r = Pearson correlation coefficient

All the Pearson correlations referred to in the results were visualised using graphs (see Addendum A).

Chapter 5: Discussion

5.1 Introduction

Traditionally breeding values for animals were estimated using only pedigree and phenotypic information (Meuwissen *et al.*, 2001; Goddard & Hayes, 2007). Once SNP information became available, it was expected that including SNP information in the breeding value estimation would accelerate genetic gain compared to only using pedigree information (Meuwissen *et al.*, 2001). GS has several advantages such as reducing the cost of progeny testing, increasing the accuracy of breeding values and being able to estimate the breeding values of animals that do not have phenotypic information (Meuwissen *et al.*, 2001; Schaeffer, 2006; Goddard, 2012). GS has been implemented in dairy and beef cattle and several studies across the world have demonstrated the advantages of using genomic information to estimate GEBVs in dairy and beef cattle breeds (Meuwissen *et al.*, 2001; Schaeffer, 2006; Harris *et al.*, 2008; Garrick, 2009; Hayes *et al.*, 2009; VanRaden *et al.*, 2009; Bolormaa *et al.*, 2013; Onogi *et al.*, 2014; Júnior *et al.*, 2016).

In South Africa, the Bonsmara breed is one of the primary breeds used in feedlots (Scholtz *et al.*, 2008). Therefore, due to the economic importance of growth traits in Bonsmara cattle, the breeding value accuracies play an important role in the selection and genetic progress in the breed (SA Stud Book, 2017). Growth traits such as average daily gain, direct weaning weight, maternal weaning weight and height are measured in the Bonsmara breed and EBVs and GEBVs form part of the genetic evaluation of the breed (SA Stud Book, 2017). Bonsmara breeders have made genetic progress for traits such as weaning weight without increasing birth weight and height, through the application of breeding values and selection for functional efficiency (Bonsmara SA, 2019). Maternal weaning weight also plays an important role in direct weaning weight (University of Arkansas System, 2015). The ability of the mother to sufficiently feed her calves increases the direct weaning weight of the animal which in turn allows the farmer to increase their profit when selling weaners to the feedlot (University of Arkansas System, 2015). South African Bonsmara breeders have only been applying GEBVs in their herds since 2017 and they expressed interest in the validation of including genomic information in the breeding value estimation. The aim of this study was to assess the accuracies of EBVs and GEBVs in the selection of South African Bonsmara cattle.

5.2 The rankings of the animals vary between EBVs and GEBVs

Breeding values are used to rank animals based on their estimated genetic potential (Bourdon, 2000). In analysis I, the EBVs and the GEBVs as well as the breeding value accuracies were estimated for all the animals that had genotypic information to determine whether inclusion of genomic information influenced the ranking of animals and the breeding value accuracies. In this study, the Pearson correlation coefficients were close to 1 and these results corresponded to values obtained in a previous study on dairy cattle (Winkelman *et al.*, 2015). However, upon further investigation, the ranking of the animals did not remain consistent between the EBVs and GEBVs for all the traits and re-ranking was mostly observed for animals that were not placed in the top three positions based on their GEBV. For all the traits the top animal remained in the first place according to their GEBV and EBV except for height where re-ranking was observed. The re-ranking that was observed in height could most likely be due to only males being measured for height in the Bonsmara breed due to only a selection of males from the breed participating in growth tests that form part of the national beef cattle performance testing scheme (Bergh, 1999). Therefore, the EBVs are based on limited phenotypic information due to only being able to rely on measurements of the sires, grandsires or male offspring that participated in growth tests.

Additionally, the re-ranking that was observed for most animals that were not in the top three positions for all the traits is most likely due to the inclusion of genomic information in the breeding value estimation providing higher calculated accuracies of the true potential of the animal within the group (Meuwissen *et al.*, 2001; Calus, 2010; Goddard, 2012). These results highlighted the importance of including genomic information in the breeding value estimation to determine the most accurate ranking of an animal within a larger group. An animal with high breeding potential for a specific trait can be overlooked or culled based on its EBV, but could rank as one of the top 10 animals in the herd based on its GEBV.

The accuracies of the GEBVs were higher compared to the EBVs. Improvement in the accuracies varied from 2.7% to 4.5% for direct weaning weight and maternal weaning weight, respectively and 4.3% to 5.3% for height and average daily gain, respectively. The results of this study are similar to previous studies that found that using genomic information to estimate breeding values increased the accuracies of the breeding values compared to only using pedigree relationships (Meuwissen *et al.*, 2001; Calus, 2010; Goddard, 2012; Onogi *et al.*, 2014). This is due to the different methods that are used to obtain EBVs and GEBVs. The ssGBLUP method that is used to estimate the GEBVs and the BLUP method that is used to estimate the EBVs use different relationship matrixes (Calus, 2010; Forni *et*

al., 2011; Meuwissen *et al.*, 2016). The ssGBLUP method uses the inverse of the H-matrix which relies on genomic relationships whereas the BLUP method uses the inverse of the A-matrix which relies on pedigree relationships (Calus, 2010; Forni *et al.*, 2011; Meuwissen *et al.*, 2016). Literature indicated that the ssGBLUP method leads to breeding values with a higher accuracy compared to using the BLUP method, because the genomic relationship matrix gives a more accurate representation of the genomic information that is shared between siblings (Meuwissen *et al.*, 2016).

Furthermore, heritability of a trait influences the accuracies of genomic predictions and should be considered in the breeding value estimations (Calus, 2010). In this study, maternal weaning weight with the lowest heritability also had the lowest mean breeding value accuracies whereas height with the highest heritability had the highest mean EBV and GEBV accuracies compared to the other traits. This corresponds to previous studies that indicated that traits with high heritabilities had higher GEBV accuracies and reliabilities compared to traits with low heritabilities (Hayes *et al.,* 2009; Bolormaa *et al.,* 2013). In this study, when the traits with similar reference population sizes were compared, the improvement in the mean breeding value accuracies from EBVs to GEBVs were larger for the traits with low heritabilities than for the traits with higher heritabilities.

Height and average daily gain had similar reference population sizes with average daily gain having a lower heritability than height. Average daily gain had a 1% higher increase in the mean breeding value accuracies from EBVs to GEBVs compared to height. Additionally, the same pattern was observed when direct weaning weight and maternal weaning weight with the same reference population size was compared. Maternal weaning weight with the lowest heritability compared to direct weaning weight, had a 1.8% increase in the mean breeding value accuracies from EBVs to GEBVs compared to direct weaning weight. These results correspond to a previous study that also obtained results where the lower heritability trait had a larger increase in breeding value accuracies from EBVs to GEBVs compared to the trait with a higher heritability (Onogi *et al.* 2014). These results indicate that low heritability traits have a slightly larger increase in breeding value accuracies when genomic information is included in the breeding value estimation, which could be an indication that genomic information has a larger influence on the breeding value accuracies of low heritability traits compared to high heritability traits, emphasising the impact of genomic information on the breeding values of lowly heritable traits.

5.3 The predictive abilities of the models increased when genomic information was included

Validation populations consisting of the 500 youngest animals with phenotypic and genotypic information were used in analysis II. Analysis II aimed to determine the predictive abilities of the models of the four traits by estimating breeding values using various data sources (see Table 3.5) and comparing the various breeding values and the yield deviations. The predictive abilities of the TPA and GPA models for all the traits were the lowest compared to all the other breeding value models. These results correspond to a study performed on Northern American Holstein bulls that indicated that the average reliability of the parent averages were 23% lower compared to the reliability of GEBVs (Hayes *et al.,* 2009; VanRaden *et al.,* 2009). The TPA models only relied on phenotypic and pedigree information of the parents and the GPA models relied on phenotypic, genotypic and pedigree information of the parents to estimate breeding values and therefore the breeding values were based on parental information as opposed to the animals' own information.

When the model was adjusted to include genomic information of the validation animal using PAG, the predictive abilities of the models for height, direct weaning weight and maternal weaning weight increased. These results corresponded to literature that indicated that genomic information increases the reliability of breeding values (Harris *et al.*, 2008; Hayes *et al.*, 2009; VanRaden *et al.*, 2009). However, a decrease in the predictive ability of the model for average daily gain was observed when genomic information of the animal was included in the model. In this study, the reference population for average daily gain consisted of only 1748 animals and could be the limiting factor. The heritability for average daily gain was 0.26. Literature indicated that a reference population of more than 4000 animals is required to increase the accuracy of genomic predictions for a trait with a heritability close to 0.2 (Goddard, 2012; Oldenbroek & Van der Waaij, 2015). The size of the reference population for average daily gain is therefore a limiting factor for estimating the breeding values of genotyped animals without phenotypic information. Average daily gain can only be recorded in intensive feeding tests which is part of central testing in the national beef cattle performance testing scheme (Bergh, 1999; Mokoena *et al.*, 1999). Therefore, more animals should be included in intensive feeding tests where average daily gain can be recorded and animals can be genotyped.

The EBV models for all the traits had higher predictive abilities than the TPA, GPA and PAG models. This was observed due to the animals' phenotypes being added to the model. The predictive abilities of the GEBV models for height, average daily gain and direct weaning weight had the highest predictive abilities compared to all the other breeding value models.

This indicated that GEBVs are the most accurate in predicting the true genetic potential of an animal compared to all the other breeding values. The increase in the predictive ability when genomic information was added to the breeding value estimation is most likely due to the relationship matrix used to estimate the GEBVs. The ssGBLUP method used to estimate the GEBVs first take the genomic relationships between the animals into account followed by the pedigree information to determine the remaining relationships of non-genotyped offspring of the genotyped animals (Meuwissen *et al.*, 2016). The H-inverse matrix provides a more accurate representation of the relationships between animals compared to the inverse of the A-matrix (Meuwissen *et al.*, 2016). Therefore when genomic information is used to estimate breeding values, the breeding values are a more accurate representation of the true genetic potential of the animal (Meuwissen *et al.*, 2016).

The predictive ability of the GEBV model for maternal weaning weight was lower compared to the EBV model's predictive ability. This is due to the validation population consisting of the 500 youngest animals with phenotypic and genotypic information. The maternal weaning weight of the youngest animals cannot be accurately estimated due to the validation population having a low number of offspring. Therefore, a different method should be used to study the influence that genomic information has on the predictive ability of the model for maternal weaning weight. The heritability of the traits also influenced the predictive abilities of the models. Traits with higher heritabilities such as height (h^2 = 0.33) and average daily gain (h^2 = 0.26) had higher predictive abilities for the EBV and GEBV models, compared to traits with lower heritabilities such as maternal (h^2 = 0.15) and direct weaning weight (h^2 = 0.24). These results indicate that the heritability of a trait influences the predictive ability of the breeding value model which corresponds to literature indicating that heritability of a trait influences the accuracy of genomic predictions (Hayes *et al.,* 2009; Calus, 2010; Bolormaa *et al.,* 2013).

A previous study predicted that implementing genomic selection at a young age could double the rate of genetic gain compared to implementing genomic selection at a later stage in life (Schaeffer, 2006). Additionally, it has also been shown that genomic selection is the most advantageous when the phenotype of interest cannot be observed at a young age when the animals become sexually mature (Goddard, 2012). Genomic information influenced the predictive abilities of the PAG models for height, direct weaning weight and maternal weaning weight. Therefore, genotyping young animals that do not have measurements for these traits and using the genomic information to estimate the breeding values for the Bonsmara breed will be beneficial. In the case where there is limited information available for young animals, the genomic information will provide a good

indication of the potential of the animal at a young age compared to only relying on the information of the parents.

Additionally, the predictive abilities of the GEBV models for height, direct weaning weight and average daily gain indicated that estimating GEBVs for animals were the best measure of the true potential of the animal compared to the other breeding values. Therefore, in the Bonsmara breed it would be ideal to estimate GEBVs for animals of all ages using the information available. Although, the same was not true for maternal weaning weight, the age of the validation population may have played a role in these results and therefore a cross-validation method may be necessary to show the importance of using genomic information to estimate breeding values for maternal weaning weight.

5.4 Conclusion

This study clearly confirmed the value of genomic information in the estimation of breeding values. Ranking of animals were influenced by including genomic information in the breeding value estimation and should be taken into account for accurate selection. An improvement in the accuracies of the breeding values was demonstrated. Additionally, genomic information also increased the predictive abilities of the breeding value models with the GEBV having the highest benefit for improvement of selection accuracy in the South African Bonsmara.

Chapter 6: Conclusion and recommendations

6.1 Conclusion

In South Africa, it is mandatory to register Bonsmara animals and to record the performance of each of these animals (Bosman *et al.*, 2017). Therefore, large datasets containing phenotypic and pedigree data for the Bonsmara breed were available and facilitated the estimation of EBVs for the breed (SA Stud Book, 2017; Bosman *et al.*, 2017; SA Stud Book, 2019). However, studies indicated the value of genomic information for increased breeding value accuracies (Meuwissen *et al.*, 2001; Schaeffer, 2006). Therefore, the BGP was founded in 2015 for establishing reference populations for 16 beef breeds in South Africa that would allow the estimation of GEBVs (SA Stud Book, 2017; Van Marle-Köster & Visser, 2018a). The large datasets with phenotypic and pedigree data available for the Bonsmara breed and the genotypes from the BGP facilitated the establishment of the reference population for the Bonsmara breed (SA Stud Book, 2017). Once a reference population was established, GEBVs could be estimated for the Bonsmara breed (SA Stud Book, 2017). GEBVs was implemented and applied by the Bonsmara breeders since 2017 (SA Stud Book, 2017).

This study was performed in order to assess the accuracies of EBVs and GEBVs for selection in Bonsmara cattle for traits with medium and high heritability. Analysis I focused on all the Bonsmara animals that had genotypic information available. EBVs and GEBVs together with breeding value accuracies were estimated for the growth traits with the aim to determine if genomic information influenced the ranking of animals and the breeding value accuracies. Additionally, analysis II was performed to determine if genomic information influenced the predictive ability of the models for the growth traits. In analysis II, validation populations made up of the 500 youngest animals with phenotypic and genotypic information for the growth traits were established. Breeding values using various data sources and yield deviations were estimated. The various breeding values and the yield deviations were correlated in order to determine how the absence and presence of genomic information influenced the predictive ability of the breeding value models for the growth traits.

This study demonstrated that genomic information does influence the rankings of the animals, the accuracies of the breeding values and the predictive abilities of the breeding value models for the Bonsmara breed. Therefore, it is necessary to include genomic information in the breeding value estimation for Bonsmara cattle. The potential for genetic

gain is higher when genomic information is included in the breeding value estimation for growth traits of Bonsmara cattle. Breeders could also potentially make selection decisions using GEBVs at a younger age in their herds and could therefore increase the genetic gain compared to using traditional breeding values and EBVs. GS has the potential to accelerate genetic progress with more accurate selection at a younger age.

Additionally, it would also be advantageous to genotype older animals and stud animals, to ensure that selection decisions are made using the most accurate breeding values. The influence of genomic information on the breeding value models for maternal weaning weight is not that clear and additional research will have to be performed to determine if genomic information improves the breeding value accuracy of maternal weaning weight. However, based on the results for the other three traits, the indication is that it is worthwhile to genotype animals of all ages and to include the genotypic information in the breeding value estimation for growth traits for Bonsmara cattle.

6.2 Recommendations

In future studies, more animals that have phenotypic information for average daily gain should be genotyped in order to determine how many animals with genotypic and phenotypic information is needed to ensure that the reference population size is conducive to estimate PAGs for young animals and GEBVs with higher accuracies. Based on the results for average daily gain with a medium heritability, it is expected that the current reference population for traits with low heritability such as fertility traits will also not be sufficient to estimate PAGs for young animals and GEBVs with high accuracies. Therefore, further investigations are necessary to determine the minimum reference population size that is conducive to estimate breeding values with genomic information for young Bonsmara animals without measurements for traits with low and medium heritabilities.

Additionally, it would be informative to determine how the rate of genetic improvement varies between herds that are selected based on their EBVs compared to herds that are selected based on their GEBVs. This would give a clear indication whether the rate of genetic improvement increases when GEBVs are used for selection compared to EBVs. This information could also help the breeder to know which breeding values are worthwhile to use in order to reach the breeding goal at a faster pace within their herds.

References

Affymetrix., 2017. Axiom™ Genome-Wide BOS 1 Array Plate. https://www.affymetrix.com/products_services/arrays/specific/axiom_gwas_bovine.affx. Accessed 18 February 2020.

Aguilar, I., Tsuruta, S., Masuda, Y., Lourenco, D. A. L., Legarra, A. & Misztal, I., 2018. BLUPF90 suite of programs for animal breeding with focus on genomics. Pages 11–16 in Proceedings of the world congress on genetics applied to livestock production.

Aitken, N., Smith, S., Schwarz, C. & Morin, P. A., 2004. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. Mol. Ecol. 13, 1423–1431.

Amer, P. R., Simm, G., Keane, M. G., Diskin, M. G. & Wickham, B. W., 2001. Breeding objectives for beef cattle in Ireland. Livest. Prod. Sci. 67, 223–239.

Amin, N., Van Duijn, C. M. & Aulchenko, Y. S., 2007. A genomic background based method for association analysis in related individuals. PLoS One 2, e1274.

Bell, A. E., 1977. Heritability in retrospect. J. Hered. 68, 297–300.

Bergh, L., 1999. The national beef cattle performance testing scheme. Pages 81–90 in Beef breeding in South Africa.

Berry, D. P., Kearney, F. & Harris, B. L., 2009. Genomic selection in Ireland. Interbull Bull., 29–34.

Berry, D. P., Garcia, J. F. & Garrick, D. J., 2016. Development and implementation of genomic predictions in beef cattle. Anim. Front. 6, 32–38.

Blasco, A. & Toro, M. A., 2014. A short critical history of the application of genomics to animal breeding. Livest. Sci. 166, 4–9.

Boichard, D., Ducrocq, V., Croiseau, P. & Fritz, S., 2016. Genomic selection in domestic animals: principles, applications and perspectives. C. R. Biol. 339, 274–277.

Bolormaa, S., Pryce, J. E., Kemper, K., Savin, K., Hayes, B. J., Barendse, W., Zhang, Y., Reich, C. M., Mason, B. A., Bunch, R. J., Harrison, B. E., Reverter, A., Herd, R. M., Tier, B., Graser, H. U. & Goddard, M. E., 2013. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. J. Anim. Sci. 91, 3088–3104.

Bonsma, J. C., 1980. Cross-Breeding, Breed Creation and the Genesis of the Bonsmara. Pages 90–110 in Livestock Production - A Global Approach.

Bonsmara SA., 2019. Your comprehensive guide. http://bonsmara.co.za/wp-content/uploads/2019/09/Bonsmara-Comprehensive-Guide.pdf. Accessed 18 February 2020.

Bosman, L., Van Marle-Köster, E., Van der Westhuizen, R. R., Visser, C. & Berry, D. P.,

2017. Short communication: Population structure of the South African Bonsmara beef breed using high density single nucleotide polymorphism genotypes. Livest. Sci. 197, 102–105.

Botstein, D., White, R. L., Skolnick, M. & Davis, R. W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32, 314–331.

Bourdon, R. M., 2000. Understanding Animal Breeding. 2nd ed. Prentice Hall.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. & Buckler, E. S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633–2635.

Calus, M. P. L., 2010. Genomic breeding value prediction: methods and procedures. Animal 4, 157–164.

Chen, L., Vinsky, M. & Li, C., 2015. Accuracy of predicting genomic breeding values for carcass merit traits in Angus and Charolais beef cattle. Anim. Genet. 46, 55–59.

Crump, S. L., 1946. The estimation of variance components in analysis of variance. Biometrics Bull. 2, 7–11.

Crump, S. L., 1951. The present status of variance component analysis. Biometrics 7, 1–16.

De Roos, A. P. W., Hayes, B. J., Spelman, R. J. & Goddard, M. E., 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179, 1503–1512.

Department of Agriculture Forestry and Fisheries., 2017. A profile of the South African beef market value chain. https://www.nda.agric.za/doaDev/sideMenu/Marketing/Annual%20Publications/Commodity%20Profiles/Beef%20Market%20Value%20Chain%20Profile%202017.pdf. Accessed 15 March 2020.

Eck, S. H., Benet-Pagès, A., Flisikowski, K., Meitinger, T., Fries, R. & Strom, T. M., 2009. Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. Genome Biol. 10, 82.

Eisenhart, C., 1947. The assumptions underlying the analysis of variance. Biometrics 3, 1–21.

Elsik, C. G., Tellam, R. L. & Worley, K. C., 2009. The genome sequence of taurine cattle: A window to ruminant biology and evolution. Science. 324, 522–528.

Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., Mason, B. A. & Goddard, M. E., 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95, 4114–4129.

Everett, R. W., Quaas, R. L. & McClintock, A. E., 1979. Daughters' maternal grandsires in

sire evaluation. J. Dairy Sci. 62, 1304–1313.

Forni, S., Aguilar, I. & Misztal, I., 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet. Sel. Evol. 43, 1–7.

García-ruiz, A., Cole, J. B., VanRaden, P. M., Wiggans, G. R., Ruiz-lópez, F. J. & Van Tassel, C. P., 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc. Natl. Acad. Sci. 113, E3995-E4004.

Garrick, D. J., 2009. The nature and scope of some whole-genome analysis in US beef cattle. Pages 92–102 in Proceedings of the Beef Improvement Federation 41st Annual Research Symposium .

Garrick, D. J., 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. Genet. Sel. Evol. 43, 17.

Garrick, D. J. & Golden, B. L., 2009. Producing and using genetic evaluations in the United States beef industry of today. J. Anim. Sci. 87, E11–E18.

Garrick, D. J., Garrick, D. P. & Golden, B. L., 2018. An introduction to BOLT software for genetic and genomic evaluations. Page 973 in Proceedings of the 11th World Congress on Genetics Applied to Livestock Production.

Gilmour, A. R., Thompson, R., Cullis, B. R. & Welham, S. J., 2002. ASREML estimates variance matrices from multivariate data using the animal model. Pages 28–05 in Proceedings of the 7th World Congress on Genetics Applied to Livestock Production.

Goddard, M. E. & Hayes, B. J., 2007. Genomic selection. J. Anim. Breed. Genet. 124, 323–330.

Goddard, M. E., 2012. Uses of genomics in livestock agriculture. Anim. Prod. Sci. 52, 73–77.

González-Recio, O., Gianola, D., Long, N., Weigel, K. A., Rosa, G. J. & Avendaño, S., 2008. Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in broilers. Genetics 178, 2305–2313.

Gorjanc, G., Bijma, P. & Hickey, J. M., 2015. Reliability of pedigree-based and genomic evaluations in selected populations. Genet. Sel. Evol. 47, 65.

Graser, H. U., Smith, S. P. & Tier, B., 1987. A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. J. Anim. Sci. 64, 1362–1370.

Gray, I. C., Campbell, D. A. & Spurr, N. K., 2000. Single nucleotide polymorphisms as tools in human genetics. Hum. Mol. Genet. 9, 2403–2408.

Harris, B. L., Johnson, D. L. & Spelman, R. J., 2008. Genomic selection in New Zealand and the implications for national genetic evaluation. Pages 325–330 in Proceedings of the Interbull Meeting, Niagara Falls, Canada.

Harvey, W. R., 1960. Least-squares analysis of data with unequal subclass numbers. Agricultural Research Service, United States Department of Agriculture.

Harvey, W. R., 1977. Users guide for LSML76. Mixed model least squares and maximum likelihood computer program.Ohio State University.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92, 433–443.

Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J. & Goddard, M. E., 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genet. 6, e1001139.

Hazel, L. N. & Lush, J. L., 1942. The efficiency of three methods of selection. J. Hered. 33, 393–399.

Hazel, L. N., 1943. The genetic basis for constructing selection indexes. Genetics 28, 476–490.

Henderson, C. R., 1948. Estimation of general, specific and maternal combining abilities in crosses among inbred lines of swine. Unpublished PhD thesis. Iowa State University Library, Ames, Iowa.

Henderson, C. R., 1953. Estimation of variance and covariance components. Biometrics 9, 226–252.

Henderson, C. R., 1975. Best Linear Unbiased Estimation and Prediction under a selection model. Biometrics 31, 423–447.

Henderson, C. R., 1984. Applications of Linear Models in Animal Breeding Models. University of Guelph.

Hill, W. G., 2014. Applications of population genetics to animal breeding, from Wright, Fisher and Lush to genomic prediction. Genetics 196, 1–16.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., Bork, P., Burt, D. W., Groenen, M. A. M., Delany, M. E., Dodgson, J. B., Chinwalla, A. T., Cliften, P. F., Clifton, S. W., Delehaunty, K. D., Fronick, C., Fulton, R. S., Graves, T. A., Kremitzki, C., Layman, D., Magrini, V., McPherson, J. D., Miner, T. L., Minx, P., Nash, W. E., Nhan, M. N., Nelson, J. O., Oddy, L. G., Pohl, C. S., Randall-Maher, J., Smith, S. M., Wallis, J. W., Yang, S. P., Romanov, M. N., Rondelli, C. M., Paton, B., Smith, J., Morrice, D., Daniels, L., Tempest, H. G., Robertson, L., Masabanda, J. S., Griffin, D. K., Vignal, A., Fillon, V., Jacobbson, L., Kerje, S., Andersson, L., Crooijmans, R. P. M., Aerts, J., van der Poel, J. J., Ellegren, H., Caldwell, R. B., Hubbard, S. J., Grafham, D. V., Kierzek, A. M., McLaren, S. R., Overton, I. M., Arakawa, H., Beattie, K. J., Bezzubov, Y., Boardman, P. E., Bonfield, J. K., Croning, M. D. R., Davies, R. M., Francis, M. D., Humphray, S. J., Scott, C. E., Taylor, R. G., Tickle, C., Brown, W. R. A.,

Rogers, J., Buerstedde, J. M., Wilson, S. A., Stubbs, L., Ovcharenko, I., Gordon, L., Lucas, S., Miller, M. M., Inoko, H., Shiina, T., Kaufman, J., Salomonsen, J., Skjoedt, K., Wong, G. K. S., Wang, J., Liu, B., Wang, J., Yu, J., Yang, H., Nefedov, M., Koriabine, M., Dejong, P. J., Goodstadt, L., Webber, C., Dickens, N. J., Letunic, I., Suyama, M., Torrents, D., von Mering, C., Zdobnov, E. M., Makova, K., Nekrutenko, A., Elnitski, L., Eswara, P., King, D. C., Yang, S., Tyekucheva, S., Radakrishnan, A., Harris, R. S., Chiaromonte, F., Taylor, J., He, J., Rijnkels, M., Griffiths-Jones, S., Ureta-Vidal, A., Hoffman, M. M., Severin, J., Searle, S. M. J., Law, A. S., Speed, D., Waddington, D., Cheng, Z., Tuzun, E., Eichler, E., Bao, Z., Flicek, P., Shteynberg, D. D., Brent, M. R., Bye, J. M., Huckle, E. J., Chatterji, S., Dewey, C., Pachter, L., Kouranov, A., Mourelatos, Z., Hatzigeorgiou, A. G., Paterson, A. H., Ivarie, R., Brandstrom, M., Axelsson, E., Backstrom, N., Berlin, S., Webster, M. T., Pourquie, O., Reymond, A., Ucla, C., Antonarakis, S. E., Long, M., Emerson, J. J., Betrán, E., Dupanloup, I., Kaessmann, H., Hinrichs, A. S., Bejerano, G., Furey, T. S., Harte, R. A., Raney, B., Siepel, A., James Kent, W., Haussler, D., Eyras, E., Castelo, R., Abril, J. F., Castellano, S., Camara, F., Parra, G., Guigo, R., Bourque, G., Tesler, G., Pevzner, P. A., Smit, A., Fulton, L. A., Mardis, E. R. & Wilson, R. K., 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716.

Hofer, A., 1998. Variance component estimation in animal breeding: a review. J. Anim. Breed. Genet. 115, 247–265.

Hudson, G. F. S. & Schaeffer, L. R., 1984. Monte Carlo comparison of sire evaluation models in populations subject to selection and nonrandom mating. J. Dairy Sci. 67, 1264–1272.

Illumina., 2020. All products. https://www.illumina.com/products/all-products.html. Accessed 16 February 2020.

Júnior, G. A. F., Rosa, G. J., Valente, B. D., Carvalheiro, R., Baldi, F., Garcia, D. A., Gordo, D. G. M., Espigolan, R., Takada, L., Tonussi, R. L., De Andrade, W. B. F., Magalhães, A. F. B., Chardulo, L. A. L., Tonhati, H. & De Albuquerque, L. G., 2016. Genomic prediction of breeding values for carcass traits in Nellore cattle. Genet. Sel. Evol. 48, 1–8.

Kizilkaya, K., Fernando, R. L. & Garrick, D. J., 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J. Anim. Sci. 88, 544–551.

Koivula, M., Strandén, I., Su, G. & Mäntysaari, E. A., 2012. Different methods to calculate genomic predictions—Comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach

(H-BLUP). J. Dairy Sci. 95, 4065–4073.

Koopaee, H. K. & Koshkoiyeh, A. E., 2014. SNPs genotyping technologies and their applications in farm animals breeding programs: review. Brazilian Arch. Biol. Technol. 57, 87–95.

Legarra, A. & Misztal, I., 2008. Technical Note: Computing strategies in genome-wide selection. J. Dairy Sci. 91, 360–366.

Legarra, A., Robert-Granié, C., Manfredi, E. & Elsen, J. M., 2008. Performance of genomic selection in mice. Genetics 180, 611–618.

Legarra, A., Christensen, O. F., Aguilar, I. & Misztal, I., 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166, 54–65.

Legarra, A. & Reverter, A., 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet. Sel. Evol. 50, 53.

Leutenegger, A. L., Prum, B., Génin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F. & Thompson, E. A., 2003. Estimation of the inbreeding coefficient through use of genomic data. Am. Soc. Hum. Genet. 73, 516–523.

Lidauer, M., Matilainen, K., Mäntysaari, E., Pitkänen, T., Taskinen, M. & Strandén, I., 2015. MiX99 - Solving large mixed model equations.

Lidauer, M., Matilainen, K., Mäntysaari, E., Pitkänen, T., Taskinen, M. & Strandén, I., 2017. Technical reference guide for Mix99 solver.

Lourenco, D. A. L., Fragomeni, B. O., Tsuruta, S., Aguilar, I., Zumbach, B., Hawken, R. J., Legarra, A. & Misztal, I., 2015. Accuracy of estimated breeding values with genomic information on males, females, or both: an example on broiler chicken. Genet. Sel. Evol. 47, 56.

Lush, J. L., 1935. The inheritance of productivity in farm livestock. Part V. Discussion of preceding contributions. Emp. J. Exp. Agric. 3, 25–30.

Lush, J. L., 1936. Genetic aspects of the Danish system of progeny-testing swine. Res. Bull. 18, 1.

Lush, J. L., 1937. Animal breeding plans. Collegiate Press Inc.

Madsen, P., Su, G., Labouriau, R. & Christensen, O. F., 2006. DMU-a package for analyzing multivariate mixed models. Page 247 in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production.

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P. L., Sonstegard, T. S. & Van Tassel, C. P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4, e5350.

McParland, S., Kearney, J. F., Rath, M. & Berry, D. P., 2007. Inbreeding trends and pedigree

analysis of Irish dairy and beef cattle populations. J. Anim. Sci. 85, 322–331.

Meissner, H. H., Scholtz, M. M. & Palmer, A. R., 2013. Sustainability of the South African Livestock Sector towards 2050 Part 1: Worth and impact of the sector. S. Afr. J. Anim. Sci. 43, 282–297.

Meuwissen, T. H. E. & Goddard, M. E., 1996. The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. 28, 161–176.

Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Meuwissen, T. H. E. & Goddard, M. E., 2010. Accurate prediction of genetic values for complex traits by Whole-Genome resequencing. Genetics 185, 623–631.

Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E., 2016. Genomic selection: A paradigm shift in animal breeding. Anim. Front. 6, 6–14.

Meyer, K., 1986. Restricted maximum likelihood to estimate genetic parameters - in practice. Pages 454–459 in Proceedings 3rd World Congress on Genetics Applied to Livestock Production. University of Nebraska, Lincoln NE.

Meyer, K., 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. Genet. Sel. Evol. 21, 317–340.

Meyer, K., 1991. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. Genet. Sel. Evol. 23, 67–83.

Mokoena, M. R., Townsend, R. F. & Kirsten, J. F., 1999. Cattle improvement schemes in South Africa: Measuring the returns to research investments. Agrekon 38, 1–12.

Mouse Genome Sequencing Consortium., 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562.

Mrode, R. A., 1996. Linear models for the prediction of animal breeding values.

Mullis, K., Faloona, F., Scharf, S., Saiki, R. K., Horn, G. T. & Erlich, H., 1986. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. Pages 263–273 in Cold Spring Harbor Symposia on Quantitative Biology. Cold Spring Harbor Laboratory Press.

Neves, H. H., Carvalheiro, R., O'Brien, A. M. P., Utsunomiya, Y. T., Do Carmo, A. S., Schenkel, F. S., Sölkner, J., McEwan, J. C., Van Tassell, C. P., Cole, J. B., da Silva, M. V., Queiroz, S. A., Sonstegard, T. S. & Garcia, J. F., 2014. Accuracy of genomic predictions in *Bos indicus* (Nellore) cattle. Genet. Sel. Evol. 46, 17.

Nevondo, T. T., Chaminuka, P., Nhundu, K. & Liebenberg, F., 2019. Economic returns from investment in beef cattle improvement research in South Africa. Agrekon, 1–12.

Ni, G., Kipp, S., Simianer, H. & Erbe, M., 2017. Accuracy of genomic breeding values revisited: Assessment of two established approaches and a novel one to determine the

accuracy in two-step genomic prediction. J. Anim. Breed. Genet. 134, 242–255.

Oldenbroek, K. & Van der Waaij, L., 2015. The size of the reference population. Page 187 in Textbook animal breeding and genetics for BSc students.

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

Patterson, H. D. & Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545–554.

Queenan, K., Sobratee, N., Davids, R., Mabhaudhi, T., Chimonyo, M., Slotow, R., Shankar, B. & Häsler, B., 2020. A systems analysis and conceptual system dynamics model of the livestock-derived food system in South Africa: A tool for policy guidance. J. Agric. Food Syst. Community Dev. 9, 1–30.

Red Meat Producers' Organisation., 2017. Manual for livestock production. Agri Connect, Pretoria.

Robertson, A. & Rendel, J., 1954. The performance of heifers got by artificial insemination. J. Agric. Sci. 44, 184–192.

Robertson, A., Stewart, A. & Ashton, E. D., 1956. The progeny assessment of dairy sires for milk: the use of contemporary comparisons. Pages 43–50 in Proceedings of the British Society of Animal Production.

Rothschild, M. F. & Plastow, G. S., 2014. Applications of genomics to improve livestock in the developing world. Livest. Sci. 166, 76–83.

SA Stud Book., 2017. Genomic breeding values for Bonsmara. http://www.studbook.co.za/images/photos/News-BonsmaraGEBVs.pdf. Accessed 27 August 2019.

SA Stud Book., 2018. 2018 Logix elite news 1. http://www.studbook.co.za/images/photos/Logix-Elite-Newsletter-2018-April.pdf. Accessed 24 November 2019.

SA Stud Book., 2019. SA Stud Book Annual Bonsmara Report 2019.

Schaeffer, L. R., 1983. Effectiveness of model for cow evaluation intraherd. J. Dairy Sci. 66, 874–880.

Schaeffer, L. R., 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123, 218–223.

Schefers, J. M. & Weigel, K. A., 2012. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. Anim. Front. 2, 4–9.

Scholtz, M. M., Bester, J., Mamabolo, J. M. & Ramsay, K. A., 2008. Results of the national cattle survey undertaken in South Africa, with emphasis on beef. Appl. Anim. Husb. Rural Dev. 1, 1–9.

Seidel, G. E., 2010. Brief introduction to whole-genome selection in cattle using single nucleotide polymorphisms. Reprod. Fertil. Dev. 22, 138–144.

Sun, C., Madsen, P., Nielsen, U. S., Zhang, Y., Lund, M. S. & Su, G., 2009. Comparison between a sire model and an animal model for genetic evaluation of fertility traits in Danish Holstein population. J. Dairy Sci. 92, 4063–4071.

Taylor, J. F., Taylor, K. H. & Decker, J. E., 2016. Holsteins are the genomic selection poster cows. Proc. Natl. Acad. Sci. 113, 7690–7692.

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

U. N., 2019. World Population Prospects 2019: Data Booklet. New York United Nations Department for Economic Affairs and Social Affairs.

University of Arkansas System., 2015. Beef cattle production. https://www.uaex.edu/publications/pdf/mp184/MP184.pdf. Accessed 12 March 2020.

VanRaden, P. M., 2007. Genomic measures of relationship and inbreeding. Interbull Bull. 91, 4414–4423.

VanRaden, P. M., 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91, 4414–4423.

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F. & Schenkel, F. S., 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92, 16–24.

VanRaden, P. M., O'Connell, J. R., Wiggans, G. R. & Weigel, K. A., 2011. Genomic evaluations with many more genotypes. Genet. Sel. Evol. 43, 10.

Van der Westhuizen, B. & Mostert, B., 2020. Authentic South African genomic breeding values for the local dairy industry. Stockfarm, 47.

Van Marle-Köster, E., Visser, C. & Berry, D. P., 2013. A review of genomic selection - implications for the South African beef and dairy cattle industries. S. Afr. J. Anim. Sci. 43, 1–17.

Van Marle-Köster, E. & Visser, C., 2018a. Genetic improvement in South African livestock: can genomics bridge the gap between the developed and developing sectors? Front. Genet. 9, 1–12.

Van Marle-Köster, E. & Visser, C., 2018b. Genomics for the advancement of livestock production: A South African perspective. S. Afr. J. Anim. Sci. 48, 808-817.

Van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., Haudenschild, C. D., Moore, S. S., Warren, W. C. & Sonstegard, T. S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat. Methods 5, 247–252.

Vermeulen, H., Kirsten, J. & Sartorius, K., 2008. Contracting arrangements in agribusiness

procurement practices in South Africa. Agrekon 47, 198–221.

Vignal, A., Milan, D., SanCristobal, M. & Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol 34, 275–305.

Weigel, K. A., VanRaden, P. M., Norman, H. D. & Grosu, H., 2017. A 100-Year Review: Methods and impact of genetic selection in dairy cattle—From daughter–dam comparisons to deep learning algorithms. J. Dairy Sci. 100, 10234–10250.

Weller, P., Jeffreys, A. J., Wilson, V. & Blanchetot, A., 1984. Organization of the human myoglobin gene. EMBO J. 3, 439–446.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. & Tingey, S. V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18, 6531–6535.

Winkelman, A. M., Johnson, D. L. & Harris, B. L., 2015. Application of genomic evaluation to dairy cattle in New Zealand. J. Dairy Sci. 98, 659–675.

Wright, S., 1922. Coefficients of inbreeding and relationship. Am. Nat. 56, 330–338.

Addendum A

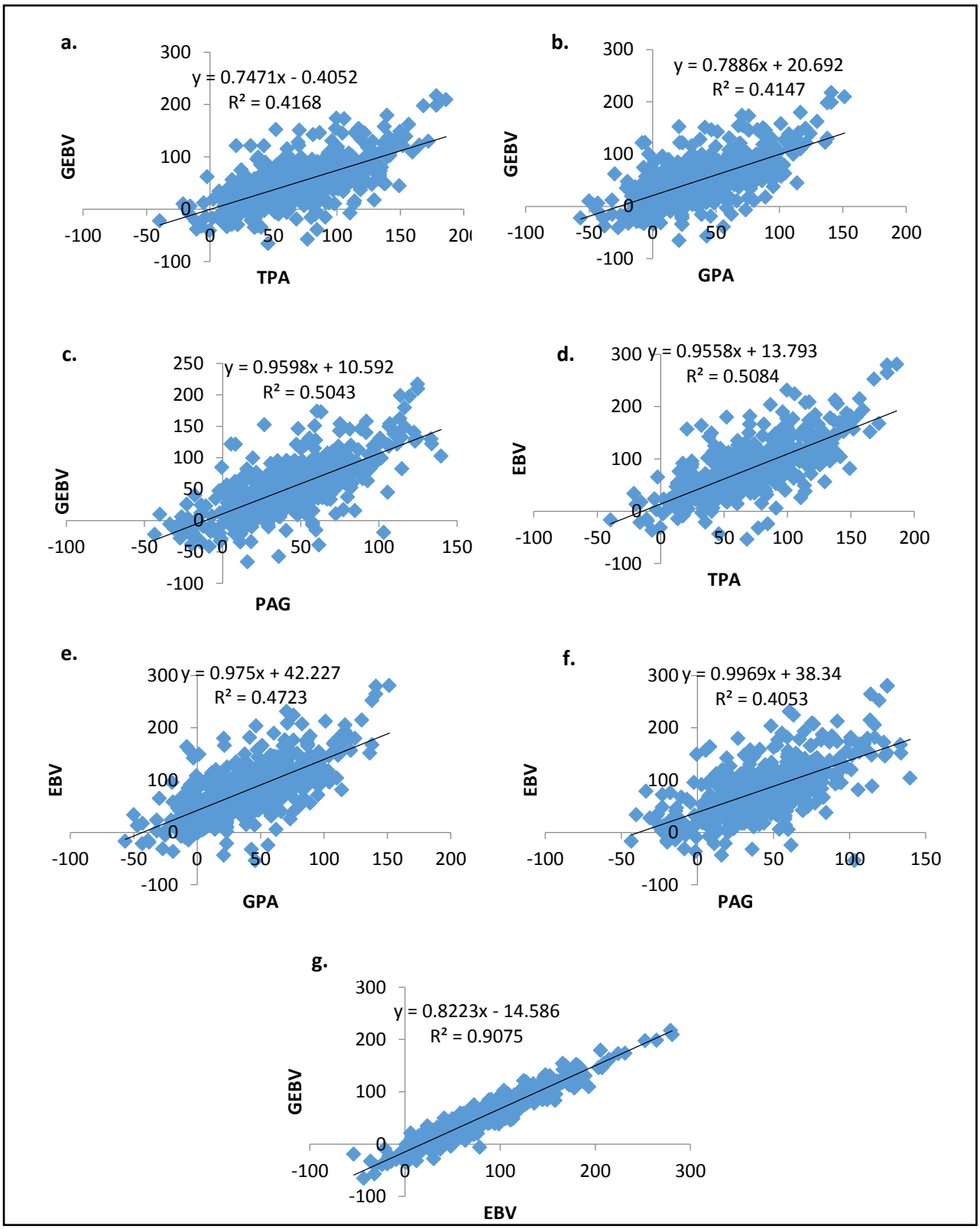

Figure A1. Pearson correlations of (a) TPA and GEBVs, (b) GPA and GEBVs (c) PAG and GEBVs, (d) TPA and EBVs, (e) GPA and EBVs, (f) PAG and EBVs, (g) EBVs and GEBVs for average daily gain.

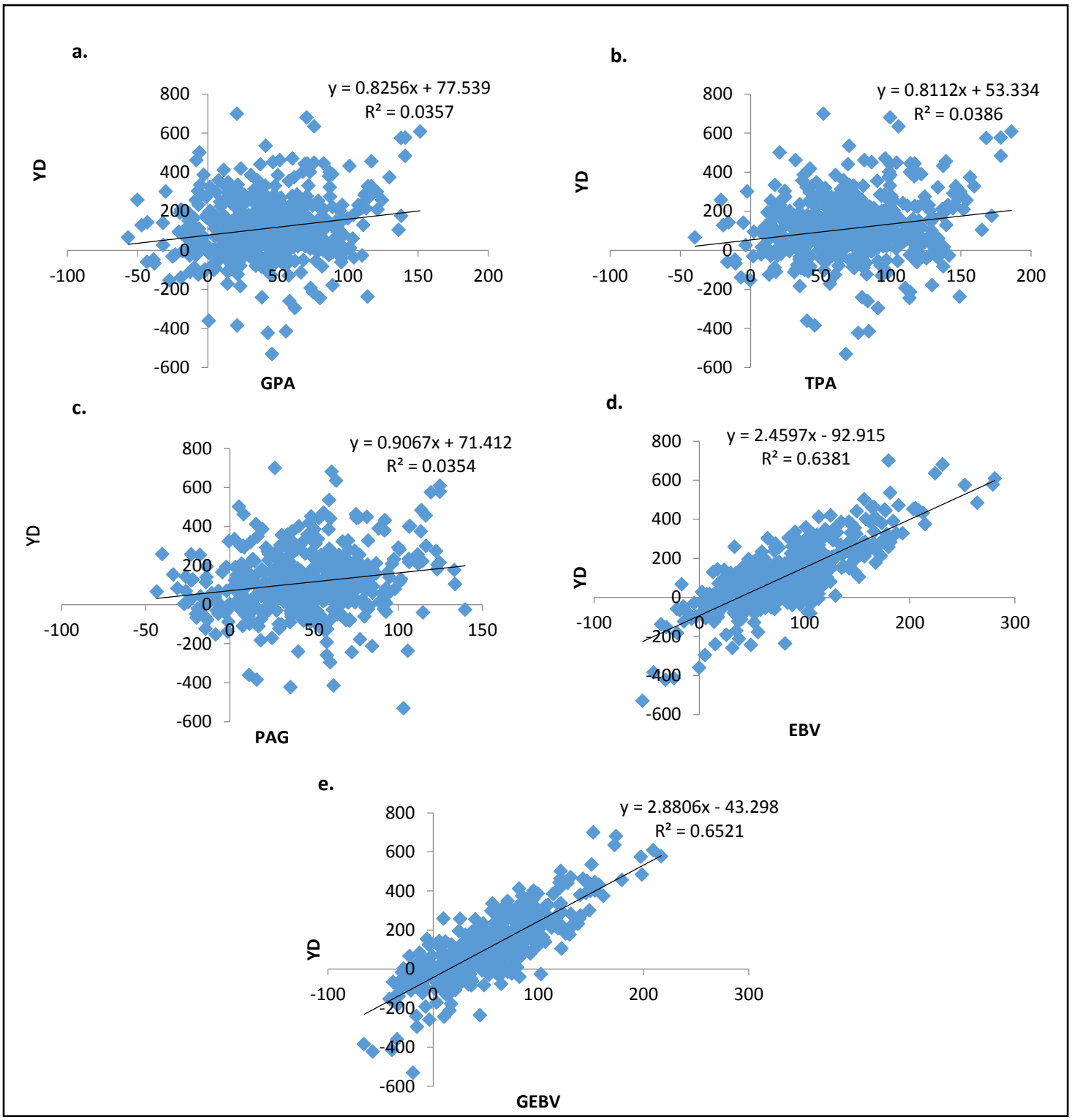

Figure A2. Pearson correlations of (a) GPA and YDs, (b) TPA and YDs, (c) PAG and YDs, (d) EBV and YDs, (e) GEBV and YDs for average daily gain.

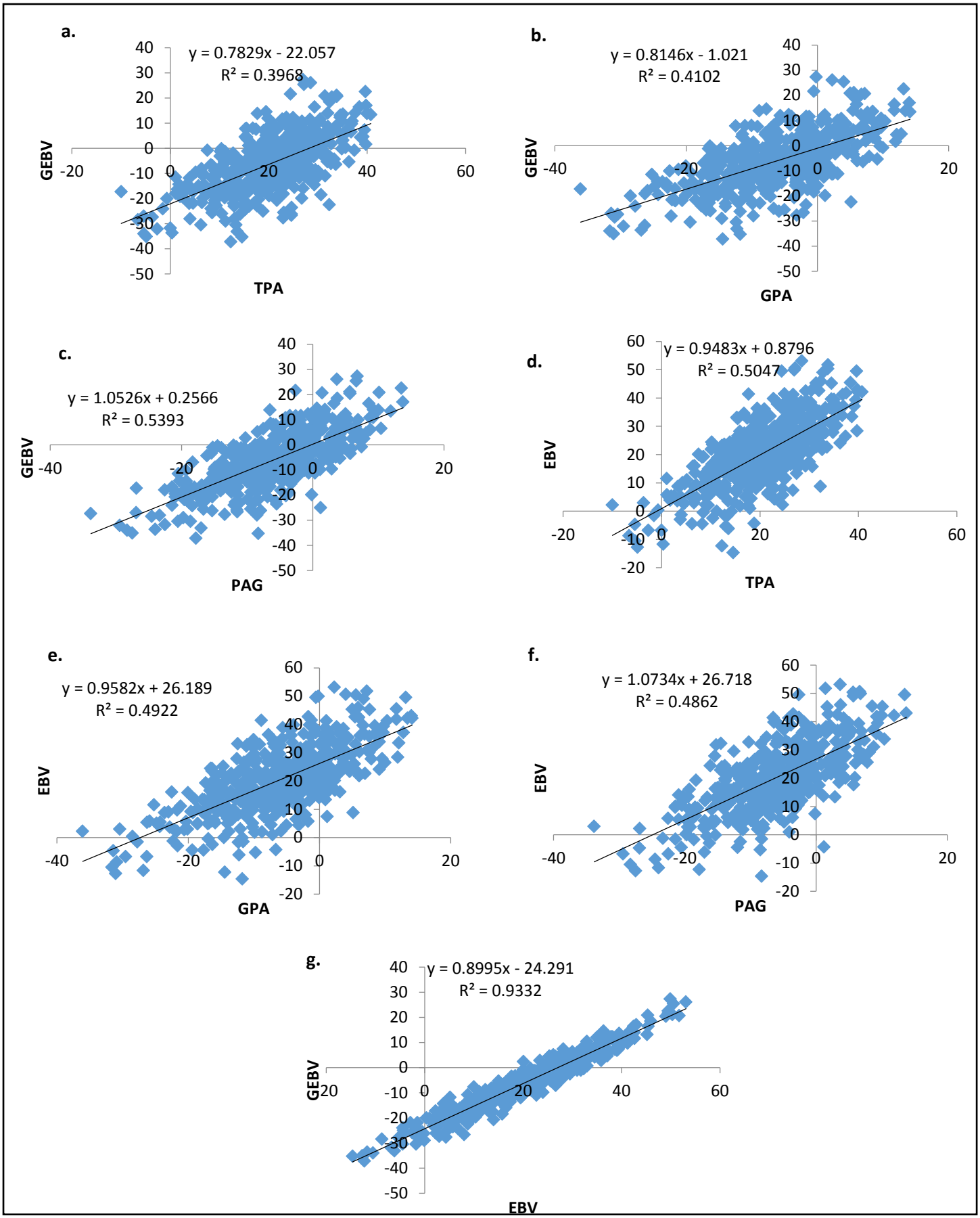

Figure A3. Pearson correlations of (a) TPA and GEBVs, (b) GPA and GEBVs (c) PAG and GEBVs, (d) TPA and EBVs, (e) GPA and EBVs, (f) PAG and EBVs, (g) EBVs and GEBVs for height.

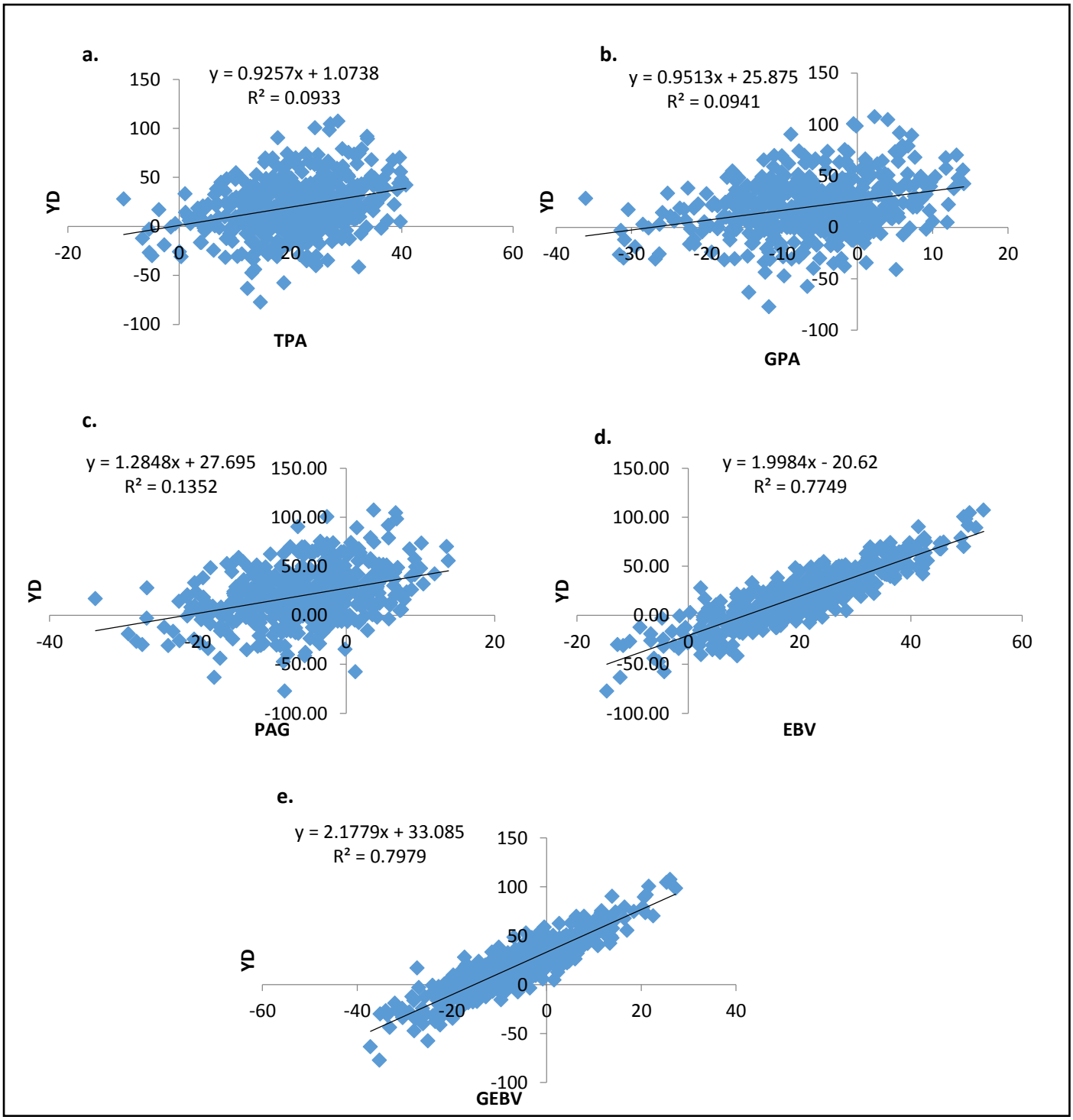

Figure A4. Pearson correlations of (a) TPA and YDs, (b) GPA and YDs, (c) PAG and YDs, (d) EBV and YDs, (e) GEBV and YDs for height.

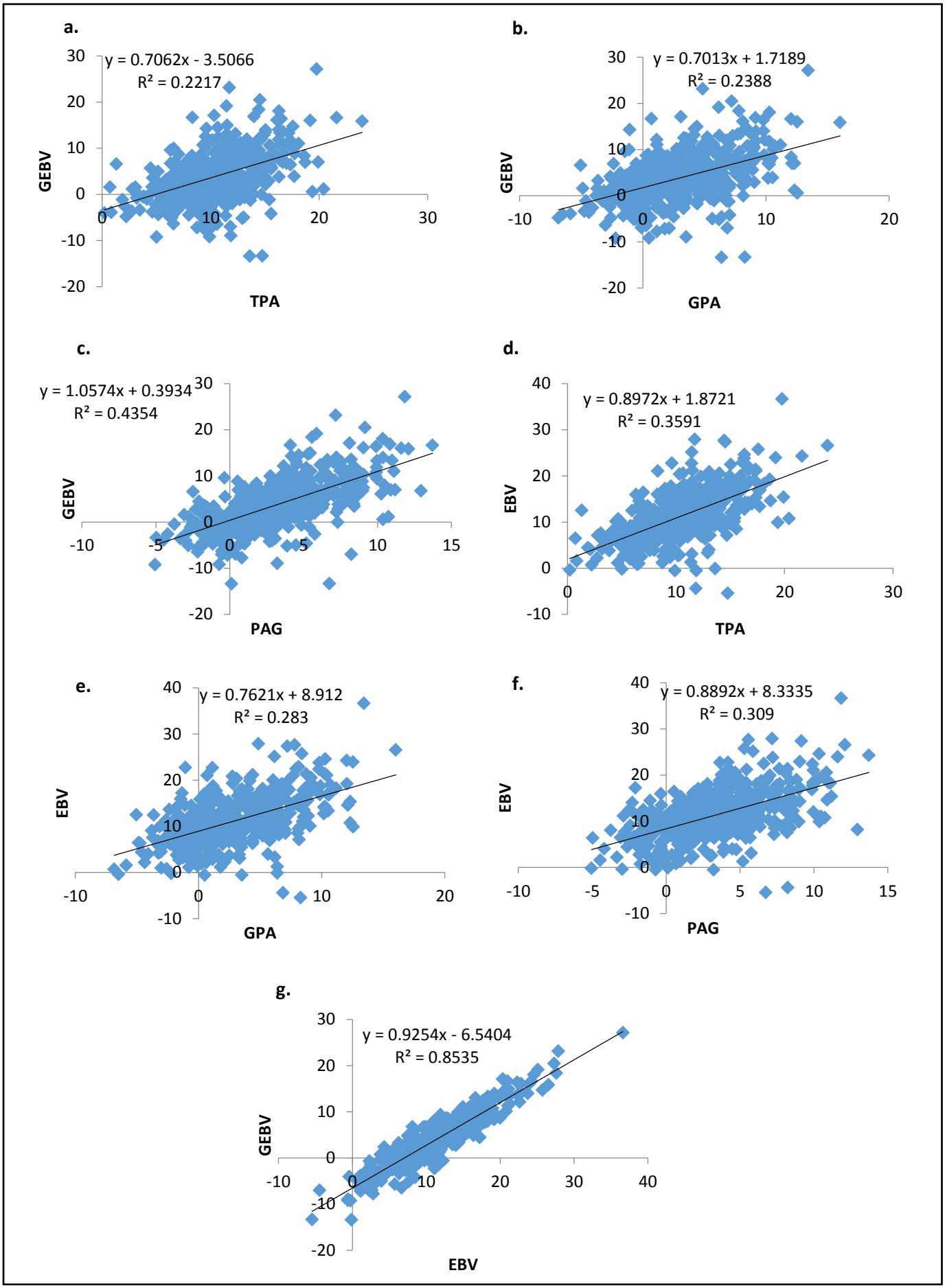

Figure A5. Correlations of (a) TPA and GEBVs, (b) GPA and GEBVs (c) PAG and GEBVs, (d) TPA and EBVs, (e) GPA and EBVs, (f) PAG and EBVs, (g) EBVs and GEBVs for direct weaning weight.

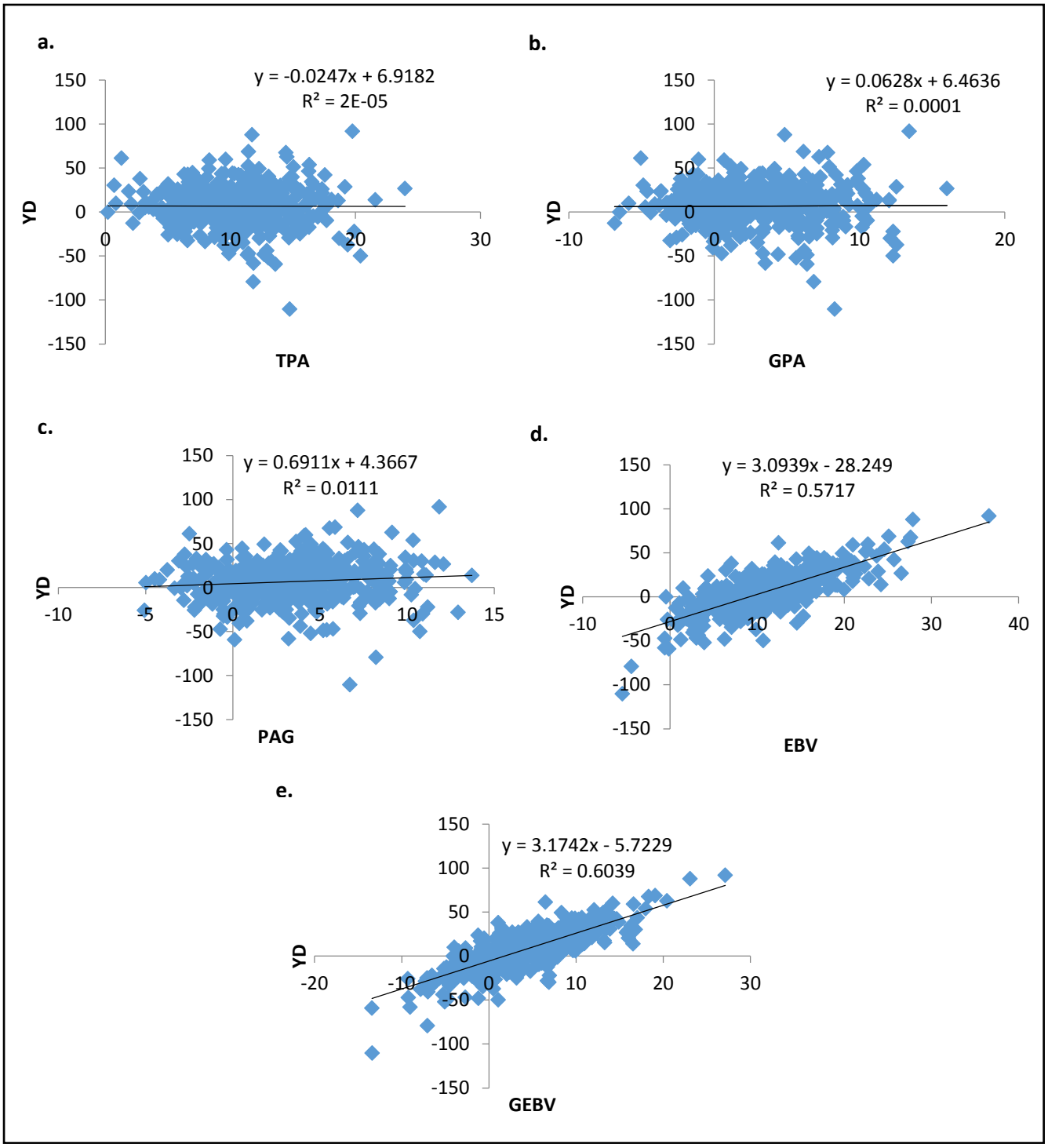

Figure A6. Correlations of (a) TPA and YDs, (b) GPA and YDs, (c) PAG and YDs, (d) EBV and YDs, (e) GEBV and YDs for direct weaning weight.

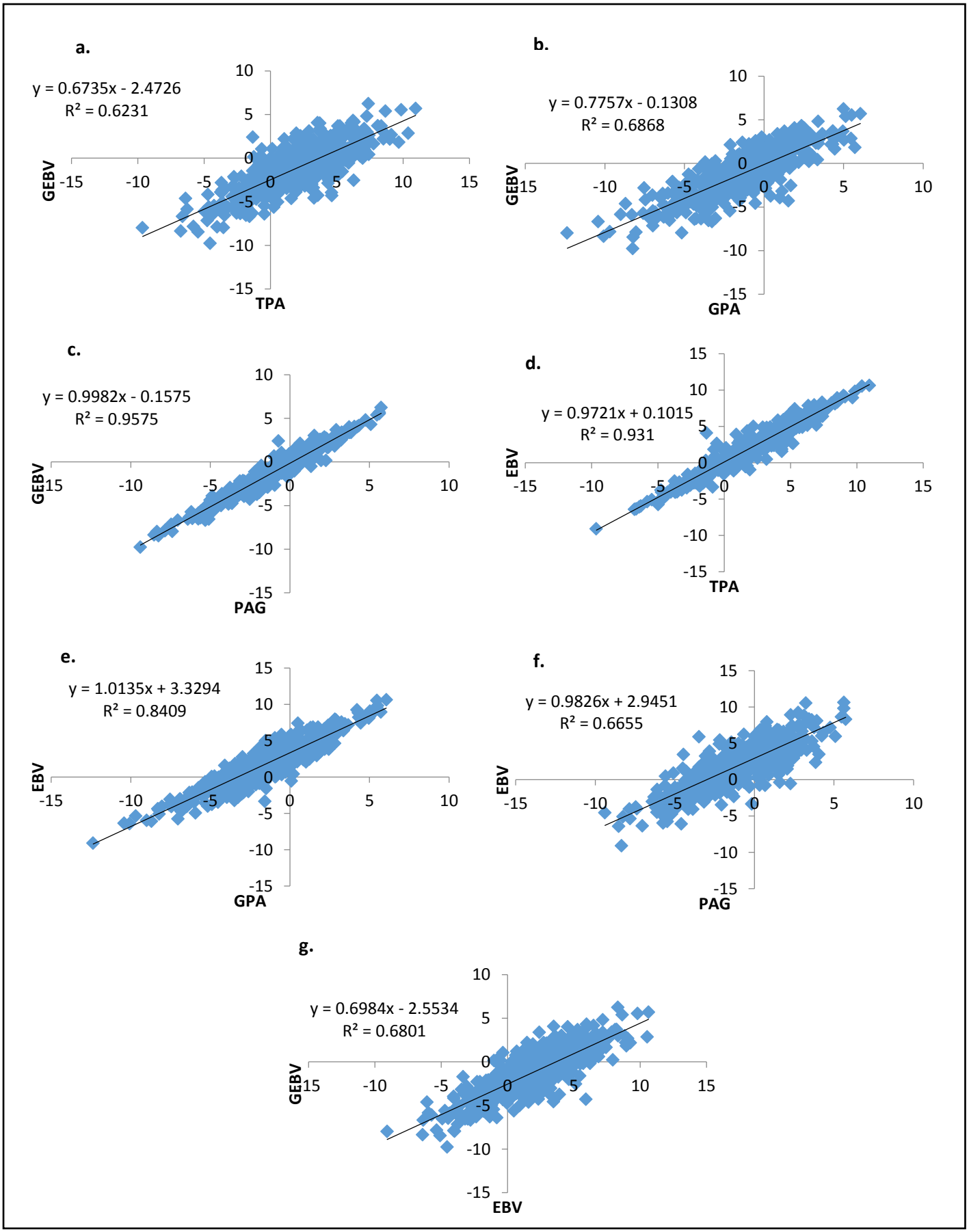

Figure A7. Correlations of (a) TPA and GEBVs, (b) GPA and GEBVs (c) PAG and GEBVs, (d) TPA and EBVs, (e) GPA and EBVs, (f) PAG and EBVs, (g) EBVs and GEBVs for maternal weaning weight.

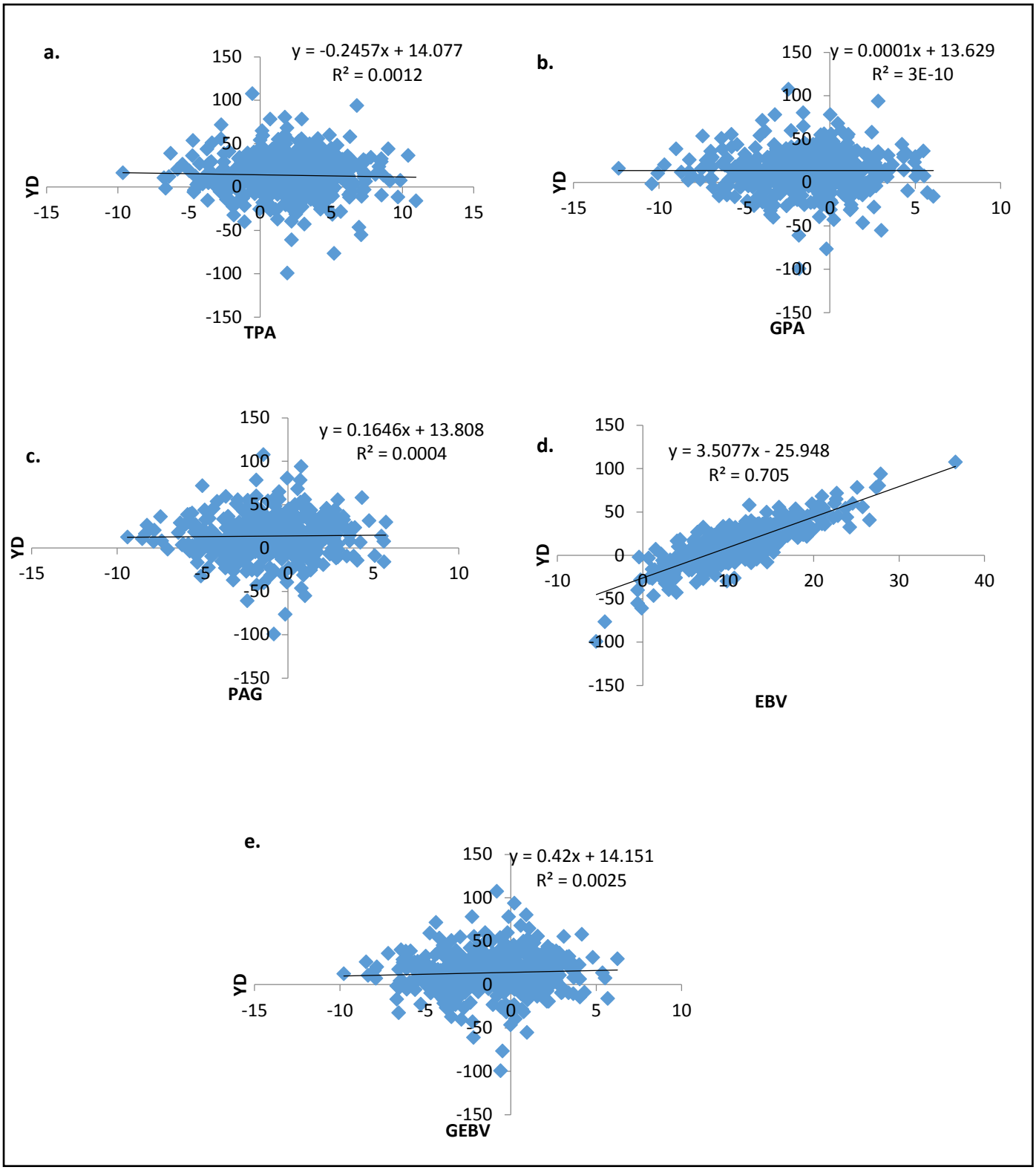

Figure A8. Correlations of (a) TPA and YDs, (b) GPA and YDs, (c) PAG and YDs, (d) EBV and YDs, (e) GEBV and YDs for maternal weaning weight.